



HAL
open science

Proceedings of ICCSA 2009 - 3rd International Conference on Complex Systems and Applications

Cyrille Bertelle, Xinzhi Liu, M A Aziz-Alaoui

► **To cite this version:**

Cyrille Bertelle, Xinzhi Liu, M A Aziz-Alaoui (Dir.). Proceedings of ICCSA 2009 - 3rd International Conference on Complex Systems and Applications. 2009. hal-03313772

HAL Id: hal-03313772

<https://hal.science/hal-03313772>

Submitted on 4 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ICCSA 2009

June 29 - July 02

3rd International Conference on
**Complex Systems
and Applications**

Conference Proceedings General Sessions

Cyrille Bertelle, Xinzhi Liu, M.A.Aziz-Alaoui (eds)



Université du Havre

ICCSA 2009

**Third International Conference on Complex
Systems and Applications**

General Sessions

June 29 – July 02 2009

Le Havre, Normandy, France

Cyrille Bertelle, Xinzhi Liu and M.A. Aziz-Alaoui (eds)

Advisory Committee:

- Auger P., Académie des Sciences, France
- Bourguine P., ISC, Paris, France
- Chen G., City University of Hong-Kong
- Chua O.L., University of California at Berkeley, USA
- Demongeot J., IUF, TIMC Grenoble, France
- Françoise J.-P., UPMC, Paris 6, France
- Grebogi C., University of Aberdeen, UK
- Pumain D., ISC, Paris, France
- Rossler O., IPTC, Tübingen, Germany
- Siljak D.D., Santa Clara University, USA

General Chairs:

- Cyrille Bertelle, University of Le Havre, France
- Xinzhi Liu, University of Waterloo, Canada
- M.A. Aziz-Alaoui, University of Le Havre, France

Local Organization Committee:

- M. Aziz-Alaoui, LMAH, Le Havre, France
- Stefan Balev, LITIS, Le Havre, France
- Julien Baudry, LITIS, Le Havre, France
- Cyrille Bertelle, LITIS, Le Havre, France
- Nathalie Corson, LMAH, Le Havre, France
- Antoine Dutot, LITIS, Le Havre, France
- Rawan Ghnemmat, LITIS, Le Havre, France
- Frédéric Guinand, LITIS, Le Havre, France
- Rabah Labbas, LMAH, Le Havre, France
- Djamila Moulay, LMAH, Le Havre, France
- Damien Olivier, LITIS, Le Havre, France
- Yoann Pigné, LITIS, Le Havre, France
- Claire Roussin, LITIS-LMAH, Le Havre, France
- Nathalie Verdière, LMAH, Le Havre, France
- Adnan Yassine, LMAH, Le Havre, France

Regional Organization Committee:

- Abdulrab H., INSA Rouen, France
- Bouzid M., University of Caen, France
- Caputo J.-G., INSA Rouen, France
- Cherrier E., ENSI of Caen, France
- Daude E., University of Rouen, France
- Elissalde B., University of Rouen, France
- Letellier C., University of Rouen, France
- Mutabazi I., University of Le Havre, France

International Scientific Committee:

- Afraimovich V., IICU, University of San Luis Potisi, Mexico
- Ahmed N.U., University of Ottawa, Canada
- Akhmet M.U., Middle East Technical University, Turkey
- Ayesha A., De Montfort University, UK
- Banos A., LIV, Strasbourg, France
- Barbot J.P., ENSEA, Cergy, France
- Basin M., Autonomous University of Nuevo Leon, Mexico

- Belykh I., Department of Mathematics and statistics, Georgia State University, USA
- Biswas S.K., Temple University, USA
- Bollt E.M., Clarkson University, USA
- Boukas E. K., Polytechnique de Montreal, Canada
- Boutayeb M., Strasbourg University, France
- Chavalarias D., CREA, Ecole polytechnique, France
- Collet P., Centre de Physique théorique, Ecole polytechnique, Paris
- Cotsaftis M., LTME/ECE, Paris, France
- Courbage M., Université Paris Diderot Paris 7, France
- Danca M.F., Avram Iancu University, Romania
- Datta B.N., Northern Illinois University, USA
- Diehl M., K.U.Leuven, Belgium
- Duan, Z., Peking University, China
- Duchamp G.H.E., Paris XIII University, France
- Elwakil A.S., University of Sharjah, Emirates
- Gerdtts M., University of Birmingham at Edgbaston, UK
- Ginoux J.-M., PROTEE, Toulon, France
- Gulyas L., Lorand Eotvos University, Hungary
- Han Q., Central Queensland University, Australia
- Ho D., City University of Hong Kong
- Hong, L., Wright State University, USA
- Ishitobi, M., Kumamoto University, Japan
- Jost C., University of Paul Sabatier, Toulouse, France
- Kolumban G., Budapest University of Technology and Economics, Hungary
- Kurths J., Postdam University, Germany
- Lam L. University of Hong Kong
- Letellier C., University of Rouen, France
- Li B.-L., California University, USA
- Li T., Tsinghua University, China
- Lozi R., Nice University, France
- Lu Jinhu, Chinese Academy of Sciences, China
- Mao X., University of Strathclyde, UK
- Migorski S., Jagiellonian University, Poland
- Momani S., Mutah University, Jordan
- Mordukhovich, B.S., Wayne State University, USA
- Müller J.-P., CIRAD, Montpellier, France
- Odibat Z., Al Balqa Applied University, Jordan
- Paolini E., Universidad Nacional del Sur, Argentina
- Respondek W., INSA Rouen, France
- Rossetto B., PROTEE, Toulon, France
- Sanjuan M., Department of physics, Universidad Rey Juan Carlos, Madrid, Espana
- Shen, X., University of Waterloo, Canada
- Smith R., North Carolina State University, USA
- Tang W.K.S., City University of Hong Kong, China
- Teo K.L., Curtin University of Technology, Australia
- van Wyk M., University of the Witwatersrand, South Africa
- Volny D., University of Rouen, France
- Wang P.K.C., University of California at Los Angeles, USA
- Ying H., Wayne State University, USA
- Yong J., University of Central Florida, USA
- Yu P., University of Western Ontario, Canada
- Yue W., Konan University, Kobe, Japan

Contents

Contents	iii
1 Plenary Talks	1
Memristor: 37 years later	
Leon O. CHUA	2
Coupled mixed-mode dynamics and associated propagations	
Jean-Pierre FRANÇOISE	3
Complex systems in pathologic processes	
Jacques DEMONGEOT	4
Reduction of complexity of dynamical systems: applications to fishery modelling	
Pierre AUGER	5
Evolution theory and the place of man	
Otto ROSSLER	6
Multi-Scale Dynamics Reconstruction	
Paul BOURGINE	7
Understanding complex systems: a survey of phenomenological, physical and structural approaches	
Michel COTSAFTIS	8
Cascades of period doublings galore	
James A. YORKE	9
Swarm Intelligence, Swarm-bots and Swarmanoids	
Marco DORIGO	10
Synchronization and complex networks: are such theories useful for Earth sciences?	
Jürgen KURTHS, J. DONGES, N. MARWAN, Y. ZOU	11
Collective motion: from minimal models to starling flocks	
Hugues CHATÉ	12
2 Dynamical Systems, Synchronization	13
Applications of subpositive definite matrices for stability of general non-linear systems	
Feng LIU, Guodong SHI, Zhiqing WENG	15

Chaotic pseudo random number generators via ultra weak coupling of chaotic maps and double threshold sampling sequences	
René LOZI	20
Periodic solution for fourth-order discrete Hamiltonian systems	
Qiong MENG	25
Homoclinic orbits for a class of nonperiodic Hamiltonian systems	
Jian DING, Guizhen FENG, Weili WU	29
Frequency bifurcation by singularities of congruence	
Roger TAULEIGNE	34
Calculation of the largest Lyapunov exponent for a particular hybrid dynamical systems class	
Céline QUÉMARD	40
Forced synchronization of coupled hyperchaotic oscillators : application to an authentication process	
Laurent LAVAL	41
Observers design for a new weakly coupled map function	
Sébastien HÉNAFF, Ina TARALOVA, René LOZI	47
Periodic dynamical systems in the phase space	
Bruno ROSSETTO, Jean-Marc GINOUX	51
Network synchronization of unified chaotic systems in master-slave coupling	
Cesar CRUZ-HERNANDEZ, Rosa Martha LÓPEZ-GUTIÉRREZ, Everardo INZUNZA-GONZÁLEZ, Liliana CARDOZA-AVENDAÑO	56
Differential geometry applied to dynamical systems	
Jean-Marc GINOUX, Bruno ROSSETTO	61
High gain observer based synchronization for a class of time-delay chaotic systems. Application to secure communications	
Estelle CHERRIER, Mohammed M'SAAD	65
Observer based approach for synchronization in complex dynamical networks	
Mohamed BOUTAYEB, Ali ZEMOUCHE, Jean-Jacques SLOITINE	70
LPV approach for the stabilization of a class of dynamical systems	
Ali ZEMOUCHE, Mohamed BOUTAYEB	74
Mathematical study for a tuberculosis model with two differential infectivity and n classes of latent	
Samuel BOWONG, Yves EMVUDU, Dany Pascal MOUALEU, Jean Jules TEWA	79
Complex dynamics of cellular automata emerging in chaotic rules	
Genaro J. MARTÍNEZ, Andrew ADAMATZKY, Ramon ALONSO-SANZ	91
Bifurcations of a neuron oscillator	
Nathalie CORSON	100
The chikungunya epidemic: modelling and dynamics	
Djamila MOULAY	105
Synchronization in inhibitory networks of bursting neurons	
Igor BELYKH, Sajiya JALIL, Andrey SHILNIKOV	110
3 Signal Processing	111
A novel data pre-processing method on automatic determinig of sleep stages: k-means clustering based feature weighting	
Salih GÜNEŞ, Kemal POLAT, Şebnem YUSUNKAYA, Mehmet DURSUN	112

Spectral kurtosis and wavelets' entropy for transients' enhancement: application to ter- mite detection	
Juan José GONZÁLEZ DE LA ROSA, Antonio MORENO-MUNOZ, Antolino GALLEGO, Rosa PIOTRKOWSKI, Enrique CASTRO	118
Modeling and simulation about rarefaction wave gun interior ballistics applying inertial brechblock	
Zhi JIANZHUANG, Zheng JIAN, Di CHANGCHUN	124
The effect of AR model degrees to early detection of rheumatoid arthritis disease	
Ali Osman ÖZKAN, Sadik KARA, Ali SALLI, Mehmet Emin SAKARYA, Serter GÜMÜŞ, Ayse ÜNAL, Salih GÜNEŞ	125
4 Nonlinear Optimization	131
Nonlinear regression analysis of optimal conditions for a physics-chemical process	
Vyacheslav RUSANOV, Sergey AGAFONOV, Alexey DANEEV, Dmitry SHARPINSKIY, Timofey UDILOV	132
H_∞ optimal model reduction of complex systems using particle swarm optimization	
Maamar BETTAYEB, Reem SALIM	138
H2 optimal model reduction of dynamic systems with time-delay using particle swarm optimization	
Reem SALIM, Maamar BETTAYEB	144
Optimization algorithm for combined economic and emission dispatch with security con- straints	
Nurettin ÇETINKAYA	150
Bandwidth optimization of a PIFA antenna using hybrid genetic algorithms	
Mohammad Riyad AMEERUDDEN, Harry C. S. RUGHOOPUTH	154
5 Heuristics, Decision and Optimization	159
Artificial neural network training models in prediction of concrete compressive strength using Euclidean normalization method	
Akdemir BAYRAM, Salih GÜNEŞ, A GENÇ	160
Concept clustering on interval concept lattice	
Wen ZHOU, Yan ZHAO, Wei LIU, Dong XU, Zongtian LIU, Bofeng ZHANG, Mingjun XIN	166
Dagda, a load-balanced middleware to distribute complex systems simulations	
Antoine DUTOT, Damien OLIVIER, Guilhelm SAVIN	171
GeoCells model : European Structural Funds and regional interactions, which conver- gences for the European regions?	
Bernard ELISSALDE, Patrice LANGLOIS, Dominique GOYAT	175
Hybrid approaches for stock price prediction	
Wen QINGHUA, Yang ZEHONG, Song YIXU	202
A new method to improve network exchanges of multimedia stream	
Bechir ALAYA, Claude DUVALLET, Bruno SADEG	206
Index	213

Plenary Talks

Contents

Memristor: 37 years later	
Leon O. CHUA	2
Coupled mixed-mode dynamics and associated propagations	
Jean-Pierre FRANÇOISE	3
Complex systems in pathologic processes	
Jacques DEMONGEOT	4
Reduction of complexity of dynamical systems: applications to fishery modelling	
Pierre AUGER	5
Evolution theory and the place of man	
Otto ROSSLER	6
Multi-Scale Dynamics Reconstruction	
Paul BOURGINE	7
Understanding complex systems: a survey of phenomenological, physical and structural approaches	
Michel COTSAFTIS	8
Cascades of period doublings galore	
James A. YORKE	9
Swarm Intelligence, Swarm-bots and Swarmanoids	
Marco DORIGO	10
Synchronization and complex networks: are such theories useful for Earth sciences?	
Jürgen KURTHS, J. DONGES, N. MARWAN, Y. ZOU	11
Collective motion: from minimal models to starling flocks	
Hugues CHATÉ	12

Memristor: 37 years later

Leon O. CHUA

University of California, N.O.E.L Berkeley
USA

Coupled mixed-mode dynamics and associated propagations

Jean-Pierre FRANÇOISE

UMR 7598, Université P.-M. Curie, Paris6
France

Complex systems in pathologic processes

Jacques DEMONGEOT

TIMC-IMAG, Grenoble
France

The phenotypic symptoms in many pathologies result from the interactions between various elements of an organism at numerous levels - genetic disorders due to mutations or translocations, cell dedifferentiation, vascular neogenesis,... - and these emerging properties are giving medical signs at certain opportune times, the "kairos" (in the terminology proposed by Herophilus). We discuss possible mathematical models of pathologic processes underlying infectious and oncogenetic diseases, in which disorders are visible after a certain evolution and we suggest some measurements for accelerating the diagnostic phase.

Reduction of complexity of dynamical systems: applications to fishery modelling

Pierre AUGER

Académie des Sciences, Paris
France

Population and community models are complex in the sense that they usually take into account many variables and parameters. We present methods allowing the reduction of the system which may become, in this way, mathematically more tractable. The reduction of the proposed system is undertaken with the help of aggregation methods which aim at studying the relationships between a large class of complex systems, in which many variables are involved, and their corresponding reduced or aggregated systems, governed by a few variables. The kind of aggregation methods that we consider is based on time scale separation methods. In order to illustrate the methods in population and community dynamics, I present an application of these aggregation methods in the context of fishery modelling. We present a spatial stock-fishing effort model with several fishing sites. We study the effects of fast price variation as a result of demand and supply on the global dynamics of the fishery.

Evolution theory and the place of man

Otto ROSSLER

University of Tübingen
Germany

Max Scheler invented the phrase "The place of man in the cosmos" in 1928. Physicist-biologist Teilhard de Chardin was still young at the time. The question is important because of the great risk that the human species presents to the planet since the 20th century, with the almost atomic-war in the 1960s, and the current century's first decade with the almost-risk of Armageddon in the LHC. Is man a biological creature alone, despite his jump out of biology? The brain equation allows to pinpoint the biological place of man (and of all other non-plantlike intelligences and the brain in the genome, the doubtlessly highest intelligence). But all these intelligences are good-natured. They are like the angels of traditional Aquinean philosophy - without responsibility. The jump-out, by the human toddler, is a jump into the lap of point Omega. It can be deadly, because for the first (and perhaps last) time, a species has on the level of the individual assumed personal responsibility for other individuals (and species and ecosystems and planets). Schopenhauer would call it "pity", Lévinas "exteriority". What it really is is an appreciation of benevolence being shown towards oneself (and a reciprocating by one's showing benevolence). Only a creator of a higher-level world can fully appreciate this. Is science ready to deal with the mechanism without missing (by overlooking it) the implicit infinite light?

Multi-Scale Dynamics Reconstruction

Paul BOURGINE

CREA & RNSC, Paris
France

Understanding complex systems: a survey of phenomenological, physical and structural approaches

Michel COTSAFTIS

ECE, Paris
France

Modern Technology advance and much more accurate observations from highly improved sensors have been unravelling recently the existence of a huge class of systems in almost all domains of research exhibiting properties outside classical Science main stream, from non strict causality to emergence phenomenon capturing today the attention. In order to understand these unexpected occurrences, some of which having huge impact on industrial and human activity on the Earth, different ways of approach have been started, mainly in the line of previous methods already in use at the places where the study was undertaken, and borrowed by other new places entering the subject. Very contrasted results have been obtained to date with sometimes confusion as concerns their interpretation, often due to a lack of agreement on the very definition of the object “complex system”.

To help for its understanding, it is proposed in the survey to review the various methods presently in use which are shown to belong to three main streams inherited from the past. Keys will be given for locating their potential in the big landscape picture already available from collected information, principally by balancing the result they can access to with the nature of the information they are working with. It will mainly be shown that if the three lines may have their own merit, some targets are completely out of reach in a specific line. To facilitate a common understanding, the presentation will be starting from very fundamental and elementary concept, out of which the split in the approaches is more easily understandable. The conclusion will aim at providing some possible directions for future developments, as well as a new global picture of the Universe emerging today from the addition of discussed complex effects.

Cascades of period doublings galore

James A Yorke

*Institute for Physical Sciences and Technology
University of Maryland
College Park, MD 20742
USA*

A Period Doubling Cascade is something many people know about, especially in view of Feigenbaum's scaling law. But there is much more to know. For example little has been written on why they exist. When you see a bifurcation diagram, you see a cascade in each window as a parameter is varied. Here is a key fact: There are an infinite number of cascades when there is one. In work with Evelyn Sander, we find that the amount of chaos in the system as the parameter approaches infinity determines which cascades exist. Here there is potential to characterize the chaos as the parameter approaches infinity for various systems. That will explain its pattern of cascades. We have proved there is an invariance, namely that large local perturbations cannot destroy Cascades. The perturbations simply shift the position of the cascade.

Swarm Intelligence, Swarm-bots and Swarmanoids

Marco DORIGO

IRIDIA, Free University of Brussels
Belgium

Swarm intelligence is the discipline that deals with natural and artificial systems composed of many individuals that coordinate using decentralized control and self-organization. In particular, it focuses on the collective behaviors that result from the local interactions of the individuals with each other and with their environment. The characterizing property of a swarm intelligence system is its ability to act in a coordinated way without the presence of a coordinator or of an external controller.

Swarm robotics could be defined as the application of swarm intelligence principles to the control of groups of robots.

In this talk I will discuss results of Swarm-bots, an experiment in swarm robotics. A swarm-bot is an artifact composed of a swarm of assembled s-bots. The s-bots are mobile robots capable of connecting to, and disconnecting from, other s-bots. In the swarm-bot form, the s-bots are attached to each other and, when needed, become a single robotic system that can move and change its shape. S-bots have relatively simple sensors and motors and limited computational capabilities. A swarm-bot can solve problems that cannot be solved by s-bots alone. In the talk, I will shortly describe the s-bots hardware and the methodology we followed to develop algorithms for their control. Then I will focus on the capabilities of the swarm-bot robotic system by showing video recordings of some of the many experiments we performed to study coordinated movement, path formation, self-assembly, collective transport, shape formation, and other collective behaviors.

I will conclude presenting initial results of the Swarmanoid experiment, an extension of swarm-bot to 3-dimensional environments.

Synchronization and complex networks: are such theories useful for Earth sciences?

Jürgen KURTHS, J. DONGES, N. MARWAN, Y. ZOU

University of Postdam
Germany

Synchronization phenomena are abundant in nature, science, engineering and social life, such as in organ pipes, fireflies and even in the mechanics of bridges. But synchronization was first recognized by Christiaan Huygens in 1665 for coupled pendulum clocks; this was the beginning of nonlinear sciences. In the last two decades, this concept has been successfully extended to more complex systems, as identification of teleconnections in the climate system.

Complex networks were firstly studied by Leonhard Euler in 1736 when he solved the Königsberger Brückenproblem. Recent research has revealed a rich and complicated network topology in various model systems as well as in several fields of applications, such as transportation and social networks, or the WWW. It will be discussed whether this approach can lead to useful new insights into rather large complex systems or whether it is fashionable only to interpret various phenomena from this viewpoint and publish papers on that.

Many promising approaches have already lead to useful applications, e.g. immunization problems (spreading of diseases), functioning of biological/physiological processes as protein networks, brain dynamics, or functioning of social networks as network of vehicle traffic in a region or air traffic.

A challenging task is to understand the implications of such network structures on the functional organization of the system Earth. We show especially that the climate system can be interpreted as a highly dynamic network. This approach gives new insights into the vulnerability of the system Earth.

References Pikovsky, A., M. Rosenblum, and J. Kurths, *Synchronization – A Universal Concept in Nonlinear Sciences*, Cambridge University Press 2001. Maraun, D., and J. Kurths, *Geophys. Res. Lett.* 2005, 32, L15709. Osipov, G.V., J. Kurths, and C. Zhou, *Synchronization in Oscillatory Networks*, Springer Complexity, Berlin 2007. Arenas, A., A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, *Phys. Reports* 2008, 469, 93. Donges, J., Y. Zou, N. Marwan, and J. Kurths, *Europ. Phys. J. ST* 2009 (in press).

Collective motion: from minimal models to starling flocks

Hugues CHATÉ

CEA, Saclay
France

The collective properties of self-propelled particles trying to align are often striking, even in the simplest setting where the fluid in which they evolve is neglected and no interaction keeps them together. I will first describe these remarkable emerging properties of minimal models for collective motion, stressing their genericity and universality. In a second part, I will introduce a cohesion force in an effort to get closer, always with minimal ingredients, to the case of flocks of starlings, as observed recently within the Starflag project.

Dynamical Systems, Synchronization

Contents

Applications of subpositive definite matrices for stability of general non-linear systems Feng LIU, Guodong SHI, Zhiqing WENG	15
Chaotic pseudo random number generators via ultra weak coupling of chaotic maps and double threshold sampling sequences René LOZI	20
Periodic solution for fourth-order discrete Hamiltonian systems Qiong MENG	25
Homoclinic orbits for a class of nonperiodic Hamiltonian systems Jian DING, Guizhen FENG, Weili WU	29
Frequency bifurcation by singularities of congruence Roger TAULEIGNE	34
Calculation of the largest Lyapunov exponent for a particular hybrid dynamical systems class Céline QUÉMARD	40
Forced synchronization of coupled hyperchaotic oscillators : application to an authen- tication process Laurent LAVAL	41
Observers design for a new weakly coupled map function Sébastien HÉNAFF, Ina TARALOVA, René LOZI	47
Periodic dynamical systems in the phase space Bruno ROSSETTO, Jean-Marc GINOUX	51
Network synchronization of unified chaotic systems in master-slave coupling Cesar CRUZ-HERNANDEZ, Rosa Martha LÓPEZ-GUTIÉRREZ, Everardo INZUNZA-GONZÁLEZ, Liliana CARDOZA-AVENDAÑO	56
Differential geometry applied to dynamical systems Jean-Marc GINOUX, Bruno ROSSETTO	61
High gain observer based synchronization for a class of time-delay chaotic systems. Application to secure communications Estelle CHERRIER, Mohammed M'SAAD	65
Observer based approach for synchronization in complex dynamical networks Mohamed BOUTAYEB, Ali ZEMOUCHE, Jean-Jacques SLOTINE	70

LPV approach for the stabilization of a class of dynamical systems	
Ali ZEMOUCHE, Mohamed BOUTAYEB	74
Mathematical study for a tuberculosis model with two differential infectivity and n	
classes of latent	
Samuel BOWONG, Yves EMVUDU, Dany Pascal MOUALEU, Jean Jules TEWA	79
Complex dynamics of cellular automata emerging in chaotic rules	
Genaro J. MARTÍNEZ, Andrew ADAMATZKY, Ramon ALONSO-SANZ	91
Bifurcations of a neuron oscillator	
Nathalie CORSON	100
The chikungunya epidemic: modelling and dynamics	
Djamila MOULAY	105
Synchronization in inhibitory networks of bursting neurons	
Igor BELYKH, Sajiya JALIL, Andrey SHILNIKOV	110

Applications of Subpositive Definite Matrices for Stability of General Non-Linear Systems

Feng Liu , Guodong Shi, Zhiqing Weng

Abstract—Stability of general non-linear systems at positive equilibrium state is an important problem,many articles have rather thoroughly researched it, and some criterions have been established based on M matrix and positive definite matrix. In this paper,subpositive definite matrices are led into general non-linear systems, and some new criterions for general non-linear systems on overall stability are acquired. These results improve or expand some existing criterions.

Index Terms—General non-linear systems, stability, subpositive definite matrix

I. INTRODUCTION

Stability of systems at positive equilibrium state is an important problem, and many articles have established some criterions for Volterra systems, Gilpin-Ayala systems and general systems based on M matrix and positive definite matrix, see [1~8]. In this paper, Subpositive definite matrices are led into general non-linear systems, and some new criterions for general non-linear systems on overall stability at positive equilibrium position are acquired. For this reason, firstly the definition of subpositive (subnegative) definite matrix and the explaining of some signs are presented as follows:

Definition[9] Suppose A is a n -order matrix,if $\forall 0 \neq x \in R^n$,and always $f = x^T A x > 0 (< 0)$,then A is called Subpositive definite matrix(subnegative definite matrix) (A is not always symmetry matrix) .

It is easy to prove that A is subpositive definite matrix(subnegative definite matrix) if and only if $A + A^T$ is symmetry positive definite matrix(negative definite matrix).

Let $x = col(x_1, x_2, \dots, x_n)$;

$$R_+^n = \{x \mid x \in R^n, x_i > 0, i = 1 \sim n\}$$

$$f(x) = col(f_1, f_2, \dots, f_n).$$

Manuscript received December 15, 2008. This work was supported by National Natural Science Foundation of China (60674018); and supported by the Natural Sciences Research Foundation of Department of Education of Jiangsu Province in China (08KJD510003).

Feng Liu is with the Department of Basic Courses, Jiangsu Teachers University of Technology, Changzhou 213001, P.R. China (e-mail: liufeng200099@vip.sina.com).

Guodong Shi is with the Jiangsu Teachers University of Technology, Changzhou 213001, P.R. China (e-mail: sgd@jstu.edu.cn).

Zhiqing Weng is with the Changzhou Textile Garment Institute, Changzhou 213164, P.R. China (e-mail: njrwengzq@163.com).

II. MAIN THEOREMS AND PROOF

Consider the general non-linear systems [5,6]

$$\frac{dx_i}{dt} = x_i f_i(x_1, x_2, \dots, x_n) (i = 1 \sim n) \quad (1)$$

where $f_i(x_1, x_2, \dots, x_n)$ continues in $x \geq 0$, and suppose System (1) have positive equilibrium state $x = x^* > 0$, namely, $f_i(x_1^*, x_2^*, \dots, x_n^*) = 0 (i = 1 \sim n)$.

Theorem 1 As for system(1), if $(x - x^*)^T f(x)$ is subnegative definite in R_+^n , then the positive equilibrium state $x = x^* > 0$ of System(1) is of overall stabilization in R_+^n .

Proof As for system (1), make transform

$$y_i = \ln \frac{x_i}{x_i^*} (i = 1, 2, \dots, n),$$

then it can map $x > 0$ into R^n space, and with this transform, System (1) can be rewritten as

$$\frac{dy_i}{dt} = f_i(x_1^* e^{y_1}, x_2^* e^{y_2}, \dots, x_n^* e^{y_n}) (i = 1, 2, \dots, n) \quad (2)$$

and then overall stability of the positive equilibrium state $x = x^* > 0$ of system(1) in R_+^n is equivalent to overall stability of the ordinary solution of system(2) in R^n .

As for system(2), make

$$V(y) = \sum_{i=1}^n x_i^* (e^{y_i} - y_i - 1),$$

then we have

$$\begin{aligned} \left. \frac{dV(y)}{dt} \right|_{(2)} &= \sum_{i=1}^n x_i^* (e^{y_i} - 1) f_i^*(y) \\ &= \sum_{i=1}^n (x_i - x_i^*) f_i(x) \\ &= (x - x^*)^T f(x) < 0 \quad (\text{if } x \neq x^*) \end{aligned}$$

so that the ordinary solution of System(2) in R^n is of overall stabilization, thus, the positive equilibrium state $x = x^* > 0$ of System (1) is of overall stabilization in R_+^n .

Theorem 2 As for system(1), if there exists subnegative definite matrix $G = (g_{ij})_{n \times n}$ in R_+^n , such that

$$\frac{\partial f_i(x)}{\partial x_i} \leq g_{ii} = \text{const},$$

$$\left| \frac{\partial f_i(x)}{\partial x_j} \right| \leq g_{ij} = \text{const} \quad (i \neq j, i, j = 1 \sim n),$$

then the positive equilibrium state $x = x^* > 0$ of System (1) is of overall stabilization in R_+^n .

Proof Because

$$f(x^*) = 0, \frac{d}{d\lambda} f(\lambda x) = \left(\frac{\partial}{\partial x} f(\lambda x)(x - x^*) \right),$$

where $x = \text{col}(x_1, x_2, \dots, x_n)$, $f(x) = \text{col}(f_1, f_2, \dots, f_n)$, thereby

$$f(x) = \left(\int_0^1 \frac{\partial}{\partial x} f(\lambda x) d\lambda \right) (x - x^*) = \left(\frac{\partial}{\partial x} f(\tilde{x}) \right) (x - x^*),$$

As for System(2), make $V(y) = \sum_{i=1}^n x_i^* (e^{y_i} - y_i - 1)$, since

$$\begin{aligned} \left. \frac{dV(y)}{dt} \right|_{(2)} &= (x - x^*)^T f(x) \\ &= (x - x^*)^T \frac{\partial}{\partial x} f(\tilde{x})(x - x^*) \\ &= \sum_{i=1}^n (x_i - x_i^*) \frac{\partial f_i(\tilde{x})}{\partial x_i} (x_i - x_i^*) \\ &\quad + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n (x_i - x_i^*) \frac{\partial f_i(\tilde{x})}{\partial x_j} (x_j - x_j^*) \\ &\leq \sum_{i=1}^n (x_i - x_i^*)^2 g_{ii} + \\ &\quad \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n |x_i - x_i^*| \cdot |g_{ij}| \cdot |x_j - x_j^*| \\ &= (x - x^*)^T G (x - x^*) < 0 \quad (\text{if } x \neq x^*) \end{aligned}$$

so that the ordinary solution of System (2) in R^n is of overall stabilization, thereby the positive equilibrium state $x = x^* > 0$ of System (1) is of overall stabilization in R_+^n .

Theorem 3 As for System(1), if there exists matrix

$G = (g_{ij})_{n \times n}$ in R_+^n such that

$$\frac{\partial f_i(x)}{\partial x_i} \leq g_{ii} = \text{const},$$

$$\left| \frac{\partial f_i(x)}{\partial x_j} \right| \leq g_{ij} = \text{const} \quad (i \neq j, i, j = 1, 2, \dots, n),$$

moreover 1) $g_{ii} < 0 (i = 1, 2, \dots, n)$; 2) There exists constants $\xi_i^{(j)} > 0 (i \neq j, i, j = 1, 2, \dots, n)$, such that

$$\sum_{\substack{j=1 \\ j \neq i}}^n \xi_i^{(j)} \leq 1$$

and

$$\begin{vmatrix} 2\xi_i^{(j)} g_{ii} & g_{ij} + g_{ji} \\ g_{ij} + g_{ji} & 2\xi_j^{(i)} g_{jj} \end{vmatrix} > 0, (i \neq j, i, j = 1, 2, \dots, n).$$

Then the positive equilibrium state $x = x^* > 0$ of System (1) is of overall stabilization in R_+^n .

Proof Since

$$\begin{vmatrix} 2\xi_i^{(j)} g_{ii} & g_{ij} + g_{ji} \\ g_{ij} + g_{ji} & 2\xi_j^{(i)} g_{jj} \end{vmatrix} > 0,$$

therefore there exists $0 < \varepsilon \ll 1$ such that

$$\begin{aligned} 2(G + G^T) \leq & \begin{vmatrix} 2\xi_i^{(j)} g_{ii} + \varepsilon & g_{ij} + g_{ji} \\ g_{ij} + g_{ji} & 2\xi_j^{(i)} g_{jj} + \varepsilon \end{vmatrix} > 0, \\ & \begin{bmatrix} 2\xi_1^{(2)} g_{11} + \varepsilon & g_{12} + g_{21} & 0 & \cdots & 0 \\ g_{12} + g_{21} & 2\xi_2^{(1)} g_{22} + \varepsilon & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots \\ & + \begin{bmatrix} 2\xi_1^{(n)} g_{11} + \varepsilon & 0 & \cdots & 0 & g_{1n} + g_{n1} \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{1n} + g_{n1} & 0 & \cdots & 0 & 2\xi_n^{(1)} g_{nn} + \varepsilon \end{bmatrix} \\ & + \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 2\xi_2^{(3)} g_{22} + \varepsilon & g_{23} + g_{32} & \cdots & 0 \\ 0 & g_{23} + g_{32} & 2\xi_3^{(2)} g_{33} + \varepsilon & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \\ & + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 2\xi_2^{(n)} g_{22} + \varepsilon & \cdots & g_{2n} + g_{n2} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & g_{2n} + g_{n2} & \cdots & 2\xi_n^{(2)} g_{nn} + \varepsilon \end{bmatrix} \\ & + \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 2\xi_{n-1}^{(n)} g_{n-1, n-1} + \varepsilon & g_{n-1, n} + g_{n, n-1} \\ 0 & \cdots & g_{n-1, n} + g_{n, n-1} & 2\xi_n^{(n-1)} g_{nn} + \varepsilon \end{bmatrix} \\ & - \text{diag}((n-1)\varepsilon, (n-1)\varepsilon, \dots, (n-1)\varepsilon). \end{aligned}$$

Except the last item

$$(-\text{diag}((n-1)\varepsilon, (n-1)\varepsilon, \dots, (n-1)\varepsilon))$$

at the rightside of this inequations is negative definite, others

are half negative definite ($0 < \varepsilon \ll 1$), but their sum is negative definite. Therefore, it can be deduced that $G + G^T$ is negative definite, so that $G = (g_{ij})_{n \times n}$ is subnegative definite matrix, from theorem 2, the positive equilibrium state $x = x^* > 0$ of System (1) is of overall stabilization in R_+^n .

Theorem 4 As for System(1), if there exists matrix $G = (g_{ij})_{n \times n}$ in R_+^n such that

$$\frac{\partial f_i(x)}{\partial x_i} \leq g_{ii} = \text{const},$$

$$\left| \frac{\partial f_i(x)}{\partial x_j} \right| \leq g_{ij} = \text{const} \quad (i \neq j, i, j = 1, 2, \dots, n)$$

and 1) $g_{ii} < 0 (i = 1, 2, \dots, n)$;

2) $\frac{2}{n-1} \sqrt{g_{ii} g_{jj}} > |g_{ij} + g_{ji}|, i, j = 1, 2, \dots, n, j \neq i$,

then the positive equilibrium state $x = x^* > 0$ of System (1) is of overall stabilization in R_+^n .

Proof According to the condition 2), to the arbitrary positive number α, β , there always exists:

$$\begin{aligned} (-\alpha g_{ii} - \beta g_{jj}) / (n-1) &\geq 2\sqrt{\alpha\beta g_{ii} g_{jj}} / (n-1) \\ &> |g_{ij} + g_{ji}| \cdot \sqrt{\alpha\beta} \end{aligned}$$

Thereby

$$(\alpha g_{ii} + \beta g_{jj}) / (n-1) < -|g_{ij} + g_{ji}| \cdot \sqrt{\alpha\beta} \quad (3)$$

So that for the arbitrary $x = \text{col}(x_1, x_2, \dots, x_n) \neq 0$, there exists the following two complexions:

(a) There is only one $x_{i_0} \neq 0$ in $x_i (i = 1, 2, \dots, n)$, the rest is $x_i = 0 (i \neq i_0)$, and then

$$x^T G x = g_{i_0 i_0} x_{i_0}^2 < 0.$$

(b) There are two or more $x_i \neq 0$ in $x_i (i = 1 \sim n)$, we might as well suppose there exists a set of $x_i \neq 0, x_j \neq 0$, according to (3), there exists:

$$\sum_{i < j} (g_{ii} x_i^2 + g_{jj} x_j^2) / (n-1) < -\sum_{i < j} |g_{ij} + g_{ji}| \cdot |x_i x_j|$$

thereby

$$\sum_i g_{ii} x_i^2 < -\sum_{i < j} |g_{ij} + g_{ji}| \cdot |x_i x_j| \leq -\sum_{i < j} (g_{ij} + g_{ji}) x_i x_j,$$

namely

$$x^T G x = \sum_{i,j} g_{ij} x_i x_j < 0.$$

Sum up (a) and (b) above, it can be deduced that $G = (g_{ij})_{n \times n}$ is subnegative definite matrix, from Theorem 2, it can be concluded that the positive equilibrium state $x = x^* > 0$ of System(1) is of overall stabilization in R_+^n .

we consider three classes of non-linear ecosystems as follows [4,5,6]:

$$\frac{dx_i}{dt} = x_i f_i(x_1, x_2, \dots, x_n) (i = 1, 2, \dots, n) \quad (4)$$

where $f_i(x_1, x_2, \dots, x_n)$ continues in $x \geq 0$, and suppose System (4) have positive equilibrium state $x = x^* > 0$, namely, $f_i(x_1^*, x_2^*, \dots, x_n^*) = 0 (i = 1, 2, \dots, n)$;

$$\frac{dx_i}{dt} = r_i x_i \left[1 - \left(\frac{x_i}{k_i} \right)^{\theta_i} - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \left(\frac{x_j}{k_j} \right) \right] \quad (i = 1, 2, \dots, n) \quad (5)$$

where $r_i > 0, k_i > 0 (i = 1, 2, \dots, n)$;

$$\frac{dx_i}{dt} = r_i x_i \left[1 - \sum_{j=1}^n e_{ij} \left(\frac{x_j}{k_j} \right)^{\theta_j} \right] \quad (i = 1, 2, \dots, n) \quad (6)$$

where $r_i > 0, \theta_i > 0, e_{ii} = 1, e_{ij}$ is constant and $e_{ij} \geq 0 (i, j = 1, 2, \dots, n, i \neq j)$.

Theorem 5 As for System(4), if there exists matrix $G = (g_{ij})_{n \times n}$ in R_+^n such that

$$\frac{\partial f_i(x)}{\partial x_i} \leq g_{ii} = \text{const},$$

$$\left| \frac{\partial f_i(x)}{\partial x_j} \right| \leq g_{ij} = \text{const} \quad (i \neq j, i, j = 1, 2, \dots, n)$$

and 1) $g_{ii} < 0 (i = 1, 2, \dots, n)$; 2) for arbitrary $i \neq j$,

$$4g_{ii} g_{jj} > (\Lambda_i + \Lambda'_i)(\Lambda_j + \Lambda'_j),$$

where $\Lambda_i = \sum_{j \neq i} |g_{ij}|, \Lambda'_i = \sum_{j \neq i} |g_{ji}|$. Then the positive

equilibrium state $x = x^* > 0$ of System (4) is of overall stabilization in R_+^n .

Proof Let

$$B = (G + G^T) / 2 = (b_{ij})_{n \times n},$$

from condition 2), we have

$$b_{ii} b_{jj} > \tilde{\Lambda}_i \tilde{\Lambda}_j,$$

where

$$\tilde{\Lambda}_i = \sum_{j \neq i} |b_{ij}|.$$

(a) If $|b_{ii}| > \tilde{\Lambda}_i, i = 1, 2, \dots, n$, namely B is diagonally dominant symmetrical matrix and $b_{ii} < 0 (i = 1, 2, \dots, n)$, thereby B is negative definite matrix, so that G is subnegative definite matrix.

(b) If as for some $k, |b_{kk}| \leq \tilde{\Lambda}_k$, then as for arbitrary $i \neq k$, there exists d such that

$$|b_{ii}| / \tilde{\Lambda}_i > d > \tilde{\Lambda}_k / |b_{kk}| \geq 1.$$

Make

$$D = \text{diag}(d_i \mid d_k = d; d_i = 1, i \neq k),$$

$$C = BD = (c_{ij})_{n \times n}$$

then

$$\begin{aligned} |c_{kk}| &= d |b_{kk}| > \tilde{\Lambda}_k = \sum_{j \neq k} |c_{kj}|, |c_{ii}| \\ &= |b_{ii}| > d \tilde{\Lambda}_i \geq d |b_{ik}| + \sum_{j \neq k} |b_{ij}| \\ &= \sum_{j \neq i} |c_{ij}| \end{aligned}$$

namely, matrix C is strict diagonally dominant, thereby matrix B is generalized strict diagonally dominant and $b_{ii} < 0 (i=1,2,\dots,n)$, so that G is subnegative definite matrix..

Sum up (a) and (b) above, it can be deduced that $G = (g_{ij})_{n \times n}$ is subnegative definite matrix.

As for system (4), make transform

$$y_i = \ln \frac{x_i}{x_i^*} \quad (i=1,2,\dots,n),$$

then it can map $x > 0$ into R^n space, and with this transform, System (4) can be rewritten as

$$\frac{dy_i}{dt} = f_i(x_1^* e^{y_1}, x_2^* e^{y_2}, \dots, x_n^* e^{y_n}) \quad (i=1,2,\dots,n) \quad (7)$$

and then overall stability of the positive equilibrium state $x = x^* > 0$ of system(1) in R_+^n is equivalent to overall stability of the ordinary solution of system(7) in R^n .

As for System(7), make

$$V(y) = \sum_{i=1}^n x_i^* (e^{y_i} - y_i - 1),$$

since

$$\begin{aligned} \left. \frac{dV(y)}{dt} \right|_{(7)} &= (x - x^*)^T f(x) \\ &= (x - x^*)^T \frac{\partial}{\partial x} f(\tilde{x})(x - x^*) \\ &= \sum_{i=1}^n (x_i - x_i^*) \frac{\partial f_i(\tilde{x})}{\partial x_i} (x_i - x_i^*) \\ &\quad + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n (x_i - x_i^*) \frac{\partial f_i(\tilde{x})}{\partial x_j} (x_j - x_j^*) \\ &\leq \sum_{i=1}^n (x_i - x_i^*)^2 g_{ii} \\ &\quad + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n |x_i - x_i^*| \cdot |g_{ij}| \cdot |x_j - x_j^*| \\ &= (x - x^*)^T G(x - x^*) < 0 \quad (\text{if } x \neq x^*) \end{aligned}$$

so that the ordinary solution of System (7) in R^n is of overall stabilization, thereby the positive equilibrium state $x = x^* > 0$ of System (4) is of overall stabilization in R_+^n .

Theorem 6 Suppose System (5) has positive equilibrium state $x = x^* > 0$, let

$$\beta_{ii} = \left(\frac{x_i^*}{k_i} \right)^{\theta_i - 1}, \quad \beta_{ij} = a_{ij} \quad (i, j = 1, 2, \dots, n, j \neq i),$$

if $\theta_i \geq 1 (i=1,2,\dots,n)$, and $B = (-\beta_{ij})_{n \times n}$ is subnegative definite matrix, then the positive equilibrium state $x = x^* > 0$ of system (5) is of overall stabilization in R_+^n .

Proof Since $x = x^*$ is equilibrium state of system(5), so that

$$\left(\frac{x_i^*}{k_i} \right)^{\theta_i} + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \left(\frac{x_j^*}{k_j} \right) = 1 \quad (i=1,2,\dots,n) \quad (8)$$

From (5) and (8), we obtain

$$\frac{dx_i}{dt} = r_i x_i \left[\left(\frac{x_i}{k_i} \right)^{\theta_i} - \left(\frac{x_i}{k_i} \right)^{\theta_i} - \sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij}}{k_j} (x_j - x_j^*) \right]. \quad (9)$$

Make Lyapunov function

$$V(x) = \sum_{i=1}^n \frac{1}{r_i} [x_i - x_i^* - x_i^* \ln \left(\frac{x_i}{x_i^*} \right)].$$

If $x_i \geq x_i^*$, from $\theta_i \geq 1$, it can be known that

$$\begin{aligned} \left(\frac{x_i}{x_i^*} \right)^{\theta_i} &\geq \left(\frac{x_i}{x_i^*} \right), \text{ thereby} \\ (x_i - x_i^*) (x_i^{\theta_i} - x_i^{*\theta_i}) &= (x_i - x_i^*) x_i^{*\theta_i} \left(\left(\frac{x_i}{x_i^*} \right)^{\theta_i} - 1 \right) \\ &\geq (x_i - x_i^*) x_i^{*\theta_i} \left(\frac{x_i}{x_i^*} - 1 \right) = x_i^{*\theta_i - 1} (x_i - x_i^*)^2; \end{aligned}$$

If $x_i \leq x_i^*$, similarly, it can be proven

$$(x_i - x_i^*) (x_i^{\theta_i} - x_i^{*\theta_i}) \geq x_i^{*\theta_i - 1} (x_i - x_i^*)^2.$$

Then

$$\begin{aligned} \left. \frac{dV(x)}{dt} \right|_{(6)} &= - \sum_{i=1}^n \frac{1}{k_i^{\theta_i}} (x_i - x_i^*) (x_i^{\theta_i} - x_i^{*\theta_i}) \\ &\quad - \sum_{\substack{i,j=1 \\ j \neq i}}^n \frac{a_{ij}}{k_j} (x_i - x_i^*) (x_j - x_j^*) \\ &\leq - \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} (x_i - x_i^*) (x_j - x_j^*) \\ &= (x - x^*)^T B(x - x^*) < 0 \quad (\text{if } x \neq x^*) \end{aligned}$$

so that the positive equilibrium state $x = x^* > 0$ of System (5) is of overall stabilization in R_+^n .

Theorem 7 Suppose System (6) has positive equilibrium state $x = x^* > 0$, if $E = (e_{ij})_{n \times n}$ is subpositive definite matrix, then the positive equilibrium state $x = x^* > 0$ of System(6) is of overall stabilization in R_+^n .

Proof Let

$$y_i = \frac{x_i}{k_i} \quad (i = 1 \sim n),$$

then (6) can be rewritten as

$$\frac{dy_i}{dt} = r_i y_i \left[1 - \sum_{j=1}^n e_{ij} y_j^{\theta_j} \right] \quad (10)$$

from $\sum_{j=1}^n e_{ij} y_j^{*\theta_j} = 1$, and make

$$V(y) = \sum_{i=1}^n \frac{1}{r_i} \left[\left(\frac{y_i^{\theta_i} - y_i^{*\theta_i}}{\theta_i} \right) - y_i^{*\theta_i} \ln \left(\frac{y_i}{y_i^*} \right) \right],$$

then

$$\left. \frac{dV(y)}{dt} \right|_{(10)} = -(z - z^*)^T E(z - z^*) < 0 \quad (\text{if } z \neq z^*)$$

where $z_i = y_i^{\theta_i}$. Therefore, the positive equilibrium state

$$z = z^* = (y_1^{*\theta_1}, y_2^{*\theta_2}, \dots, y_n^{*\theta_n})$$

of System (10) is of overall stabilization in R_+^n . Thus, the positive equilibrium state $x = x^* > 0$ of System(6) is of overall stabilization in R_+^n .

III. CONCLUSION

In this paper, subpositive definite matrices are led into general non-linear ecosystems, and some new criterions for general non-linear ecosystems on overall stability based on subpositive definite matrices are acquired. These results improve or expand some existing criterions.

REFERENCES

- [1] Li Jibin & Chen Lansun, *Biology and Mathematics*, Sichuan Education Publishing Company, Cheng Dou, 1986.
- [2] Araki M. & Kodno B, Stability and transient behavior of composite non-linear systems, *IEEE Trans.Autom.Control*, 17(1972),537-541.
- [3] Goh,B.S, Sector stability of a complex ecosystem model, *Math Biosci*, 40(1978),157-166.
- [4] Liao Xiaoxin & Jia Li, Stability in Gilpin-Ayala competition.models with diffusion, *Nonlinear Analysis.Theory,Methods & Appl*, 20(10)1997,1751-1758.
- [5] Liao Xiaoxin, Stability of general ecological systems and neural networks systems. *Proc of The First World Congress of Nonlinear Analyst*, 1996, 1325-1340.
- [6] Kang Qingkai & Liao Xiaoxin, Dissipation, boundedness and persistence of general ecological systems,*Nonlinear Analysis Theory Method & Appl*, 25(11)1995, 1237-1250.
- [7] Zhao Suxia, On absolute stability of direct regulated system, *Acta Mathematica Scientia*, 22(4)1979,404-419.
- [8] Zhang Binggen, Stability of ecological systems with over one step food cycle, *Acta Mathematica Applicatae Scientia*, 6(2)1983,236-239.
- [9] Tu Boxun, The theory of subpositive definite matrices(I), *Acta Mathematica Scientia*, 33(4)1990,462-471.
- [10] Goh,B.S, Sector stability of a complex ecosystem model, *Math Biosci*, 40(1978),157-166.
- [11] Liu Feng, Some applications of subpositive definite matrices for stability of Gilpin-Ayala systems, *Mathematics in Practice and Theory*, 35(8)2005.,123-126
- [12] Ge Zhaoqiang, Liu Feng, Zhao G.uangshe, Pole assignment for a class of first order coupled generalized systems, *Chinese Journal of Engineering Mathematic*, 22(1) 2005,147—150.

- [13] Ge Zhaoqiang and Ma Yonghao, Pole assignment for first order parameter distributed generalized systems,*Chinese Annals of Mathematics, Series A*, 22(6)2001,729--734.
- [14] F.Liu, J.S.Wang. Some New Criterions for Stability of Hopfield Type Neural Network Systems Based on Subpositive Definite Matrices.*Proc. of the 6th World Congress on Intelligent Control and Automation*, 14(2006), 2691-2694.
- [15] F.Liu, Z.Q.Ge.Some New Criterions for Stability of Volterra Ecosystems Based on Subpositive Definite Matrix.*Pure and Applied Mathematics*, 22(2006),198-203.
- [16] F.liu.Z Q Ge,J S Wang.Some New Criterions for Stability of Generalized Cohen-Grossberg Systems Based on Subpositive Definite Matrix. *Dynamics of Continuous, Discrete and Impulsive Systems SeriesA: mathematics Analysis 14* (2007) ,569-572

Chaotic Pseudo Random Number Generators via Ultra Weak Coupling of Chaotic Maps and Double Threshold Sampling Sequences.

René Lozi

Abstract—Generation of random or pseudorandom numbers, nowadays, is a key feature of industrial mathematics. Pseudorandom or chaotic numbers are used in many areas of contemporary technology such as modern communication systems and engineering applications. More and more European or US patents using discrete mappings for this purpose are obtained by researchers of discrete dynamical systems [1], [2]. Efficient Chaotic Pseudo Random Number Generators (CPRNG) have been recently introduced. They use the ultra weak multidimensional coupling of p I -dimensional dynamical systems which preserve the chaotic properties of the continuous models in numerical experiments. Together with chaotic sampling and mixing processes, ultra weak coupling leads to families of (CPRNG) which are noteworthy [3], [4].

In this paper we improve again these families using a double threshold chaotic sampling instead of a single one.

We analyze numerically the properties of these new families and underline their very high qualities and usefulness as CPRNG when very long series are computed.

Index Terms—Chaos, Discrete time systems, Floating point arithmetic, Random number generation.

I. INTRODUCTION

Efficient Chaotic Pseudo Random Number Generators (CPRNG) have been recently introduced. The idea of applying discrete chaotic dynamical systems, intrinsically, exploits the property of extreme sensitivity of trajectories to small changes of initial conditions. They use the ultra weak multidimensional coupling of p I -dimensional dynamical systems which preserve the chaotic properties of the continuous models in numerical experiments. The process of chaotic sampling and mixing of chaotic sequences, which is pivotal for these families, works perfectly in numerical simulation when floating point (or double precision) numbers are handled by a computer.

It is noteworthy that these families of very weakly coupled maps are more powerful than the usual formulas used to generate chaotic sequences mainly because only additions and

multiplications are used in the computation process; no division being required. Moreover the computations are done using floating point or double precision numbers, allowing the use of the powerful Floating Point Unit (FPU) of the modern microprocessors (built by both Intel and Advanced Micro Devices (AMD)). In addition, a large part of the computations can be parallelized taking advantage of the multicore microprocessors which appear on the market of laptop computers.

In this paper we improve the properties of these families using a double threshold chaotic sampling instead of a single one. The genuine map f used as one-dimensional dynamical systems to generate them is henceforth perfectly hidden.

II. ULTRA WEAK MULTIDIMENSIONAL COUPLING

A. System of p -Coupled Symmetric Tent Map

When a dynamical system is realized on a computer using floating point or double precision numbers, the computation is of a discretization, where finite machine arithmetic replaces continuum state space. For chaotic dynamical systems, the discretization often has collapsing effects to a fixed point or to short cycles [5], [6]. In order to preserve the chaotic properties of the continuous models in numerical experiments we have recently introduced an ultra weak multidimensional coupling of p one-dimensional dynamical systems which is noteworthy [7].

In this specific case we have chosen as an example the symmetric tent map defined by

$$f_a(x) = 1 - a|x| \quad (1)$$

with the value $a = 2$, later denoted simply as f , even though others chaotic map of the interval (as the logistic map) can be used for the same purpose. The dynamical system associated to this one dimensional map is defined by the equation on the interval $\mathbf{J} = [-1, 1] \subset \mathbb{R}$ [8].

$$x_{n+1} = 1 - a|x_n| \quad (2)$$

The system of p -coupled dynamical systems is then:

$$X_{n+1} = F(X_n) = A \cdot \underline{f}(X_n) \quad (3)$$

with

R. Lozi is with the Laboratory J. A. Dieudonné, UMR CNRS 6621, University of Nice Sophia-Antipolis, 06108 Nice Cedex 02, France and the Institut Universitaire de Formation des Maîtres Célestin Freinet-académie de Nice, University of Nice-Sophia-Antipolis, 89 avenue George V, 06046 Nice Cedex 1, France (corresponding author to provide phone: 04-93-53-75-08; e-mailrlozi@unice.fr).

$$X = \begin{pmatrix} x^1 \\ \vdots \\ x^p \end{pmatrix}, \quad \underline{f}(X) = \begin{pmatrix} f(x^1) \\ \vdots \\ f(x^p) \end{pmatrix} \quad (4)$$

and

$$A = \begin{pmatrix} 1-(p-1)\varepsilon_1 & \varepsilon_1 & \cdots & \varepsilon_1 & \varepsilon_1 \\ \varepsilon_2 & 1-(p-1)\varepsilon_2 & \cdots & \varepsilon_2 & \varepsilon_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \varepsilon_p & \cdots & \cdots & \varepsilon_p & 1-(p-1)\varepsilon_p \end{pmatrix} \quad (5)$$

F is a map of \mathbf{J}^p into itself.

Several combinations can be given for the relative values of the ε_i , in this paper we choose

$$\varepsilon_i = i \varepsilon_1 \quad i = 2, \dots, p \quad (6)$$

The matrix A is always a stochastic matrix iff the coupling constants ε_i verify

$$0 \leq \varepsilon_i \leq \frac{1}{p-1} \quad (7)$$

When $\varepsilon_i = 0$ the maps are decoupled, when $\varepsilon_i = \frac{1}{p-1}$ they are fully cross coupled. Generally, researchers do not consider very small values of ε_i because it seems that the maps are quasi decoupled with those values and no special effect of the coupling is expected. In fact it is not the case and ultra small coupling constant (as small as 10^{-7} for floating point numbers or 10^{-14} for double precision numbers), allows the construction of very long periodic orbits, leading to sterling chaotic generators. Moreover each component of these numbers belonging to \mathbb{R}^p is equally distributed over the finite interval $\mathbf{J} \subset \mathbb{R}$. Numerical computations show that this distribution is obtained with a very good approximation. They have also the property that the length of the periods of the numerically observed orbits is very large [7].

B. Chaotic Pseudo-Random Generators

However chaotic numbers are not pseudo-random numbers because the plot of the couples of iterated points (x_n, x_{n+1}) in the phase plane reveals the map f used as one-dimensional dynamical systems to generate them.

Nevertheless we have recently introduced a family of Enhanced Chaotic Pseudo Random Number Generators (CPRNG) in order to compute very fast long series of pseudorandom numbers with desktop computer [9]. This family is based on the previous ultra weak coupling which is enhanced in order to conceal the chaotic genuine function.

In the aim of hiding f in the phase space (x_n^l, x_{n+1}^l) two mechanisms are used. The pivotal idea of the first one mechanism is to sample chaotically the sequence $(x_0^l, x_1^l, x_2^l, \dots, x_n^l, x_{n+1}^l, \dots)$ generated by the l -th component x^l , selecting x_n^m every time the value x_n^m of the m -th component x^m , is strictly greater than a threshold $T \in \mathbf{J}$, with $l \neq m$, for $1 \leq l, m \leq p$.

A second mechanism can improve the unpredictability of the chaotic sequence generated as above, using synergistically all the components of the vector X , instead of two. This simple mechanism is based on the chaotic mixing of the p -1 sequences $(x_0^1, x_1^1, x_2^1, \dots, x_n^1, x_{n+1}^1, \dots)$, $(x_0^2, x_1^2, x_2^2, \dots, x_n^2, x_{n+1}^2, \dots)$, \dots , $(x_0^{p-1}, x_1^{p-1}, x_2^{p-1}, \dots, x_n^{p-1}, x_{n+1}^{p-1}, \dots)$ using the last one $(x_0^p, x_1^p, x_2^p, \dots, x_n^p, x_{n+1}^p, \dots)$ with respect to a given partition T_1, T_2, \dots, T_{p-1} of \mathbf{J} , to distribute the iterated points.

C. Numerical Results

As an example we explicit both mechanisms taking 4-coupled equations for (3). The value of x_n^4 commands the chaotic sampling and the mixing processes as follows.

$$\text{Let us set three threshold values } T_1, T_2 \text{ and } T_3 \\ -1 < T_1 < T_2 < T_3 < 1 \quad (8)$$

we sample and mix together chaotically the sequences $(x_0^1, x_1^1, x_2^1, \dots, x_n^1, x_{n+1}^1, \dots)$, $(x_0^2, x_1^2, x_2^2, \dots, x_n^2, x_{n+1}^2, \dots)$ and $(x_0^3, x_1^3, x_2^3, \dots, x_n^3, x_{n+1}^3, \dots)$ defining $(\overline{x_0}, \overline{x_1}, \overline{x_2}, \dots, \overline{x_q}, \overline{x_{q+1}}, \dots)$ by

$$\overline{x_q} = \begin{cases} x_n^1 & \text{iff } x_n^4 \in]T_1, T_2[\\ x_n^2 & \text{iff } x_n^4 \in [T_2, T_3[\\ x_n^3 & \text{iff } x_n^4 \in [T_3, 1[\end{cases} \quad (9)$$

Numerical results about chaotic numbers produced by (3) - (9) show that they are equally distributed over the interval \mathbf{J} .

In order to compute numerically an approximation of the invariant measure also called the probability distribution function $P_N(x)$ linked to the I -dimensional map f we consider a regular partition of M small intervals (boxes) r_i of \mathbf{J} .

$$\mathbf{J} = \bigcup_0^{M-1} r_i \quad (10)$$

$$r_i = [s_i, s_{i+1}[\quad , i = 0, M-2 \quad (11)$$

$$r_{M-1} = [s_{M-1}, I] \quad (12)$$

$$s_i = -1 + \frac{2i}{M} \quad i = 0, M \quad (13)$$

the length of each box is

$$s_{i+1} - s_i = \frac{2}{M} \quad (14)$$

All iterates $f^{(n)}(x)$ belonging to these boxes are collected (after a transient regime of Q iterations decided *a priori*, i.e. the first Q iterates are neglected). Once the computation of $N+Q$ iterates is completed, the relative number of iterates with respect to N/M in each box r_i represents the value $P_N(s_i)$. The approximated $P_N(x)$ defined in this article is then a step function, with M steps. As M may vary, we define

$$P_{M,N}(s_i) = \frac{1}{2} \frac{M}{N} (\#r_i) \quad (15)$$

where $\#r_i$ is the number of iterates belonging to the interval r_i and the constant $1/2$ allows the normalisation of $P_{M,N}(x)$ on the interval \mathbf{J} .

$$P_{M,N}(x) = P_{M,N}(s_i) \quad \forall x \in r_i \quad (16)$$

In the case of coupled maps, we are more interested by the distribution of each component x^1, \dots, x^p of X rather than the distribution of the variable X itself in \mathbf{J}^p . We then consider the approximated probability distribution function $P_N(x^j)$ associated to one among several components of $F(X)$ defined by (3) which are one-dimensional maps.

The discrepancies E_1 (in norm L_1) and E_2 (in norm L_2) between $P_{N_{disc}, N_{iter}}(x)$ and the Lebesgue measure which is the invariant measure associated to the symmetric tent map, are defined by

$$E_1(N_{disc}, N_{iter}) = \left\| P_{N_{disc}, N_{iter}}(x) - 0.5 \right\|_{L_1} \quad (17)$$

$$E_2(N_{disc}, N_{iter}) = \left\| P_{N_{disc}, N_{iter}}(x) - 0.5 \right\|_{L_2} \quad (18)$$

In the same way an approximation of the correlation distribution function $C_N(x, y)$ is to obtained numerically building a regular partition of M^2 small squares (boxes) of \mathbf{J}^2 imbedded in the phase subspace (x^j, x^m)

$$r_{i,j} = [s_i, s_{i+1}] \times [t_j, t_{j+1}] \quad , \quad i, j = 0, M-2 \quad (19)$$

$$r_{M-1,j} = [s_{M-1}, I] \times [t_j, t_{j+1}] \quad , \quad j = 0, M-2 \quad (20)$$

$$r_{i,M-1} = [s_i, s_{i+1}] \times [t_{M-1}, I] \quad , \quad i = 0, M-2 \quad (21)$$

$$r_{M-1,M-1} = [s_{M-1}, I] \times [t_{M-1}, I] \quad (22)$$

$$s_i = -1 + \frac{2i}{M}, t_j = -1 + \frac{2j}{M}, i, j = 0, M \quad (23)$$

the measure of the area of each box is

$$(s_{i+1} - s_i) \cdot (t_{j+1} - t_j) = \left(\frac{2}{M}\right)^2 \quad (24)$$

Once $N + Q$ iterated points (x_n^j, x_n^m) belonging to these boxes are collected the relative number of iterates with respect to N/M^2 in each box $r_{i,j}$ represents the value $C_N(s_i, t_j)$. The approximated probability distribution function $C_N(x, y)$ defined here is then a 2-dimensional step function, with M^2 steps. As M can take several values in the next sections, we define

$$C_{M,N}(s_i, t_j) = \frac{1}{4} \frac{M^2}{N} (\#r_{i,j}) \quad (25)$$

where $\#r_{i,j}$ is the number of iterates belonging to the square $r_{i,j}$ and the constant $1/4$ allows the normalisation of $C_{M,N}(x, y)$ on the square \mathbf{J}^2 .

$$C_{M,N}(x, y) = C_{M,N}(s_i, t_j) \quad \forall (x, y) \in r_{i,j} \quad (26)$$

The discrepancies E_{C_1} in norm L_1 between $C_{N_{disc}, N_{iter}}(x, y)$ and the uniform distribution on the square is defined by

$$E_{C_1}(N_{disc}, N_{iter}) = \left\| C_{N_{disc}, N_{iter}}(x, y) - 0.25 \right\|_{L_1} \quad (27)$$

Finally let $AC_{M,N}(x, y)$ be the autocorrelation distribution function which is the correlation function $C_{M,N}(x, y)$ of (26) defined in the phase space (x_n^j, x_{n+1}^j) instead of the phase space (x^j, x^m) . In order to control that the enhanced chaotic numbers $(\overline{x_0}, \overline{x_1}, \overline{x_2}, \dots, \overline{x_q}, \overline{x_{q+1}}, \dots)$ are uncorrelated, we plot them in the phase subspace $(\overline{x_n}, \overline{x_{n+1}})$ and we check if they are uniformly distributed in the square \mathbf{J}^2 and if f is concealed.

Fig. 1 shows the values of $E_{AC_1}(N_{disc}, NSampl_{iter})$ for a system of 4-coupled equations when the three components x^1, x^2, x^3 are mixed and sampled by x^4 for the threshold values $T_1 = 0.98, T_2 = 0.987, T_3 = 0.994$ or $T_1 = 0.998, T_2 = 0.9987, T_3 = 0.9994$.

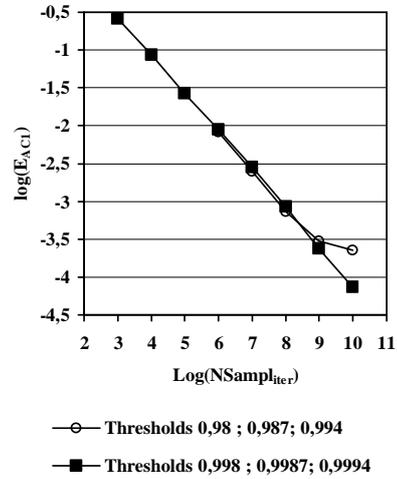


Figure 1. Error of $E_{AC_1}(N_{disc}, NSampl_{iter})$ $N_{disc}=10^2 \times 10^2$, $NSampl_{iter} = 10^3$ to 10^{10} , $\varepsilon_i = i \cdot \varepsilon_1$, $\varepsilon_1 = 10^{-14}$.

N_{iter}	$NSampl_{iter}$	$E_{AC_1}(N_{disc}, NSampl_{iter})$ 4-coupled equation $T_1 = 0.998,$ $T_2 = 0.9987, T_3 = 0.9994$
10^5	93	0.68924731
10^6	1015	0.25881773
10^7	10,139	0.086706776
10^8	100,465	0.026815309
10^9	1,000,549	0.0089111078
10^{10}	9,998,814	0.0027932033
10^{11}	100,001,892	0.00085967214
10^{12}	999,945,728	0.0002346851
10^{13}	10,000,046,137	0.000073234736

Table 1. Error of $E_{AC_1}(N_{disc}, NSampl_{iter})$ for a system of 4 coupled-equations when the three components x^1, x^2, x^3 are mixed and sampled by x^4 for the threshold values $T_1 = 0.998, T_2 = 0.9987, T_3 = 0.9994$.

III. DOUBLE THRESHOLD CHAOTIC SAMPLING

A. Improved CPRNG

One can again improve the CPRG previously introduced with respect to the infinity norm instead of the L_1 or L_2 norms because the L_∞ norm is more sensitive than the others to reveal the concealed f . For this aim, consider first that in the phase space (x'_n, x'_{n+1}) the graph of the chaotically sampled chaotic numbers is a mix of the graphs of all the $f^{(i)}$ (Fig. 2).

It is obvious as showed on Fig. 3 that for $r = 1$ if $M = 1$ or 2 , $AC_{M,N}(x, y)$ is constant and normalized on the square hence $E_{AC_\infty}(M, N) = E_{AC_1}(M, N) = E_{AC_2}(M, N) = 0$.

The autocorrelation function is different from zero only if $M > 2$ (Fig. 4).

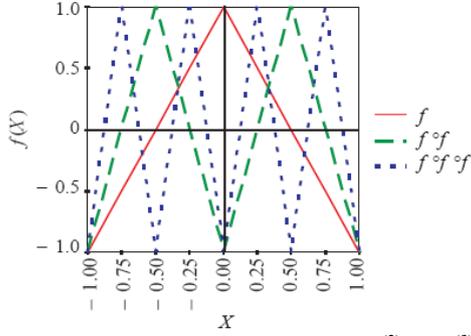


Figure 2. Graphs of the symmetric tent map $f, f^{(2)}$ and $f^{(3)}$ on the interval $[-1, 1]$.

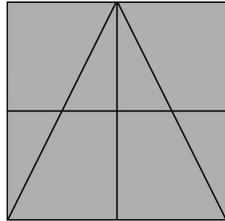


Figure 3. In shaded regions the autocorrelation distribution $AC_{M,N}(x, y)$ is constant for the symmetric tent map f on the interval $[-1, 1]$ for $M = 1$ or 2 .

In the same way as displayed on Fig. 5, 6 and 7, $E_{AC_\infty}(M, N) = E_{AC_1}(M, N) = E_{AC_2}(M, N) = 0$ for $f^{(i)}$ iff $M < 2^i$. Hence for a given M , if we cancel the contribution of all the $f^{(i)}$ for $2^i < M$, it is not possible to identify the genuine function f .

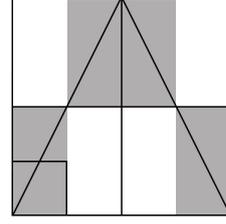


Figure 4. Regions where the autocorrelation distribution $AC_{M,N}(x, y)$ is constant for the symmetric tent map f are shaded, for $M = 4$. (The square on the bottom left of the graph shows the size of the $r_{i,j}$ box). $AC_{M,N}(x, y)$ vanishes on the white regions.

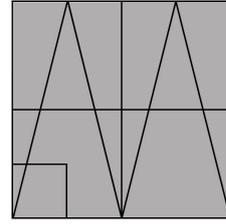


Figure 5. In shaded regions the autocorrelation distribution $AC_{M,N}(x, y)$ is constant for the symmetric tent map $f^{(2)}$ on the interval $[-1, 1]$ for $M = 1, 2$ and 4 .

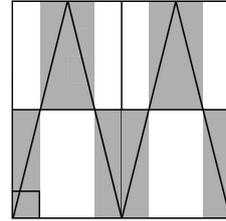


Figure 6. Regions where the autocorrelation distribution $AC_{M,N}(x, y)$ is constant for the symmetric tent map $f^{(2)}$ are shaded for $M = 8$.

B. Algorithm and Numerical Results

We describe again the algorithm of the double threshold chaotic sampling in the case of 4-coupled equations.

Consider the sequence $(\overline{x_0}, \overline{x_1}, \overline{x_2}, \dots, \overline{x_q}, \overline{x_{q+1}}, \dots)$ we want to mix and sample. For each $q-1$ there exists $n_{(q-1)}$ in the original sequence. We introduce a second threshold $N' \in \mathbb{N}$ and then we define:

$$\overline{x_q} = \begin{cases} x_n^1 & \text{iff } x_n^4 \in]T_1, T_2[\text{ and } n - n_{(q-1)} > N' \\ x_n^2 & \text{iff } x_n^4 \in [T_2, T_3[\text{ and } n - n_{(q-1)} > N' \\ x_n^3 & \text{iff } x_n^4 \in [T_3, 1[\text{ and } n - n_{(q-1)} > N' \end{cases} \quad (28)$$

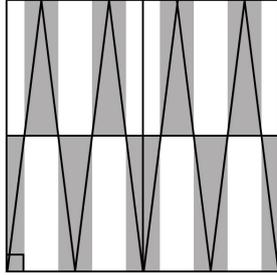


Figure 7. Regions where the autocorrelation distribution $AC_{M,N}(x, y)$ is constant for the symmetric tent map $f^{(3)}$ are shaded for $M = 16$.

As shown previously [9] the errors in L_1 or L_2 norms decrease with the number of chaotic points (as in the law of large numbers) and conversely increase with the number M of boxes used to define $AC_{M,N}(x, y)$. It is the same for the error in L_∞ norm. Fig. 8 shows that when M is greater than 2^5 , the sequence defined by (28) behaves better than the one defined by (9).

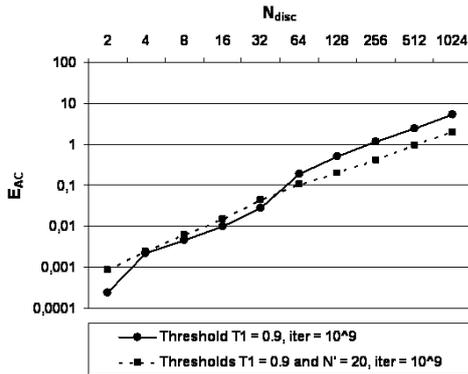


Figure 8. Error of $E_{AC_\infty}(N_{disc}, NSampl_{iter})$ $N_{disc} = 2^1$ to 2^{10} , $NSampl_{iter} = 10^9$, thresholds $T = 0.9$ and $N' = 20$, $\varepsilon_i = i \cdot \varepsilon_1$, $\varepsilon_1 = 10^{-14}$.

Fig. 9 shows that when the number of chaotic points increases the error L_∞ decreases drastically. If $N' > 100$, it is necessary to use a huge grid of $2^{100} \times 2^{100}$ boxes splitting the square \mathcal{J}^2 in order to find a trace of the genuine function f . This is numerically impossible with double precision numbers. Then the chaotic numbers appear as random numbers.

Others numerical results show the high-potency of these new CPRNG. Due to limitation of this article, they will be published elsewhere.

IV. CONCLUSION

Using a double threshold in order to sample a chaotic sequence, we have improved with respect to the infinity norm the CPRNG previously introduced. When the value of the

second threshold N' is greater than 100, it is impossible to find the genuine function used to generate the chaotic numbers. The new CPRNG family is robust versus the choice of the weak parameter of the system for $10^{-14} < \varepsilon < 10^{-5}$, allowing the use of this family in several applications as for example chaotic cryptography.

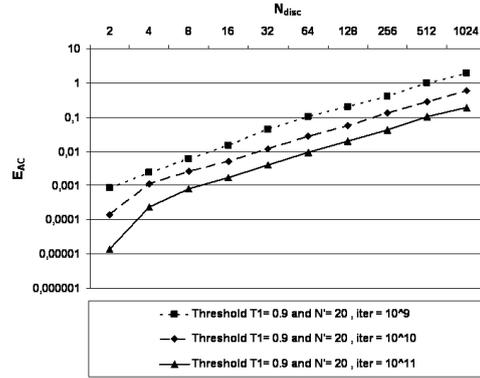


Figure 9. Error of $E_{AC_\infty}(N_{disc}, NSampl_{iter})$ $N_{disc} = 2^1$ to 2^{10} , $NSampl_{iter} = 10^9$ to 10^{11} , thresholds $T = 0.9$ and $N' = 20$, $\varepsilon_i = i \cdot \varepsilon_1$, $\varepsilon_1 = 10^{-14}$.

REFERENCES

- [1] Petersen, M. V., Sorensen, H. M., Method of generating pseudo-random numbers in an electronic device, and a method of encrypting and decrypting electronic data. *United States Patent 7170997*, 2007.
- [2] Ruggiero, D., Mascolo, D., Pedaci, I., Amato, P., Method of generating pseudo-random bits or numbers. *United States Patent Application 20060251250*, 2006.
- [3] S. Hénaff, I. Taralova, R. Lozi, Observers design for a new weakly coupled map function, preprint. <http://hal.archives-ouvertes.fr/hal-00368576/fr/>
- [4] S. Hénaff, I. Taralova, R. Lozi, Dynamical Analysis of a new statistically highly performant deterministic function for chaotic signals generation, preprint. <http://hal.archives-ouvertes.fr/hal-00368844/fr/>
- [5] O. E., Lanford III, Some informal remarks on the orbit structure of discrete approximations to chaotic maps. *Experimental Mathematics*, Vol. 7, 4, 317-324, 1998.
- [6] Gora, P., Boyarsky, A., Islam, Md. S., Bahsoun, W., Absolutely continuous invariant measures that cannot be observed experimentally. *SIAM J. Appl. Dyn. Syst.*, 5:1, 84-90 (electronic), 2006.
- [7] Lozi, R., Giga-Periodic Orbits for Weakly Coupled Tent and Logistic Discretized Maps. International Conference on Industrial and Applied Mathematics, New Delhi, december 2004, *Modern Mathematical Models, Methods and Algorithms for Real World Systems*, A.H. Siddiqi, I.S. Duff and O. Christensen (Editors), Anamaya Publishers, New Delhi, India pp. 80-124, 2006.
- [8] Sprott, J. C., *Chaos and Time-Series Analysis*. Oxford University Press, Oxford, UK, 2003.
- [9] R. Lozi, New Enhanced Chaotic Number Generators, *Indian Journal of Industrial and Applied Mathematics*, vol.1, No.1, pp.1-23, 2008.

Periodic solution for fourth-order discrete Hamiltonian systems

Qiong Meng

Abstract—The existence of periodic solution are obtained for fourth-order discrete Hamiltonian systems

$$\Delta^4 u(n-2) + \nabla F(n, u(n)) = 0, \quad \forall n \in Z$$

by using critical point theory.

Index Terms—fourth-order discrete Hamiltonian systems, periodic solution, critical point theory.

I. INTRODUCTION

Consider the fourth-order discrete Hamiltonian systems

$$\Delta^4 u(n-2) + \nabla F(n, u(n)) = 0, \quad \forall n \in Z, \quad (1)$$

where $\Delta u(n) = u(n+1) - u(n)$, $\Delta^2 u(n) = \Delta(\Delta u(n))$, $F : Z \times R^N \rightarrow R$, $F(n, x)$ is continuously differential in x for every $n \in Z$ and T -periodic in n for all $x \in R^N$, $\nabla F(n, u(n))$ denotes the gradient of $F(n, x)$ in x . We define that T is a positive integer, Z is the set of all integers and $Z[a, b] := Z \cap [a, b]$ for every $a, b \in Z$ with $a \leq b$. we are interesting in the existence of a periodic solution of (1).

In 2003, Guo and Yu [4] firstly studied the existence of periodic solutions of a second-order nonlinear difference equation by using critical point theory. Since that time, there appear many papers about periodic solutions of difference equations by using critical point theory (see [2-4, 6-8]). In this paper, we obtain some sufficient conditions for the existence of periodic solutions of (1) by using critical point theory.

Now we state our main results below.

Theorem 1. Suppose that $F(n, x)$ satisfies

(H₁) There exists a positive integer T , such that $F(t+T, x) = F(n, x)$ for all $(n, x) \in Z \times R^N$;

(H₂) There exist constants $M_1 > 0$, $M_2 > 0$, $0 \leq \alpha < 1$, such that

$$|\nabla F(n, u(n))| \leq M_1 |x|^\alpha + M_2$$

for all $(n, x) \in Z[1, T] \times R^N$;

(H₃) $|x|^{-2\alpha} \sum_{n=1}^T F(n, x) \rightarrow +\infty$ as $|x| \rightarrow +\infty$, for all $n \in Z[1, T]$.

Then (1) possess at least one periodic solution with period T .

Theorem 2. Suppose that $F(n, x)$ satisfies (H₁), (H₂) and

Qiong Meng is with the School of Mathematics Sciences and Computing Technology, Central South University, Changsha, Hunan 410083 and the Department of Mathematics, Shanxi University, Taiyuan, Shanxi 030006, PR China. e-mail: qiongmeng@sohu.com

This project was supported by the Postdoctoral Science Foundation of Central South University.

(H₄) $|x|^{-2\alpha} \sum_{n=1}^T F(n, x) \rightarrow -\infty$ as $|x| \rightarrow \infty$, for all $n \in Z[1, T]$.

Then (1) possess at least one periodic solution with period T .

Theorem 3. Suppose that $F(n, x)$ satisfies (H₁) and

(H₅) $(\nabla F(n, x), x) - 2F(n, x) \rightarrow +\infty$ as $|x| \rightarrow +\infty$ for all $n \in Z[1, T]$;

(H₆) $|x|^{-2} F(n, x) \rightarrow 0$ as $|x| \rightarrow \infty$, for all $n \in Z[1, T]$.

Then (1) possess at least one periodic solution with period T .

Theorem 4. Suppose that $F(n, x)$ satisfies (H₁) and the following:

(H₇) There exist constants $G > 0$, $0 < \beta < 2$ such that for all $(n, x) \in Z[1, T] \times R^N$ and $|x| \geq G$,

$$(\nabla F(n, x), x) \geq \beta F(n, x);$$

(H₈) $F(n, x) \rightarrow -\infty$ as $|x| \rightarrow \infty$ for all $n \in Z[1, T]$.

Then (1) possess at least one periodic solution with period T .

II. PRELIMINARY

First, we shall state some basic notations. For any given positive integer T , H_T is defined by:

$$H_T = \{u : Z \rightarrow R^N | u(n+T) = u(n), n \in Z\}.$$

H_T can be equipped with the inner product

$$\langle u, v \rangle = \sum_{n=1}^T (u(n), v(n)), \quad \forall u, v \in H_T,$$

by which norm $\|\cdot\|_{H_T}$ can be induced by:

$$\|u\| = \left(\sum_{n=1}^T |u(n)|^2 \right)^{\frac{1}{2}}, \quad \forall u \in H_T,$$

where (\cdot, \cdot) and $|\cdot|$ denote the usual inner product and the usual norm in R^N . It is easy to see that $(H_T, \langle \cdot, \cdot \rangle)$ is a finite dimension Hilbert space and linear homeomorphic to R^{NT} . We can equip H_T with another norm $\|u\|_r$ for any positive number $r > 1$, where

$$\|u\|_r = \left(\sum_{n=1}^T |u(n)|^r \right)^{\frac{1}{r}}, \quad \forall u \in H_T,$$

Obviously, $\|u\|_2 = \|u\|_{H_T}$ and $(H_T, \|\cdot\|)$ is equivalent to $(H_T, \|\cdot\|_r)$. Hence, there exist two positive constants C_1, C_2 which are not in relation to r , such that, for all $u \in H_T$,

$$C_1 \|u\|_r \leq \|u\| \leq C_2 \|u\|_r. \quad (2)$$

On the other hand, we define

$\|u\|_\infty = \sup_{n \in Z[1, T]} |u(n)|$, one can get that

$$\|u\|_\infty \leq \|u\|_r \leq T^{\frac{1}{r}} \|u\|_\infty \leq T \|u\|_\infty. \quad (3)$$

Next, we will establish the variational framework of (1) and transferring the existence of periodic solutions of (1) into the existence of critical point of some functional. For any $u, v \in H_T$, By using the methods in [7], we have a useful equality

$$\begin{aligned} & \sum_{n=1}^T (\Delta^4 u(n-2), v(n)) \\ &= \sum_{n=1}^T (\Delta^2 u(n-1), \Delta^2 v(n-1)). \end{aligned} \quad (4)$$

Now we define the functional φ defined on H_T by

$$\varphi(u) = \frac{1}{2} \sum_{n=1}^T |\Delta^2 u(n-1)|^2 + \sum_{n=1}^T F(n, u(n)). \quad (5)$$

Since $F(n, x)$ is continuously differential in x , $\varphi \in C'(H_T, \mathbb{R})$. For any $v \in H_T$, one has

$$\begin{aligned} \langle \varphi'(u), v \rangle &= \sum_{n=1}^T (\Delta^2 u(n-1), \Delta^2 v(n-1)) \\ &+ \sum_{n=1}^T (\nabla F(n, u(n)), v(n)). \end{aligned} \quad (6)$$

Then $u \in H_T$ is a critical point of φ if and only if

$$\begin{aligned} & \sum_{n=1}^T (\Delta^2 u(n-1), \Delta^2 v(n-1)) + \sum_{n=1}^T (\nabla F(n, u(n)), v(n)) \\ &= 0. \end{aligned} \quad (7)$$

It follows from (4) and (7) that

$$\sum_{n=1}^T (\Delta^4 u(n-2), v(n)) + \sum_{n=1}^T (\nabla F(n, u(n)), v(n)) = 0.$$

By the arbitrary of v , we conclude that

$$\Delta^4 u(n-2) + \nabla F(n, u(n)) = 0, \quad \forall n \in Z.$$

Since $u \in H_T$ is T -periodic and $F(n, x)$ is T -periodic in n , hence $u \in H_T$ is a critical point of φ if and only if for any $n \in Z$, $\Delta^4 u(n-2) + \nabla F(n, u(n)) = 0$. Thus the problem of finding the T -periodic solution for problem (1) is reducing to the one of seeking the critical point of functional φ on H_T .

In the final, we give lemmas which discuss the properties of finite dimensional space H_T by the operator theory.

Lemma 1.[7] As a subspace of H_T , N_k is defined by:

$$N_k := \{u \in H_T \mid -\Delta^2 u(n-1) = \lambda_k u(n)\},$$

where $\lambda_k = 2 - 2\cos k\omega$, $\omega = 2\pi/T$, $k \in Z[0, [T/2]]$,

$[\cdot]$ denotes the Gauss Function. Then we claim that:

(i) $N_k \perp N_j$, $k \neq j$, $k, j \in Z[0, [T/2]]$.

(ii) $H_T = \bigoplus_{k=0}^{[T/2]} N_k$.

It is easy to obtain by using the methods in [7].

Lemma 2. $H_k := \bigoplus_{j=0}^k N_j$, $H_k^\perp := \bigoplus_{j=k+1}^{[T/2]} N_j$, $k \in Z[0, [T/2] - 1]$, then one has:

(i) $0 \leq \sum_{n=1}^T |\Delta u^2(n-1)|^2 \leq \lambda_k^2 \|u\|^2$, $\forall u \in H_k$,

(ii) $\lambda_{k+1}^2 \|u\|^2 \leq \sum_{n=1}^T |\Delta u^2(n-1)|^2 \leq \lambda_{[T/2]}^2 \|u\|^2$, $\forall u \in H_k^\perp$.

III. MAIN PROOFS

In this section, we will prove our four theorems.

Proof of Theorem 1. Let $u = \bar{u} + \tilde{u} \in H_0 \oplus H_0^\perp$, where $H_0 = N_0$, $H_0^\perp := \bigoplus_{j=1}^{[T/2]} N_j$. By (H₂), (2), (3) and Cauchy-Schwartz inequality, we have

$$\begin{aligned} & \left| \sum_{n=1}^T [F(n, u(n)) - F(n, \bar{u})] \right| \\ &= \left| \sum_{n=1}^T \int_0^1 (\nabla F(n, \bar{u} + s\tilde{u}(n)), \tilde{u}(n)) ds \right| \\ &\leq \sum_{n=1}^T \int_0^1 (M_1 |\bar{u} + s\tilde{u}(n)|^\alpha + M_2) |\tilde{u}(n)| ds \\ &\leq 2M_1 \sum_{n=1}^T |\bar{u}|^\alpha |\tilde{u}(n)| + 2M_1 \sum_{n=1}^T |\tilde{u}(n)|^{\alpha+1} \\ &\quad + M_2 \sum_{n=1}^T |\tilde{u}(n)| \\ &\leq \frac{\lambda_1^2}{4} \|\tilde{u}\|^2 + \frac{4M_1^2 T}{\lambda_1^2} |\bar{u}|^{2\alpha} + \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}\|^{\alpha+1} \\ &\quad + M_2 T \|\tilde{u}\|. \end{aligned}$$

From (5) and Lemma 2, we get

$$\begin{aligned} \varphi(u) &= \frac{1}{2} \sum_{n=1}^T |\Delta^2 u(n-1)|^2 + \sum_{n=1}^T F(n, u(n)) \\ &= \frac{1}{2} \sum_{n=1}^T |\Delta^2 \tilde{u}(n-1)|^2 + \sum_{n=1}^T F(n, u(n)) \\ &\geq \frac{\lambda_1^2}{2} \|\tilde{u}\|^2 + \sum_{n=1}^T [F(n, u(n)) - F(n, \bar{u})] \\ &\quad + \sum_{n=1}^T F(n, \bar{u}) \\ &\geq \frac{\lambda_1^2}{4} \|\tilde{u}\|^2 - \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}\|^{\alpha+1} - M_2 T \|\tilde{u}\| \\ &\quad + |\bar{u}|^{2\alpha} \left(| \bar{u} |^{-2\alpha} F(n, \bar{u}) - \frac{4M_1^2 T}{\lambda_1^2} \right) \end{aligned}$$

for all $u \in H_T$. As $\|u\| \rightarrow \infty$ if and only if $\|\tilde{u}\| + |\bar{u}| \rightarrow \infty$, the above inequality and (H₃) imply

$$\varphi(u) \rightarrow +\infty, \quad \text{as } \|u\| \rightarrow \infty.$$

That is $\varphi(u)$ is coercive. Hence There exists a bounded minimizing sequence. Since H_T is finite, a bounded minimizing sequence has a convergent subsequence. As $\varphi \in C'(H_T, R)$, there exists at least one point which φ achieves its infimum, that is $\varphi(u)$ has a critical point.

Proof of Theorem 2. First we prove that φ satisfies the *(PS)* condition. Suppose that $\{u_k\} \subset H_T$ is a consequence such that $-M_3 \leq \varphi(u_k) \leq M_3$, where $M_3 > 0$ and $\varphi'(u_k) \rightarrow 0$ as $k \rightarrow \infty$. Then for sufficiently large k and for any $u \in H_T$,

$$-\|u\| \leq \langle \varphi'(u_k), u \rangle \leq \|u\|.$$

Let $u_k = \bar{u}_k + \tilde{u}_k \in H_0 \oplus H_0^\perp$. In a way similar to the proof of Theorem 1, we have

$$\begin{aligned} & \sum_{n=1}^T \langle \nabla F(n, u_k(n)), \tilde{u}_k \rangle \\ & \leq \frac{\lambda_1^2}{4} \|\tilde{u}_k\|^2 + \frac{4M_1^2 T}{\lambda_1^2} |\bar{u}_k|^{2\alpha} \\ & \quad + \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}_k\|^{\alpha+1} + M_2 T \|\tilde{u}_k\|. \end{aligned}$$

By (6), we get

$$\begin{aligned} & \sum_{n=1}^T \langle \Delta^2 u_k(n-1), \Delta^2 \tilde{u}_k(n-1) \rangle \\ & = \langle \varphi'(u_k), \tilde{u}_k \rangle - \sum_{n=1}^T \langle \nabla F(n, u_k(n)), \tilde{u}_k \rangle \\ & \leq \|\tilde{u}_k\| + \frac{\lambda_1^2}{4} \|\tilde{u}_k\|^2 + \frac{4M_1^2 T}{\lambda_1^2} |\bar{u}_k|^{2\alpha} \\ & \quad + \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}_k\|^{\alpha+1} + M_2 T \|\tilde{u}_k\|. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} & \sum_{n=1}^T \langle \Delta^2 u_k(n-1), \Delta^2 \tilde{u}_k(n-1) \rangle \\ & = \sum_{n=1}^T \langle \Delta^2 \tilde{u}_k(n-1), \Delta^2 \tilde{u}_k(n-1) \rangle \geq \lambda_1^2 \|\tilde{u}_k\|^2. \end{aligned}$$

Thus we have

$$\begin{aligned} & \frac{4M_1^2 T}{\lambda_1^2} |\bar{u}_k|^{2\alpha} \\ & \geq \frac{3}{4} \lambda_1^2 \|\tilde{u}_k\|^2 - \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}_k\|^{\alpha+1} \\ & \quad - (1 + M_2 T) \|\tilde{u}_k\| \\ & = \frac{1}{2} \lambda_1^2 \|\tilde{u}_k\|^2 + \frac{1}{4} \lambda_1^2 \|\tilde{u}_k\|^2 \\ & \quad - \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}_k\|^{\alpha+1} - (1 + M_2 T) \|\tilde{u}_k\| \\ & \geq \frac{1}{2} \lambda_1^2 \|\tilde{u}_k\|^2 - M_4, \end{aligned}$$

where $M_4 > 0$. Hence we get for all large k

$$\|\tilde{u}_k\| \leq M_5 |\bar{u}_k|^\alpha + M_6, \quad (8)$$

where $M_5 > 0$, $M_6 > 0$.

It follows from the boundedness of $\varphi(u_k)$,

$$\begin{aligned} M_3 & \leq \varphi(u_k) \\ & = \frac{1}{2} |\Delta^2 \tilde{u}_k(n-1)|^2 + \sum_{n=1}^T F(n, \bar{u}) \\ & \quad + \sum_{n=1}^T [F(n, u_k(n)) - F(n, \bar{u})] \\ & \leq \frac{3\lambda_{[T/2]}^2}{4} \|\tilde{u}_k\|^2 + \frac{4M_1^2 T}{\lambda_{[T/2]}^2} |\bar{u}_k|^{2\alpha} \\ & \quad + \frac{2M_1}{C_1^{\alpha+1}} \|\tilde{u}_k\|^{\alpha+1} + M_2 T \|\tilde{u}_k\| + \sum_{n=1}^T F(n, \bar{u}) \\ & \leq |\bar{u}_k|^{2\alpha} \|u_k\|^{-2\alpha} \sum_{n=1}^T F(n, \bar{u}_k) \\ & \quad + M_7 + M_8(|\bar{u}_k|) \end{aligned}$$

for sufficiently large k , where constant $M_7 > 0$ and $M_8(|\bar{u}_k|) \rightarrow 0$ as $|\bar{u}_k| \rightarrow \infty$. The above inequality and (H_4) imply that $\{\bar{u}_k\}$ is bounded. Then it follows from (8) that $\{u_k\}$ is bounded. Since H_T is a finite dimensional space, there exists a subsequence of $\{u_k\}$ convergent in H_T . Thus we conclude that the *(PS)* condition is satisfied.

In order to use the saddle point theorem ([5, Theorem 4.6]), we only need to verify the following

- (I₁) $\varphi(x) \rightarrow -\infty$ as $|x| \rightarrow \infty$ in H_0 .
- (I₂) $\varphi(u) \rightarrow +\infty$ as $\|u\| \rightarrow \infty$ in H_0^\perp .

In fact, from (H_4) , for any $x \in H_0$, we have

$$\varphi(x) = \sum_{n=1}^T F(n, x) \rightarrow -\infty, \quad \text{as } |x| \rightarrow \infty.$$

Thus (I₁) is easy to verify.

Next, by (H_2) , we have for all $n \in Z[1, T]$ and $x \in R^N$,

$$\begin{aligned} F(n, x) & \leq |F(n, x) - F(n, 0)| + |F(n, 0)| \\ & \leq \int_0^1 |(\nabla F(n, sx), x)| ds + |F(n, 0)| \\ & \leq \int_0^1 |(M_1 |sx|^\alpha + M_2) |x| ds + |F(n, 0)| \\ & \leq M_1 |x|^{\alpha+1} + M_2 |x| + M_9, \end{aligned}$$

where $M_9 > 0$.

By (5) and the above inequality, for any $u \in H_0^\perp$, we have

$$\begin{aligned} & \varphi(u) \\ & = \frac{1}{2} |\Delta^2 u(n-1)|^2 + \sum_{n=1}^T F(n, u(n)) \\ & \geq \frac{\lambda_1^2}{2} \|u\|^2 - \sum_{n=1}^T (M_1 |u(n)|^{\alpha+1} + M_2 |u(n)| + M_9) \\ & \geq \frac{\lambda_1^2}{2} \|u\|^2 - \frac{M_1}{C_1^{\alpha+1}} \|u\|^{\alpha+1} - M_2 T \|u\| - M_9 T \end{aligned}$$

Since $0 \leq \alpha < 1$, we can obtain $\varphi(u) \rightarrow +\infty$ as $\|u\| \rightarrow \infty$ in H_0^\perp . The proof of Theorem 2 is complete.

Proof of Theorem 3. First we prove that φ satisfies condition *(C)* (I1), that is for every sequence $\{u_k\} \subset H_T$, $\{u_k\}$ has a

convergent subsequence if $\varphi(u_k)$ is bounded and $\|\varphi'(u_k)\|(1+\|u_k\|) \rightarrow 0$ as $k \rightarrow \infty$.

Let $\{u_k\} \subset H_T$ be a sequence such that $\{\varphi(u_k)\}$ is bounded and

$$\|\varphi'(u_k)\|(1+\|u_k\|) \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Then there exists a constant C_3 , such that

$$|\varphi(u_k)| \leq C_3, \quad \|\varphi'(u_k)\|(1+\|u_k\|) \leq C_3.$$

Thus, one has

$$\begin{aligned} 3C_3 &\geq \langle \varphi'(u_k), u_k \rangle - 2\varphi(u_k) \\ &= \sum_{n=1}^T [(\nabla F(n, u_k(n)), u_k(n)) - 2F(n, u_k(n))]. \end{aligned}$$

From (H₅), $\{u_k\}$ is bounded. So $\{u_k\}$ has a convergent subsequence and condition (C) is verified.

Next, we show that φ satisfies (I₁), (I₂).

From (H₆), we can get that for all $\varepsilon > 0$, there is a constant $G_1 > 0$, such that

$$(\nabla F(n, x), x) - 2F(n, x) > \frac{1}{\varepsilon}$$

for all $|x| > G_1$. That is

$$(\nabla F(n, sx), sx) - 2F(n, sx) > \frac{1}{\varepsilon}$$

for all $|sx| > G_1$. Then we have

$$\begin{aligned} \frac{d}{ds} \left(\frac{F(n, sx)}{s^2} \right) &= \frac{s(\nabla F(n, sx), x) - 2F(n, sx)}{s^3} \\ &> \frac{1}{\varepsilon s^3} = -\frac{d}{ds} \left(\frac{1}{2\varepsilon s^2} \right). \end{aligned}$$

Let $s > 1$,

$$\int_1^s \frac{d}{ds} \left(\frac{F(n, sx)}{s^2} \right) > -\int_1^s \frac{d}{ds} \left(\frac{1}{2\varepsilon s^2} \right).$$

So we obtain

$$\frac{F(n, sx)}{s^2} - F(n, x) > -\frac{1}{2\varepsilon s^2} + \frac{1}{2\varepsilon}.$$

Let $s \rightarrow +\infty$ in the above inequality, we have $F(n, x) < -\frac{1}{2\varepsilon}$ for all $|x| > G_1$. From the arbitrariness of ε , one can conclude that $F(n, x) \rightarrow -\infty$ as $|x| \rightarrow \infty$. Then it is similar to the proof of (I₁) in Theorem 2, we can get that $\varphi(x) \rightarrow -\infty$ as $|x| \rightarrow \infty$ in H_0 .

Take $\varepsilon = \frac{1}{4}\lambda_1^2$, by (H₆), there exists a constant $G_2 > 0$, such that for all $|x| > G_2$

$$|F(n, x)| \leq \frac{1}{4}\lambda_1^2|x|^2. \quad (9)$$

Then from Lemma 2 and (9), for any $u \in H_0^\perp$, we have

$$\begin{aligned} \varphi(u) &= \frac{1}{2}|\Delta^2 u(n-1)|^2 + \sum_{n=1}^T F(n, u(n)) \\ &\geq \frac{1}{2}\lambda_1^2\|u\|^2 - \frac{1}{4}\lambda_1^2 \sum_{n=1}^T |u(n)|^2 \\ &= \frac{1}{2}\lambda_1^2\|u\|^2 - \frac{1}{4}\lambda_1^2\|u\|^2 = \frac{1}{4}\lambda_1^2\|u\|^2 \end{aligned}$$

thus we can conclude that $\varphi(u) \rightarrow +\infty$ as $\|u\| \rightarrow \infty$ in H_0^\perp . It follows from the saddle point theorem that Theorem 3 hold. The proof of Theorem 3 is complete.

Proof of Theorem 4. To proof Theorem 4, we only need to specify that F satisfying the conditions of Theorem 4 fulfil the ones of Theorem 3. That is, we just need to show that conditions (H₇), (H₈) imply (H₅), (H₆).

It follows from (H₇), (H₈) that for all $n \in Z[1, T]$,

$$\begin{aligned} &(\nabla F(n, x), x) - 2F(n, x) \\ &= [(\nabla F(n, x), x) - \beta F(n, x)] + (\beta - 2)F(n, x) \\ &\geq (\beta - 2)F(n, x) \rightarrow +\infty \end{aligned}$$

as $|x| \rightarrow \infty$, which implies (H₅).

By (H₈), there exists a constant $G_3 > G$, such that for $|x| > G_3$, $-F(n, x) > 0$. From (H₇), we have

$$\begin{aligned} &\frac{d}{ds} \left(-\frac{F(n, s\frac{x}{|x|})}{s^\beta} \right) \\ &= \frac{-s(\nabla F(n, s\frac{x}{|x|}), \frac{x}{|x|}) + \beta F(n, s\frac{x}{|x|})}{s^{\beta+1}} \leq 0. \end{aligned}$$

Integrate both sides of the above inequality from G_3 to $|x|$ and obtain

$$-F(n, x) \leq M_{10}|x|^\beta,$$

where $M_{10} > 0$. For any $x \in R^N$, we have

$$0 < -F(n, x) \leq M_{10}|x|^\beta + M_{11},$$

where $M_{11} = \max\{-F(n, x) : |x| \leq G_3, n \in Z[1, T]\}$. Hence

$$0 \leq -\frac{F(n, x)}{|x|^2} \leq \frac{M_{10}}{|x|^{2-\beta}} + \frac{M_{11}}{|x|^2}.$$

Since $0 < \beta < 2$, we conclude that $\frac{F(n, x)}{|x|^2} \rightarrow 0$, as $|x| \rightarrow \infty$. Thus we have got (H₆). The proof of Theorem 4 is complete.

REFERENCES

- [1] G. Cerami, An existence criterion for the critical points on unbounded manifolds, Istit. Lombardo Accad. Sci. Lett. Rend. A 112 (2) (1978) 332-336 (in Italian).
- [2] Xiaochun Cai, J. Yu, Z. Guo, Existence of periodic for fourth-order difference equations, Computers and Mathematics with Applications, 50(2005) 49-55.
- [3] Z. Guo, J. Yu, The existence of periodic and subharmonic solutions of subquadratic second order difference equations, J. London Math. Soc. (2) 68 (2) (2003) 419-430.
- [4] Z. Guo, J. Yu, Existence of periodic and subharmonic solutions for second-order superlinear difference equations, Sci. China Ser. A 46 (4)(2003) 506-515.
- [5] J. Mawhin, M. Willem, Critical point theory and Hamiltonian systems, in: Applied Mathematical Sciences, vol. 74, Springer-Verlag, New York, 1989.
- [6] Y. F. Xue, C. L. Tang, Existence of a periodic solution for subquadratic second-order discrete Hamiltonian systems, Nonlinear Analysis, 67 (2007) 2072-2080.
- [7] Y. F. Xue, C. L. Tang, Multiple periodic solutions for superquadratic second-order discrete Hamiltonian systems, Applied Mathematics and Computation 196 (2008) 494-500.
- [8] J. Yu, Z. Guo, X. Zou, Periodic solutions of second order self-adjoint difference equations, J. London Math. Soc. (2) 71 (1) (2005) 146-160.

Homoclinic orbits for a class of non-periodic Hamiltonian systems

Jian Ding, Guizhen Feng, Weili Wu

Abstract—We study the existence of homoclinic orbits for the second order Hamiltonian system $\ddot{u} + V_u(t, u) = f(t)$. Let $f \in C(\mathbb{R}, \mathbb{R}^n)$ and $V(t, u) = -K(t, u) + W(t, u) \in C^1(\mathbb{R} \times \mathbb{R}^n, \mathbb{R})$, where K is a quadratic growth function and W is allowed to be asymptotically quadratic in u at infinity besides the super quadratic case. Since the system is neither autonomous nor periodic, the (PS) condition is difficult to check when we use the Mountain Pass theorem. Therefore, we approximate the homoclinic orbits by virtue of the solutions of a sequence of nil-boundary-value problems.

Index Terms—Hamiltonian systems, Homoclinic orbits, Condition (C), Asymptotically quadratic, Mountain Pass theorem, Arzela-Ascoli theorem.

I. INTRODUCTION AND THE MAIN RESULT

IN this paper, we consider the existence of homoclinic orbits for the second order Hamiltonian system:

$$\ddot{u}(t) + V_u(t, u(t)) = f(t) \quad (\text{I.1})$$

where $f \in C(\mathbb{R}, \mathbb{R}^n)$ and $V(t, u) = -K(t, u) + W(t, u) \in C^1(\mathbb{R} \times \mathbb{R}^n, \mathbb{R})$.

Let us recall that a solution $u(t)$ of (I.1) is homoclinic (to 0) if $u(t) \rightarrow 0$ and $\dot{u}(t) \rightarrow 0$ as $t \rightarrow \pm\infty$. In addition, if $u(t) \not\equiv 0$ then u is called a nontrivial homoclinic orbit.

In recent years, the existence of homoclinic orbits for (I.1) has been extensively studied by variational methods (see [1] – [5]). Most of them assumed that the nonlinearity $W(t, u)$ satisfies the Ambrosetti-Rabinowitz condition, that is, there exists $\mu > 2$ such that

$$0 < \mu W(t, u) \leq (W_u(t, u), u)$$

for all $t \in \mathbb{R}$ and $u \in \mathbb{R}^n \setminus \{0\}$. Using the critical point theory, they obtained the existence of homoclinic orbits. Later some papers studied (I.1) under some weaker super quadratic assumptions than the (A-R) condition (see [6], [7]). There were also some papers considered the sub-quadratic case ([8], [9]) and the asymptotically quadratic case ([10], [11]).

It is well known that the major difficulty is to prove the (PS) condition when one applies the Mountain Pass theorem. If $V(t, u)$ is periodic in t , the problem is a little simple and

Manuscript received March 7; revised May 10.

Jian Ding is with the Department of Mathematics, Southeast University, Nanjing, 210018, PR. China. E-mail: df2001101@126.com.

Guizhen Feng is with the Division of Basic Courses, Nanjing Institute of Industry Technology, Nanjing, 210000, PR. China. E-mail: fengguizhen@gmail.com.

Weili Wu is with the Mathematics and Physical Sciences Technology, Xuzhou Institute of Technology, Xuzhou, 221008, PR. China. E-mail: wuweili-math@126.com. His work was supported in part by the Natural Science Fund for Colleges and Universities in Jiangsu Province (07KJD110197).

there are many results. In [7] and [12], the authors considered (I.1). When $V(t, u) = -\frac{1}{2}(L(t)u, u) + W(t, u)$ is independent of t , i.e., the system (I.1) is autonomous, they obtained one homoclinic orbit as a limit of solutions of a certain sequence of periodic systems. By this method, [13] considered the case that $L(t)$ and $W(t, u)$ are periodic in t . It also assumed that $L(t)$ was positive definite and symmetric, $W(t, u)$ satisfied the (A-R) condition. [14] replaced $\frac{1}{2}(L(t)u, u)$ by $K(t, u)$ which satisfied the following so-called pinching condition,

$$b_1|u|^2 \leq K(t, u) \leq b_2|u|^2$$

for some positive constants b_1 and b_2 ; If V is neither autonomous nor periodic in t , the problem is quite different from the ones' just described for lack of compactness of the Sobolev embedding. [12] imposed a coercivity condition on L , that is, the smallest eigenvalue of $L(t)$ tends to $+\infty$ as $|t| \rightarrow \infty$. Using a variant of the Mountain Pass theorem, it obtained one homoclinic orbit of (I.1). In [4], the authors removed the technical coercivity in the case where $L(t)$ and $W(t, u)$ are even in t (see also [5], [15]). They approximated the homoclinic orbits by the solutions of nil-boundary-value problems.

In this paper, we also consider the homoclinic orbits for (I.1) by the method in [4] and [14]. But here we make no any periodicity, coercivity or even assumptions. Moreover, we consider some new growth condition being different from [4] and [14]. Precisely, we assume $W(t, u)$ may be asymptotically quadratic at infinity besides the super quadratic case. In addition, $K(t, u)$ may not satisfy the pinching condition.

We make some assumptions as follows:

(A₁) $V(t, u) = -K(t, u) + W(t, u) \in C^1(\mathbb{R} \times \mathbb{R}^n)$, and $V_u(t, u) \rightarrow 0$ as $|u| \rightarrow 0$ uniformly in $t \in \mathbb{R}$;

(A₂) there exist constants $d_1 > 0, \gamma \geq 2$ such that $W(t, u) \leq d_1|u|^\gamma$ for all $(t, u) \in \mathbb{R} \times \mathbb{R}^n$;

(A₃) there exist constants $d_2 > 0$ and $\gamma \geq \mu > \gamma - 1$ and $\beta \in L^1(\mathbb{R}, [0, \infty))$ such that

$$(W_u(t, u), u) - 2W(t, u) \geq d_2|u|^\mu - \beta(t)$$

for all $(t, u) \in \mathbb{R} \times \mathbb{R}^n$;

(A₄) there exist positive constants $T > 0, R > 1$ and $\rho_0 \leq 1$ such that for $t \in [-T, T]$, $W(t, u) \leq 0$ as $|u| \leq \rho_0$ and $W(t, u) \geq \frac{2\pi^2}{T^2}|u|^2$ as $|u| > R$;

(A₅) $K(t, 0) = 0, 0 < \underline{b} := \inf_{\{t \in [-T, T], |\xi|=1\}} K(t, \xi) \leq \bar{b} := \sup_{\{t \in [-T, T], |\xi|=1\}} K(t, \xi) \leq \frac{\pi^2}{2T^2}$, and $0 \leq (u, K_u(t, u)) \leq 2K(t, u)$ for all $(t, u) \in \mathbb{R} \times \mathbb{R}^n$.

By Sobolev embedding theorem, $H^1(\mathbb{R})$ is continuously embedded into $L^\infty(\mathbb{R})$. There exists a constant $C_\infty > 0$ such

that

$$|u|_{L^\infty} \leq C_\infty \|u\|_{H^1} \quad \text{for all } u \in H^1.$$

Additionally, we make the following assumption:
(A₆) $f \in C(\mathbb{R})$, $f(t) \not\equiv 0$, $f(t) = 0$ for $|t| \geq T$ and $\max_{t \in [-T, T]} |f(t)| \leq C_0 := \min\{\frac{1}{2}, b\} \frac{\rho_0}{2\sqrt{2T}C_\infty}$.

Our main result is the following theorem:

Theorem 1.1: If the assumptions (A₁) – (A₆) are satisfied, then the system (1.1) possesses a nontrivial homoclinic orbit.

Remark 1.1: From (A₂) and (A₄) we can easily see that W may be asymptotically quadratic in u at infinity; In addition, $f(t) \not\equiv 0$ ensures that $u(t) \equiv 0$ can't be a solution of (I.1).

Remark 1.2: There are functions satisfy the assumptions above. For example, one can easily check that if

$$K(t, x) = \begin{cases} \left(1 + \frac{1}{1+t^2}\right)x^2 & x \geq 0, \\ \left(1 + \frac{2}{1+t^2}\right)x^2 & x < 0 \end{cases}$$

and

$$W(t, x) = -2x^2 + x^3 \quad (\text{the super quadratic case})$$

or

$$W(t, x) = x^2 - 2x^{\frac{3}{2}} \quad (\text{the asymptotically quadratic case})$$

where $t, x \in \mathbb{R}$, then $V(t, x) = -K(t, x) + W(t, x)$ satisfies (A₁) – (A₅) while W doesn't satisfy the Ambrosetti-Rabinowitz condition.

II. PROOF OF THE MAIN RESULT

We shall obtain the solution of (I.1) as the limit as $k \rightarrow \infty$ of the solutions of

$$\begin{aligned} \ddot{u}(t) + V_u(t, u(t)) &= f(t) \quad \text{for } t \in (-kT, kT), \\ u(-kT) &= u(kT) = 0. \end{aligned} \quad (\text{II.1})$$

For each $k \in \mathbb{N}$ and $p \in [2, \infty)$, denote

$$L_k^p := \left\{ u : [-kT, kT] \rightarrow \mathbb{R}^n \mid |u|_{L_k^p} < \infty \right\},$$

where $|u|_{L_k^p} := \left(\int_{-kT}^{kT} |u(t)|^p dt \right)^{\frac{1}{p}}$.

Let

$$L_k^\infty := \left\{ u : [-kT, kT] \rightarrow \mathbb{R}^n \mid |u|_{L_k^\infty} := \text{ess sup}\{|u(t)|\} < \infty \right\}.$$

Set

$$E_k := \left\{ u : [-kT, kT] \rightarrow \mathbb{R}^n \mid u \text{ is absolutely continuous, } u(-kT) = u(kT) = 0 \right\},$$

equipped with the norm

$$\|u\|_k = \left(\int_{-kT}^{kT} (|\dot{u}|^2 + |u|^2) dt \right)^{\frac{1}{2}}.$$

By zero extensions, we can see $E_k \subset E_{k+1} \subset H^1$ for all $k \in \mathbb{N}$. Therefore

$$|u|_{L_k^\infty} \leq C_\infty \|u\|_k \quad \text{for all } u \in E_k.$$

We using variational methods to study (II.1) and define

$$\varphi_k(u) = \int_{-kT}^{kT} \left[\frac{1}{2} |\dot{u}(t)|^2 - V(t, u(t)) + (f(t), u(t)) \right] dt.$$

It is clear that $\varphi_k \in C^1(E_k, \mathbb{R})$ and

$$\begin{aligned} &\varphi_k'(u)v \\ &= \int_{-kT}^{kT} \left[(\dot{u}(t), \dot{v}(t)) - (V_u(t, u(t)), v(t)) + (f(t), v(t)) \right] dt. \end{aligned}$$

We all know that critical points of φ_k are classical solutions of (II.1).

Lemma 2.1: Under (A₁) – (A₆), for each $k \in \mathbb{N}$, the system (II.1) possesses a nontrivial solution.

We will prove this result via the Mountain Pass theorem by Rabinowitz in [11]. Instead of the (PS) condition, here we use condition (C). Recall a function φ satisfies condition (C) on E if any sequence $\{u_j\} \subset E$ such that $\{\varphi(u_j)\}$ is bounded and $(1 + \|u_j\|)\|\varphi'(u_j)\| \rightarrow 0$ has a convergent subsequence. The Mountain Pass theorem still holds true under condition (C) (see [16]).

Before the proof of Lemma 2.1, we need the following proposition.

Proposition 2.1: Under (A₅), for every $t \in \mathbb{R}$, the following inequalities hold:

$$\begin{aligned} K(t, u) &\geq \underline{b}|u|^2 \quad \text{if } 0 < |u| \leq 1, \\ K(t, u) &\leq \bar{b}|u|^2 \quad \text{if } |u| \geq 1. \end{aligned} \quad (\text{II.2})$$

Proof for each $t \in \mathbb{R}$ and $u \in \mathbb{R}^n$, let $f(\xi) = K(t, \frac{u}{\xi})\xi^2$, $\xi \in (0, +\infty)$, then by (A₅),

$$\begin{aligned} f'(\xi) &= K_u(t, \frac{u}{\xi})\xi^2 \left(-\frac{u}{\xi^2}\right) + K(t, \frac{u}{\xi})(2\xi) \\ &= (2K(t, \frac{u}{\xi}) - K_u(t, \frac{u}{\xi})\frac{u}{\xi})\xi \\ &\geq 0. \end{aligned}$$

So $f(\xi)$ is nondecreasing. Hence we easily obtain (II.2). \square

Proof of lemma 2.1 Under our assumptions, it is easy to see that $\varphi_k(0) = 0$.

Step 1. φ_k satisfies condition (C).

Suppose $\{u_j\} \subset E_k$, $\{\varphi_k(u_j)\}$ is bounded and $(1 + \|u_j\|)\|\varphi_k'(u_j)\| \rightarrow 0$ as $j \rightarrow \infty$. Then there is a constant $M_k > 0$ such that

$$\varphi_k(u_j) \leq M_k, \quad (1 + \|u_j\|)\|\varphi_k'(u_j)\| \leq M_k \quad (\text{II.3})$$

for all $j \in \mathbb{N}$.
By (II.3), (A₃), (A₅) and (A₆),

$$\begin{aligned}
3M_k &\geq 2\varphi_k(u_j) - \varphi'_k(u_j)u_j \\
&= \int_{-kT}^{kT} \left[2K(t, u_j(t)) - \left(K_u(t, u_j(t)), u_j(t) \right) \right] dt \\
&\quad + \int_{-kT}^{kT} \left[\left(W_u(t, u_j(t)), u_j(t) \right) - 2W(t, u_j(t)) \right] dt \\
&\quad + \int_{-kT}^{kT} (f(t), u_j(t)) dt \\
&\geq d_2 \int_{-kT}^{kT} |u_j(t)|^\mu dt - \int_{-kT}^{kT} \beta(t) dt - C_0\sqrt{2T}\|u_j\|_k \\
&\geq d_2 \int_{-kT}^{kT} |u_j(t)|^\mu dt - \beta_0 - C_0\sqrt{2T}\|u_j\|_k,
\end{aligned}$$

where $\beta_0 = \int_{-\infty}^{\infty} \beta(t) dt$.

Therefore,

$$\int_{-kT}^{kT} |u_j(t)|^\mu dt \leq \frac{1}{d_2} (3M_k + \beta_0 + C_0\sqrt{2T}\|u_j\|_k).$$

By (A₅), (A₆) and Proposition 2.1,

$$\begin{aligned}
\frac{1}{2}|\dot{u}_j|_{L_k^2}^2 &= \varphi_k(u_j) - \int_{-kT}^{kT} K(t, u_j(t)) dt \\
&\quad + \int_{-kT}^{kT} W(t, u_j(t)) dt - \int_{-kT}^{kT} (f(t), u_j(t)) dt \\
&\leq M_k - \underline{b} \int_{\{t \in [-kT, kT] \mid |u_j(t)| \leq 1\}} |u_j(t)|^2 dt \\
&\quad + d_1 \int_{-kT}^{kT} |u_j(t)|^\gamma dt + C_0\sqrt{2T}\|u_j\|_k.
\end{aligned}$$

Then one has

$$\begin{aligned}
\frac{1}{2}|\dot{u}_j|_{L_k^2}^2 + \underline{b} \int_{\{t \in [-kT, kT] \mid |u_j(t)| \leq 1\}} |u_j(t)|^2 dt \\
\leq M_k + d_1 \int_{-kT}^{kT} |u_j(t)|^\gamma dt + C_0\sqrt{2T}\|u_j\|_k.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\min\left\{\frac{1}{2}, \underline{b}\right\} \|u_j\|_k^2 \\
&\leq M_k + \underline{b} \int_{\{t \in [-kT, kT] \mid |u_j(t)| > 1\}} |u_j(t)|^2 dt \\
&\quad + d_1 \int_{-kT}^{kT} |u_j(t)|^\gamma dt + C_0\sqrt{2T}\|u_j\|_k \\
&\leq M_k + \underline{b} \int_{-kT}^{kT} |u_j(t)|^\gamma dt \\
&\quad + d_1 \int_{-kT}^{kT} |u_j(t)|^\gamma dt + C_0\sqrt{2T}\|u_j\|_k \\
&\leq M_k + (\underline{b} + d_1) |u_j|_{L_k^\infty}^{\gamma-\mu} \int_{-kT}^{kT} |u_j(t)|^\mu dt \\
&\quad + C_0\sqrt{2T}\|u_j\|_k \\
&\leq M_k + (\underline{b} + d_1) C_\infty^{\gamma-\mu} \|u_j\|_k^{\gamma-\mu} \int_{-kT}^{kT} |u_j(t)|^\mu dt \\
&\quad + C_0\sqrt{2T}\|u_j\|_k \\
&\leq M_k + \frac{1}{d_2} (\underline{b} + d_1) C_\infty^{\gamma-\mu} \|u_j\|_k^{\gamma-\mu} (3M_k + \beta_0) \\
&\quad + C_0\sqrt{2T}\|u_j\|_k + C_0\sqrt{2T}\|u_j\|_k.
\end{aligned}$$

Since $\gamma - \mu < 1$, we get $\{\|u_j\|_k\}$ is bounded. Going if necessary to a subsequence, we can assume that there exists $u \in E_k$ such that $u_j \rightarrow u$ in E_k as $j \rightarrow +\infty$, which implies $u_j \rightarrow u$ uniformly on $[-kT, kT]$.

Therefore,

$$(\varphi'_k(u_j) - \varphi'_k(u))(u_j - u) \rightarrow 0,$$

$$|u_j - u|_{L_k^2} \rightarrow 0$$

and

$$\int_{-kT}^{kT} \left(V_u(t, u_j(t)) - V_u(t, u(t)), u_j(t) - u(t) \right) dt \rightarrow 0$$

as $j \rightarrow +\infty$.

By an easy computation, we can see that

$$\begin{aligned}
(\varphi'_k(u_j) - \varphi'_k(u))(u_j - u) &= |\dot{u}_j - \dot{u}|_{L_k^2}^2 \\
&\quad - \int_{-kT}^{kT} \left(V_u(t, u_j(t)) - V_u(t, u(t)), u_j(t) - u(t) \right) dt.
\end{aligned}$$

Hence we have $|\dot{u}_j - \dot{u}|_{L_k^2} \rightarrow 0$, and so we get $u_j \rightarrow u$ in E_k .

Step 2. There are constants $\rho > 0$ and $\alpha > 0$ independent of k , such that $\varphi_k|_{S_\rho} \geq \alpha$, where $S_\rho = \{u \in E_k \mid \|u\|_k = \rho\}$.

Choose $\rho = \frac{\rho_0}{C_\infty}$, then for $u \in S_\rho$ we have $|u|_{L_k^\infty} \leq \rho_0$. Therefore, $|u| \leq \rho_0 \leq 1$ for all $t \in [-kT, kT]$, and then by

(A₄), $W(t, u) \leq 0$. Together with (II.2), we obtain

$$\begin{aligned}
\varphi_k(u) &= \int_{-kT}^{kT} \left[\frac{1}{2} |\dot{u}(t)|^2 + K(t, u(t)) - W(t, u(t)) \right] dt \\
&\quad + \int_{-kT}^{kT} (f(t), u(t)) dt \\
&\geq \frac{1}{2} \int_{-kT}^{kT} |\dot{u}(t)|^2 dt + \underline{b} \int_{-kT}^{kT} |u(t)|^2 dt \\
&\quad - C_0 \sqrt{2T} \|u(t)\|_k \\
&\geq \min\left\{ \frac{1}{2}, \underline{b} \right\} \|u(t)\|_k^2 - C_0 \sqrt{2T} \|u\|_k \\
&= \min\left\{ \frac{1}{2}, \underline{b} \right\} \rho^2 - C_0 \sqrt{2T} \rho \\
&= \min\left\{ \frac{1}{2}, \underline{b} \right\} \frac{\rho_0^2}{2C_\infty^2} \\
&\triangleq \alpha.
\end{aligned}$$

Step 3. For the ρ defined as above, there exists $e_k \in E_k$ such that $\|e_k\|_k > \rho$, $\varphi_k(e_k) \leq 0$.

By (A₄),

$$\frac{W(t, u)}{|u|^2} \geq \frac{2\pi^2}{T^2}$$

for all $|u| > R$ and $t \in [-T, T]$.

Let $\delta = \max_{\{t \in [-T, T], |u| \leq R\}} |W(t, u)|$, we obtain

$$W(t, u) \geq \frac{2\pi^2}{T^2} (|u|^2 - R^2) - \delta \quad (\text{II.4})$$

for all $u \in \mathbb{R}^n$, $t \in [-T, T]$.

Set

$$e_k(t) = \begin{cases} s \sin(\omega t) e, & t \in [-T, T] \\ 0, & t \in [-kT, kT] \setminus [-T, T] \end{cases} \quad (\text{II.5})$$

where $\omega = \frac{\pi}{T}$, $e = (1, 0, \dots, 0)$. Obviously, $e_k \in E_k$, and $\|e_k\|_k \rightarrow \infty$ as $s \rightarrow \infty$. We can assume s is large enough such that $\|e_k\|_k \geq \max\{1, \rho\}$. With (II.2) and (II.4), by an easy computation, we can obtain

$$\begin{aligned}
&\varphi_k(e_k(t)) \\
&= \int_{-kT}^{kT} \left[\frac{1}{2} |\dot{e}_k(t)|^2 + K(t, e_k(t)) - W(t, e_k(t)) \right. \\
&\quad \left. + (f(t), e_k(t)) \right] dt \\
&\leq \int_{-T}^T \frac{1}{2} |\dot{e}_k(t)|^2 dt + \bar{b} \int_{-T}^T |e_k(t)|^2 dt - \frac{2\pi^2}{T^2} \int_{-T}^T |e_k(t)|^2 dt \\
&\quad + \left(\frac{2\pi^2 R^2}{T^2} + \delta \right) 2T + C_0 \sqrt{2T} \left(\int_{-T}^T |e_k(t)|^2 dt \right)^{\frac{1}{2}} \\
&\leq \frac{1}{2} s^2 \omega^2 \int_{-T}^T |\cos(\omega t)|^2 dt \\
&\quad + \left(\bar{b} s^2 - \frac{2\pi^2 s^2}{T^2} \right) \int_{-T}^T |\sin(\omega t)|^2 dt \\
&\quad + C_0 \sqrt{2T} s \left(\int_{-T}^T |\sin(\omega t)|^2 dt \right)^{\frac{1}{2}} + 2T \left(\delta + \frac{2\pi^2 R^2}{T^2} \right) \\
&= \left(\frac{1}{2} \omega^2 + \bar{b} - \frac{2\pi^2}{T^2} \right) s^2 T + C_0 \sqrt{2T} s T^{\frac{1}{2}} + 2T \left(\delta + \frac{2\pi^2 R^2}{T^2} \right) \\
&\leq -\frac{\pi^2}{T} s^2 + \sqrt{2} C_0 T s + 2T \left(\delta + \frac{2\pi^2 R^2}{T^2} \right) \rightarrow -\infty
\end{aligned}$$

as $s \rightarrow \infty$. So for all $k \in \mathbb{N}$, we can choose a s large enough such that e_k defined as above satisfies $\|e_k\|_k > \rho$ and $\varphi_k(e_k) \leq 0$.

Therefore, by the Mountain Pass theorem, for each $k \in \mathbb{N}$, φ_k possesses a critical point $u_k \in E_k$ with value $c_k = \varphi_k(u_k) \geq \alpha$, which means u_k is a nontrivial solution of (II.1). Moreover,

$$c_k = \inf_{g \in \Gamma_k} \max_{s \in [0, 1]} \varphi_k(g(s))$$

where

$$\Gamma_k = \{g \in C([0, 1], E_k) : g(0) = 0, g(1) = e_k\}. \quad \square$$

By Lemma 2.1, we can obtain a sequence $\{u_k\}$ such that for each $k \in \mathbb{N}$,

$$\varphi_k(u_k) = c_k \geq \alpha, \quad \varphi'_k(u_k) = 0.$$

In the following, we will show that there exists a subsequence of $\{u_k\}$ which almost uniformly converges to a C^2 function.

We denote $C_{loc}^p(\mathbb{R}, \mathbb{R}^n)$ ($p \in \mathbb{N} \cup \{0\}$), the space of C^p functions on \mathbb{R} with values in \mathbb{R}^n under the topology of almost uniformly convergence of functions and all derivatives up to the order p . We have the following result:

Lemma 2.2: Let $\{u_k\}_{k \in \mathbb{N}}$ be the sequence given as above. Then it possesses a subsequence also denoted by $\{u_k\}$ and a C^2 function $u : \mathbb{R} \rightarrow \mathbb{R}^n$ such that $u_k \rightarrow u$ in $C_{loc}^2(\mathbb{R}, \mathbb{R}^n)$ as $k \rightarrow +\infty$.

Proof We will prove this lemma by virtue of the Arzela-Ascoli theorem. We first show that the sequence $\{c_k\}_{k \in \mathbb{N}}$ and $\{\|u_k\|_k\}_{k \in \mathbb{N}}$ are bounded.

For any $m < n$, $E_m \subset E_n$. Hence the set of competing paths in Γ_n is greater than Γ_m , which implies $c_n \leq c_m \leq c_1$, and then $c_k \leq c_1$ for all $k \in \mathbb{N}$.

Since

$$\varphi_k(u_k) = c_k \leq c_1$$

and

$$\varphi'_k(u_k) = 0,$$

just as proof of Step 1. in Lemma 2.1, it is easy to prove $\{u_k\}$ is bounded uniformly in k . Therefore there exists a constant $M_1 > 0$ independent of k such that

$$\|u_k\|_k \leq M_1 \quad \text{for all } k \in \mathbb{N}.$$

From

$$\begin{aligned}
|u_k(t_1) - u_k(t_2)| &\leq \int_{t_1}^{t_2} |\dot{u}_k(t)| dt \\
&\leq (t_2 - t_1)^{\frac{1}{2}} \left(\int_{t_1}^{t_2} |\dot{u}_k(t)|^2 dt \right)^{\frac{1}{2}},
\end{aligned}$$

we can see $(u_n(t))$ is equicontinuous on each $[-T_n, T_n]$. Hence by the Arzela-Ascoli Theorem, it has a uniformly convergent subsequence on each $[-T_n, T_n]$.

Let $\{u_{n_k}^1\}$ be a subsequence of $\{u_n\}$ that converges on $[-T_1, T_1]$. Then it is equicontinuous and uniformly bounded on $[-T_2, T_2]$. So we can choose a subsequence $\{u_{n_k}^2\}$ of $\{u_{n_k}^1\}$ that converges uniformly on $[-T_2, T_2]$. Repeat this procedure for all n and take the diagonal sequence $\{u_{n_k}^k\}$.

It is obvious that $\{u_{n_k}^k\}$ is a subsequence of $\{u_{n_k}^i\}$ for any $i \geq 1$. Hence it converges uniformly to a function $u(t)$ on any bounded interval.

In the following, for simplicity, we denote the subsequence $\{u_{k_i}^j\}$ also by $\{u_k\}$.

Since u_k satisfies

$$\ddot{u}_k(t) + V_u(t, u_k(t)) = f(t), \quad (\text{II.6})$$

we conclude that $\{\ddot{u}_k\}$ and then also $\{\dot{u}_k\}$, converges uniformly on any bounded interval.

It is easy to see

$$u_k(t) = \int_{-kT}^t (t-s)\ddot{u}_k(s)ds.$$

Then $u \in C^2$ and $\ddot{u}_k \rightarrow \ddot{u}$ uniformly on any bounded interval. \square

In order to prove Theorem 1.1, we need the following result from [14].

Proposition 2.2: Let $u : \mathbb{R} \rightarrow \mathbb{R}^n$ be a continuous mapping such that $\dot{u} \in L_{loc}^2(\mathbb{R}, \mathbb{R}^n)$. Then for every $t \in \mathbb{R}$ the following inequality holds:

$$|u(t)| \leq \sqrt{2} \left(\int_{t-\frac{1}{2}}^{t+\frac{1}{2}} (|u(s)|^2 + |\dot{u}(s)|^2) ds \right)^{\frac{1}{2}} \quad (\text{II.7})$$

Proof of the Theorem 1.1 We prove u is exactly our desired homoclinic solution of (I.1).

Step 1. u satisfies (I.1).

Check the argument of Lemma 2.2, we can pass to the limit in (II.6), and then $u(t)$ satisfies (I.1). Since $f \not\equiv 0$, u is nontrivial.

Step 2. $u(t) \rightarrow 0$ as $|t| \rightarrow \infty$.

Obviously, for each $i \in \mathbb{N}$ there is $k_i \in \mathbb{N}$ such that for all $k \geq k_i$ we have

$$\int_{-iT}^{iT} (|u_k(t)|^2 + |\dot{u}_k(t)|^2) dt \leq \|u_k\|_k^2 \leq M_1^2.$$

Letting $k \rightarrow +\infty$, we obtain

$$\int_{-iT}^{iT} (|u(t)|^2 + |\dot{u}(t)|^2) dt \leq M_1^2.$$

As $i \rightarrow \infty$, we have

$$\int_{-\infty}^{+\infty} (|u(t)|^2 + |\dot{u}(t)|^2) dt \leq M_1^2.$$

Hence we get

$$\int_{|t| \geq \rho} (|u(t)|^2 + |\dot{u}(t)|^2) dt \rightarrow 0 \quad (\text{II.8})$$

as $\rho \rightarrow +\infty$.

Combing (II.7) and (II.8), we conclude $u(t) \rightarrow 0$ as $|t| \rightarrow \infty$.

Step 3. $\dot{u}(t) \rightarrow 0$ as $|t| \rightarrow \infty$.

Since $u(s) \rightarrow 0$ and $f(s) = 0$ as $|s| \rightarrow \infty$, by (A₁), we have

$$\begin{aligned} & \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} |\ddot{u}(s)|^2 ds \\ &= \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} (|V_u(s, u(s))|^2 + |f(s)|^2) ds \\ & \quad - 2 \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} (V_u(s, u(s)), f(s)) ds \rightarrow 0 \end{aligned} \quad (\text{II.9})$$

as $|t| \rightarrow \infty$.

From Proposition 2.2, we get

$$|\dot{u}(t)|^2 \leq 2 \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} (|u(s)|^2 + |\dot{u}(s)|^2) ds + 2 \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} |\ddot{u}(s)|^2 ds.$$

Together with (II.8) and (II.9), we obtain $\dot{u}(t) \rightarrow 0$ as $|t| \rightarrow \infty$. \square

REFERENCES

- [1] C.O. Alves, P.C. Carriao, O.H. Miyagaki, Existence of homoclinic orbits for asymptotically periodic systems involving Duffing-like equation, *Appl. Math. Lett.* **16** (5), 639-642 (2003).
- [2] P.C. Carriao, O.H. Miyagaki, Existence of homoclinic solutions for a class of time-dependent Hamiltonian systems, *J. Math. Anal. Appl.* **230** (1), 157-172 (1999).
- [3] E. Serra, M. Tarallo, S. Terracini, Subharmonic solutions to second-order differential equations with periodic nonlinearities, *Nonlinear Anal.* **41** (5-6), 649-667 (2000).
- [4] P. Korman, A.C. Lazer, Homoclinic orbits for a class of symmetric Hamiltonian systems, *Electron. J. Differential Equations* **1994** (1), 1-10 (1994).
- [5] S.Q. Zhang, Symmetrically homoclinic orbits for symmetric Hamiltonian systems, *J. Math. Anal. Appl.* **247** (2), 645-652 (2000).
- [6] Z.Q. Ou, C.L. Tang, Existence of homoclinic solution for the second order Hamiltonian systems, *J. Math. Anal. Appl.* **291** (1), 203-213 (2004).
- [7] P.L. Felmer, E.A. De B.e. Silva, Homoclinic and periodic orbits for Hamiltonian systems, *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (4) **26** (2), 285-301 (1998).
- [8] Y.H. Ding, Existence and multiplicity results for homoclinic solutions to a class of Hamiltonian systems, *Nonlinear Anal.* **25** (11), 1095-1113 (1995).
- [9] A. Salvatore, Homoclinic orbits for a special class of nonautonomous Hamiltonian systems, in: *Proceedings of the Second World Congress of Nonlinear Analysts, Part 8* (Athens, 1996), *Nonlinear Anal.* **30** (8), 4849-4857 (1997).
- [10] Y. H. Ding and L. Jeanjean, Homoclinic orbits for a nonperiodic Hamiltonian system, *J. Differential Equations* **237** (2), 473-490 (2007).
- [11] A. Ambrosetti, P.H. Rabinowitz, Dual variational methods in critical point theory and applications, *J. Funct. Anal.* **14**, 349-381 (1973).
- [12] P.H. Rabinowitz, K. Tanaka, Some results on connecting orbits for a class of Hamiltonian systems, *Math. Z.* **206** (3), 473-499 (1991).
- [13] P.H. Rabinowitz, Homoclinic orbits for a class of Hamiltonian systems, *Proc. Roy. Soc. Edinburgh Sect. A* **114** (1-2), 33-38 (1990).
- [14] I. Marek, J. Joanna, Homoclinic solutions for a class of second order Hamiltonian systems, *J. Differential Equations* **219** (2), 375-389 (2005).
- [15] Y.L.v. C.L. Tang, Existence of even homoclinic orbits for second-order Hamiltonian systems, *Nonlinear Anal.* **67** 2189-2198 (2007).
- [16] P. Bartolo, V. Benci, D. Fortunato, Abstract critical point theorems and applications to some nonlinear problems with "strong" resonance at infinity, *Nonlinear Anal.* **7** (9) 981-1012 (1983).

Frequency bifurcation by singularities of congruence

Roger TAULIEGNE¹

Abstract— Analysis of the synchronization of oscillations is often modelled using differential equations which describe the structure of the system. The constraints which control the solution are condensed in a system of equations, of which the resolution is often very difficult. We present an alternative approach by separation of the constraints. We consider here the impact of the only constraint of the congruence of the frequencies. It induces forms which are the much attractive for the families of solutions. These forms are deployed in a space of evolution, which we define. We analyze singularities of congruence of order 1.

In the evolution space, we propose a scenario for series of bifurcations. This proposal is connected to an experimental example, based on an oscillator of Van der Pol.

Index Terms:—Bifurcations, congruence, synchronisation, geometric constraints, Van Der Pol.

I. INTRODUCTION

THE synchronization of oscillations is used in a great number of applications. This phenomenon known since Huygens [1] [2], is accompanied sometimes by series of bifurcations of which it is very difficult to clarify the general law [3][4]. On this topic, we propose an original point of view: the separation of the constraints.

Each one in our speciality, when we build the model of a class of systems, we neglect the physical phenomena which intervene little. Seldom that calls into question its credibility, the more so as the taking into account of these mechanisms will complicate its equations largely. Thus, we select certain constraints, essential, and we let us be unaware of others of them. This practice is not new.

We propose an opposite step here. Rather than to group the constraints which interact in a class of systems, in the form of a differential equation, we let us analyze only one constraint, the constraint of congruence in the interaction of the oscillations. This unicity generalizes the remarks exposed in our article to a very vast class of systems, in fact with all the systems able to couple two oscillations, some is their nature.

The goal is to distinguish the consequences induced by the constraint from congruence of those dictated by the physical laws, whose common point is energy.

II. PRESENTATION OF CONGRUENCE SINGULARITY

A. Model and definition of congruence singularity

To model the constraint of congruence and by reasons of simplicity, we will consider only oscillations of sinusoidal

form. This choice can be considered as the prolongation of the analysis to the first harmonic, used by the pioneers of the nonlinear automatic. For the same reason we will limit ourselves to superimpose as two oscillations. Consequently our

model is simple:

$$v(t) = A \sin(\omega_a t + \phi_a) + B \sin(\omega_b t + \phi_b) \quad (1)$$

To reduce the writing, we can use the report/ratio of the amplitudes K, and to describe the difference of the phases by only one variable φ , then:

$$v(t) = K \sin(\omega_a t + \varphi) + \sin(\omega_b t) \quad (2)$$

We denote:

$$\omega_a = \frac{\omega_b}{\eta}, \text{ avec } \eta = \frac{n}{n+q}; n, q \in \mathbb{N}^* \quad \omega_a = \frac{2\pi}{T_a}; \omega_b = \frac{2\pi}{T_b} \quad (3)$$

we set the congruence

$$nT_b = (n+q)T_a = T_{\text{cycle}} \Leftrightarrow \omega_a > \omega_b$$

It guaranteed the periodicity of $v(t)$. Indeed, as long as $nT_b = (n+q)T_a$, with n and $n+q$ in \mathbb{N}^* , we will have:

$$\sin\left(\frac{2\pi}{T_b} t_0\right) = \sin\left(\frac{2\pi}{T_b} (t_0 + nT_b)\right)$$

Similarly:

$$K \cdot \sin\left(\frac{2\pi}{T_a} t_0\right) = K \cdot \sin\left(\frac{2\pi}{T_a} (t_0 + (n+q)T_a)\right)$$

So:

$$v(t_0) = v(t_0 + nT_b) = v(t_0 + (n+q)T_a)$$

In addition, n and $n+q$ imply $\eta \in \mathcal{Q}$. Naturally, only the cases for $q > 0$ are relevant. The phases space is the surface of the torus T^2 .

Because of the constraint of congruence, all the solutions are closed trajectories. To facilitate their description on a plane figure, we will consider a cross-section of the torus, normal to its large ray. The totality of the trajectory will be thus projected on the co-ordinates (v, dv) , which point out the space of the phases of a simple oscillator. For K very large, $v(t)$ is slightly modulated by $\sin(\omega_b t)$. For $K \ll 1$, $v(t)$ is close to $\sin(\omega_b t)$. These two cases have not interest, but for $K \approx 1$, the two sinusoids are in phase opposition and the existence of singular points become possible.

When the two sinusoids are in phase opposition, $v(t)$ can become null, but its derivative also. It is thus relevant to observe the evolution of $v(t)$ and $dv(t)$ according to the ratio K .

For example, let us consider the case $q=1$ and n odd (fig. 1). As long as $\varphi=0$, we will have, for $t_{\text{opp}} = T/2$

$$v\left(\frac{T_c}{2}\right) = K \cdot \text{Sin}\left(\frac{2\pi}{T_a} \frac{nT_a}{2}\right) + \text{Sin}\left(\frac{2\pi}{T_b} \frac{(n+q)T_b}{2}\right)$$

This equation is verified for all n .

¹ The author is with Conservatoire National des Arts et Métiers, Paris, France and work in team ECS of ENSEA, Cergy Pontoise, France

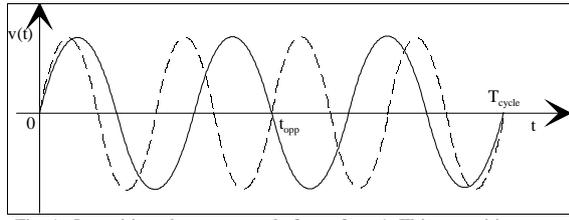


Fig. 1. Opposition phases, example for $n=3, q=1$. This opposition produces a congruence singularity.

$$v\left(\frac{T}{2}\right) = K \cdot \text{Sin}((n+1)\pi) + \text{Sin}(n\pi) = 0$$

At t_{opp} , the derivative is :

$$dv\left(\frac{T}{2}\right) = K \frac{\omega_b}{\eta} \cdot \text{Cos}((n+1)\pi) + \omega_b \cdot \text{Cos}(n\pi)$$

The signs of the circular functions are necessarily opposed, it exists a K_0 value of K which cancels this derivative:

$$K_0 = \frac{-\eta \cdot \text{cos}(n\pi)}{\text{cos}((n+1)\pi)} \Rightarrow K_0 = \frac{n}{n+q} \quad (4)$$

For this value of K , $v=dv=0$, the trajectory passes by the origin and draws a cusp (fig. 2). We will call this singularity a congruence singularity.

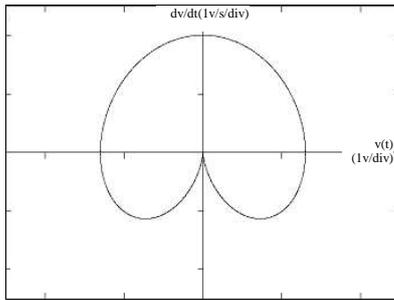


Fig. 2. The opposition phases produces a cups in the space of phases ($q=1, n=1, K=0.5$). It is a singularity of congruence.

n term of parametric functions, this singularity is a point of graining of first species.

B. Evolution space and deploymen

Each value of K defines a periodic solution. For $K > K_0$, a inner loop is formed, centred on the singularity. To observe the deployment of this form, it is necessary to trace on the same graph, a succession of solutions ordered by the values of K . One uses the reference mark (v, dv, K) in which one can follow the evolution of the orbits according to K . Consequently, we will call reference mark (v, dv, K) the space of evolution, following K . One then observes the deployment of a cone whose top is the singularity of congruence (fig. 3). We will name that the deployment of a singularity of congruence. Each solution is established with K as constant.

Remark: For $n=2$ the shape of the cone of deployment is reversed. This inversion follows that of the parity of n .

C. Main results

So that a parametric function $E(t)$ presents a turn back point of first species, it is necessary that its first non null derivative is of order pair and that the following one is odd. In our case, $E(t)$ is written as:

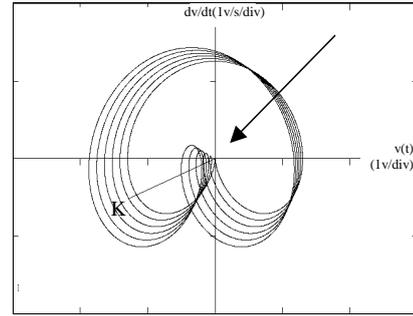


Fig. 3. Congruence singularities growing in space evolution ($q=1, n=1, K_{\text{min}}=0.5, K_{\text{max}}=0.8$).

$$E(t) = \begin{cases} v(t) \\ v'(t) \end{cases}$$

and the condition of turn back point of first species is:

$$E'(t) = 0 ; E''(t) \neq 0 \text{ et } E'''(t) \neq 0 .$$

Let t_0 and K_0 such that $v'(t_0) = 0$ et $v''(t_0) = 0 \Rightarrow E'(t_0) = 0$, then:

$$v'''(t_0) = \frac{-K}{\eta^3} \cos\left(\frac{t_0}{\eta}\right) - \cos(t_0) \neq v''(t_0) \Rightarrow v'''(t_0) \neq 0$$

Consequently: $E''(t_0) \neq 0$ et $E'''(t_0) \neq 0$. In our case, the conditions of existence of turn back point of first species are reduced to:

$$\begin{aligned} K_0 \omega_a \cos(\omega_a t_R) + \omega_b \cos(\omega_b t_R) &= 0 \\ -K_0 \omega_a^2 \sin(\omega_a t_R) + \omega_b^2 \sin(\omega_b t_R) &= 0 \end{aligned}$$

We set $\omega_b = 1$ and from the congruence constraint, we write:

$$\begin{aligned} \frac{K_0}{\eta} \cos\left(\frac{t_0}{\eta}\right) + \cos(t_0) &= 0 \\ -\frac{K_0}{\eta^2} \sin\left(\frac{t_0}{\eta}\right) - \sin(t_0) &= 0 \end{aligned}$$

To avoid rational arguments, we will pose $t_0/n = x$, from this we have:

Proposition 1: All the congruence singularities satisfy the two following conditions:

$$\frac{K_0}{\eta} \cos((n+q)x) + \cos(nx) = 0 \quad (5)$$

$$-\frac{K_0}{\eta^2} \sin((n+q)x) - \sin(nx) = 0 \quad (6)$$

Each solution of this system determines a couple (K_0, t_0) to which corresponds a congruence singularity in the space of evolution. From this proposal, we deduce the following corollary:

Corollary 1: All the singularities occur for $dv=0$.

Corollary 1 is the condition consequence of (5).

Proposition 2: No singularity exists in $t_0=0$.

Proof: The condition (5) requires that the value of each circular function is of opposite sign. For $x=0$ this condition is not satisfied. As $t_0/n = x \Rightarrow x=0 \Rightarrow t_0=0$.

A similar proof is possible with the relation (6).

Theorem 1: The number of singularities of congruence is equal to q .

Proof: Condition (5) may be rewritten as :

$$\frac{\cos((n+q)x)}{\sin(nx)} = -\frac{\eta}{K}$$

The term $\cos((n+q)x)$ is developed in a polynomial of order $n+q$, and the term $\cos(nx)$ results in a polynomial of order n . So the condition (5) can then be written in the form:

$$\frac{a_0 + a_1u^1 + a_2u^2 + \dots + a_{n+q}u^{n+q}}{b_0 + b_1u^1 + b_2u^2 + \dots + b_nu^n} = -\frac{\eta}{K}$$

From the euclidianne division, we have:

$$\frac{\eta}{K} + c_0 + c_1u^1 + c_2u^2 + \dots + c_qu^q = 0 \quad (7)$$

Considering the D'Alambert theorem this equation has q roots, that is to say as many singularities of congruencies. Notice that the condition (6) by the same method leads to the same result

For Example: With $q=3$ and $n=2$,

$$\frac{\cos((3+2)x)}{\cos(3x)} = -\frac{\eta}{K} \Rightarrow \frac{16u^5 - 20u^3 + 5u}{2u^2 - 1} = (8u^3 - 6u) + u$$

with $u=\cos(x)$. That leads to an equation of order 3, which with the obviousness has two symmetrical roots.

$$u(8u^2 - 5) + \frac{\eta}{K} = 0$$

Now, we have just to defer the values of the roots in the relation (6) for define the values of K_0 . It is thus almost always judicious to resort to the numerical approximation, for the determination of these values.

Note that the condition (6) by the same method leads to the same result.

Definition: Each q value defines a class of singularities of congruence. This value will be to name the order of the singularities of congruence.

Theorem 2, (Symetry theorem) If $\varphi=0$, each curve in the plan (v, dv) is symmetrical with axis dv .

Proof:

$$1) \frac{K}{\eta} \cos(x(n+q)+\varphi) + \cos(nx) \neq \frac{K}{\eta} \cos(-x(n+q)+\varphi) + \cos(-nx)$$

As $\cos(x(n+q)+\varphi) \neq \cos(-x(n+q)+\varphi)$ for φ not equal to 0, then $dv(x) = dv(-x) \quad \forall x$.

$$2) \frac{-K}{\eta} \sin(x(n+q)+\varphi) - \sin(nx) \neq \frac{K}{\eta} \sin(-x(n+q)+\varphi) + \sin(-nx)$$

As $-\sin(x(n+q)+\varphi) \neq \sin(-x(n+q)+\varphi)$ for φ not equal to 0, then $v(x) = -v(-x) \quad \forall x$.

If $\varphi=0$, $v(x)$ is an odd function and $dv(x)$ is an even function, from where this symmetry .

This property remains true in the space of evolution.

D. Trajectories

Considering corollary 1, in the space of evolution, each singularity is localised in the plan (v,K) by the couple (v_0, K_0) . As we have shown K_0 on n (i.e. (4)), and although in the particular case $q=n=1$, v_0 are always null, in general v_0 also depends on n . Therefore, for $q=cte$, the incrementing of n modifies the couple (v_0, K_0) .

Each new value of n defines a new of the same singularity order. It is noted that the continuation of singularity which results from it draws in the plan (v,K) a succession of aligned points. We will call trajectory this broken line. From theorem 1, for the congruence singularities of order q , we will have q distinct trajectories.

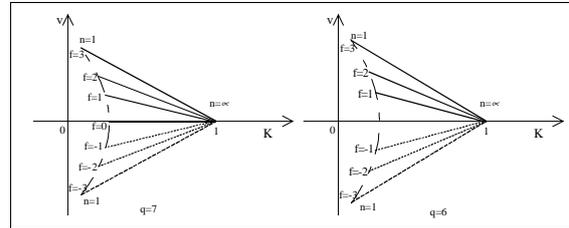


Fig. 4. Trajectories examples of the singularities of congruence, in plan (v, K) .

Définition: We will call family of singularities the whole of the singularities of congruence having a common trajectory.

This multiplicity of the singularities of congruence required the definition of a system of notation.

E. Notation of congruence singularities

It is necessary to be able to distinctly indicate all the topological singularities in the space of evolution. The characteristic of these singularities being the congruence of the oscillations, it seems natural to indicate them in a generic way by the letter C.

Theorem 1 indicates that there is q singularities of congruence for the same value of η . The number q will be called the order of the singularity and will be noted in index. In addition, for $q>1$, we will have two families of singularities, it is thus necessary to identify each family by a value of index, noted f . Because of the symmetry of singularities (Th. 2) it is relevant to sign this index of the same sign as the value of $v(t)$ corresponding to the singularity. Thus all the families of singularities to $v(t)<0$ will have a negative index. The value of f will be a relative entirety such that $f \in [0, q/2]$. For the singularities located on the axis K , $f=0$. On both sides of this axis the value of f will be increasing according to the order of appearance of the singularity. In the event of superimposed singularity, the higher cusp will have the weakest index.

Lastly, in each family a third index will number by order ascending each singularity, of the format of the notation.

$$C_{ordre(q)}(\pm famille(f), numéro(n))$$

One second quantity characterizes each singularity of congruence, the value of K which corresponds to its origin.

We will use the same indices for this second value, that is to say $K_q(\pm f, n)$.

Thus for the family of the singularities of order 1, we will write: $C_1(0, n)$; $K_1(0, n)$

III. ORDER 1 CONGRUENCE SINGULARITIES

As we showed, the moment to which occurs the singularities of order 1 is $T_{\text{cycle}}/2$. At this moment the value of K must be:

$$K_1(0, n) = \frac{n}{n+1}$$

Thus various values of K are possible the table below gives

n	$K_1(1, n)$
1	0,5
2	0,6666
3	0,75
4	0,8
5	0,8333

some examples:

When n increases the interval between two origins of singularities decrease (see fig 1 and fig. 5). This remark is probably an answer to the question: why do one more frequently observe of frequency bifurcations of a weak nature than of order high?

A. Cone of deployment

As figure 3 suggests it, the deployment of the singularity is not a cone with the direction strict, but only with the topological direction. However, there is a place of the similar singularity to a generator is easily identifiable, because he corresponds to $T_{\text{cycle}}/2$. Knowing this moment, it is easy to define the amplitude, according to the evolution of K , i.e. during the deployment of this singularity.

$$(n+1).T_b = n.T_a = T_{\text{cycle}}$$

$$v(T_{\text{cycle}}/2) = v\left(\frac{(n+1).T_a}{2}\right) = V((n+1)\pi) \text{ avec } T_a = 2.\pi$$

$$v((n+1)\pi) = K \sin((n+1)\pi) + \sin\left(\frac{n(n+1)\pi}{n+1}\right) = 0$$

This expression is always null, which shows that the point of the curve that one considers is in the plan (dv, K) . At the moment $T_{\text{cycle}}/2$, dv has for expression:

$$dv((n+1)\pi) = K \cos((n+1)\pi) + \frac{n}{n+1} \cos(n\pi)$$

The parameters n and q being fixed, this value of amplitude is a linear function of K . Because of symmetry (Th. 2), it represents a significant measurement of the ray of the cone of deployment of the singularity. It defines a generator of this cone corresponding to its intersection with the plan (v, K) , but its derivative compared to K has a constant value.

$$\frac{\partial v((n+1)\pi)}{\partial K} = \cos((n+1)\pi)$$

That shows that this generator is a line. Its slope in the plan (dv, K) is constant modulates some, therefore whatever the

value of N the slope is equal to 1, in module. All the cones of this family of singularities have even angle in the center. According to the parity of n , the sign of the slope is positive or negative. It is then possible to illustrate the overlap of the deployments of a family of singularities (fig. 5).

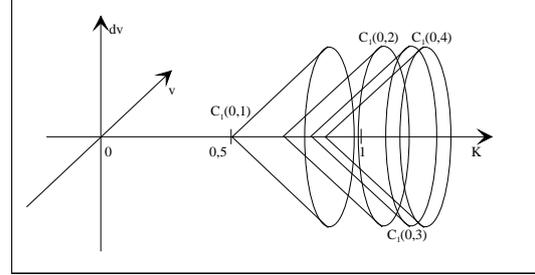


Fig. 5. Simplified representation of congruence singularities and their expansion in the evolution space.

Thus, the constraint of congruence defines a continuation of form in the space of evolution, forms which dictate a part of the behaviours of oscillating systems physique.

IV. SCENARIOS OF FREQUENCY BIFURCATION

A. Situation

The congruence of the pulsations, necessary to all synchronization, is a condition likely to lead to credible justification of the frequency bifurcation, to see to play a part of tool of design of system.

The goal is to go beyond a criterion of junction of a solution, towards a tool, near to the physical phenomena, which will clarify the total mechanism of the bifurcations, according to the structure of the system.

B. Bifurcation's assumption

In a general way this modelling of congruence cannot describe the junctions which are born from limitations of physical nature, although coupled with the constraint of congruence. But in order to outline a scenario of junctions in frequency, we will suppose:

- 1) That the physical constraints limit the size of the secondary loop to a fixed value, noted E_{seuil} .
- 2) When this limit is reached, by increase in K , the report/ratio of congruence changes.

For example, N is incremented, which authorizes the respect of congruence in spite of the growth of K

In measurement or one identifies axis v with a physical size, such as distance, tension, running, etc... and dv with derived from this size, any ray in plan (v, dv) is homogeneous with a value of energy. The threshold of bifurcation which is of energy nature, can be clarified in the plan (v, dv) . By simple graphic facility, we choose a value of dv like description of the energy threshold.

The two stated conditions are enough to model the junctions according to the growth of K (fig.6). Indeed, beyond the energy limit, the constraint of congruence is not respected any more, but our model shows that there exists, for this same value of K_{lim} another solution which restores congruence.

It is realistic to assert that, like any oscillating system, the

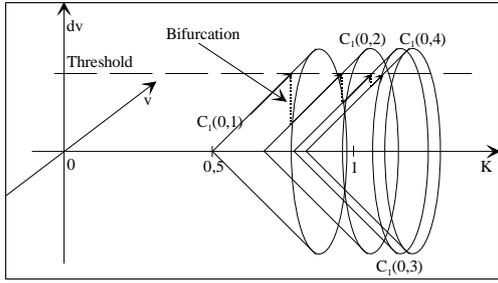


Fig. 6. Bifurcation scenarios due to K growing. Growth of K drives the trajectory on the cone of deployment of the singularity of congruence. It is geometric pressure. Arrived at the energy of threshold, resolution loses its stability, there is bifurcation.

system will tend to maximize the energy which it stores in its structure and which the periodic solution will preferably satisfy this constraint with other solutions, which they would dissipate more energy. Singularity of congruence

V. EXPERIMENTAL EXAMPLE

We propose like example an oscillator of Van Der Pol. (VDP) [3] forced by a generator of sinusoidal current (fig. 7) [5] [6].

$$\frac{d^2v}{dt^2} + \frac{g(v)}{C} \frac{dv}{dt} + \omega_0^2 v(t) = \frac{I_1 \omega_1}{C} \cos(\omega_1 t + \theta_1) \quad (8)$$

The value of E_{th} is mainly given by discontinuities of the non linear conductance (NLC), noted $g(v)$ (fig. 8). These limits being tensions, they correspond to axis v .

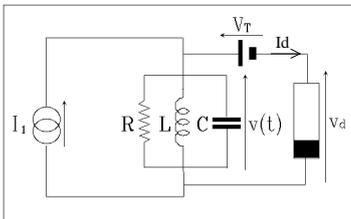


Fig. 7. Schemas of experimental circuit, with : $L=10\text{mH}$, $C=22\text{nF}$, $R=1250\Omega$, $V_T=0.4$, $g_+=0.2\text{mS}$, $g_-=-4.2\text{mS}$.

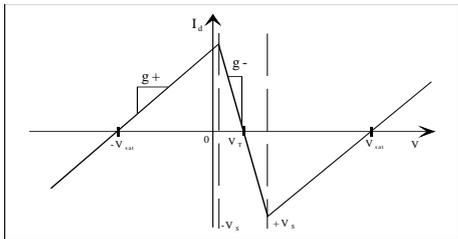


Fig. 8. Characteristic (I_d, V_d) of Non Linear Conductance (NLC).

The space of the phases is the T^2 torus, of which we will observe a section (v, dv). The order of the systems is the vector (I_1, ω_1) , to which tries to synchronize the assembly of VDP. The energy threshold E_{seuil} is translated in the space of evolution like a plan parallel with (dv, K) , this at first approximation. In any rigour, it is not a question of a plan but of a surface which has a variable curve locally.

For our analysis, we set $f_1=14\text{KHz}$ ($\omega_1=2\pi f_1$). The proper

frequency of the VDP is $f_0=10,7\text{KHz}$ ($\omega_0=2\pi f_0$), what leads to a ratio of frequencies:

$$\frac{10,7\text{KHz}}{14\text{KHz}} = 0,764$$

We are in an intermediate situation between $n=3$ and $n=4$, $q=1$.

$$\frac{n}{n+1} = \frac{3}{4} = 0,75 < \frac{f_0}{f_1} = 0,764 < \frac{4}{5} = 0,8$$

In accord with the constraint of congruence, one notes a singularity $C_1(0,3)$, whose shift compared to the origin is the consequence of a value of $\varphi \neq 0$ (fig. 9).

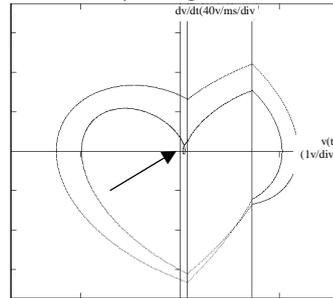


Fig. 9. Trajectory in the evolution space for K fixed. The small loop, in the zone with negative conductance, grows when I_1 augments, she leads then to an important increase of energy in the system.

The increase of I_1 , therefore the energy of one of the swing, unfolds singularity in the zone with negative conductance.

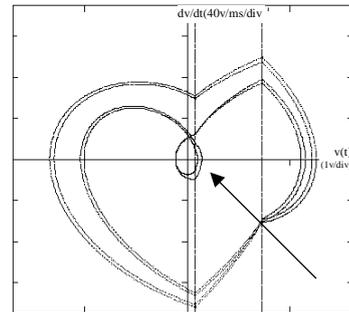


Fig. 10. Bifurcation by increasing value of I_1 . The increase of injected energy causes a doubling of cycle.

This generator of current of VDP produces a synchronous pumping of swing, of or one fork by doubling of cycle.

From the energetic argument, a very weak variation of the value of I_1 causes this junction. More precisely, this very weak variation induces increases appreciably larger in the energy injected by the NLC, defines by:

$$E_{inj} = \int_{v_{s1}(t)}^{v_2(t)} v_d(t) \cdot i_d(t) \cdot dt = g_- \int_{v_{s1}(t)}^{v_2(t)} v_d^2(t) \cdot dt$$

The analysis of this energy, developed in [7], shows that according to their initial conditions (θ, dv_s) , the segments of the trajectory which traverse the negative zone of the CNL, inject :

- 1) Very little energy (zone B, fig.11)
- 2) Definitely more in zone A, which besides presents much more a saddle node
- 3) But for certain segment which with L limiting of these two zones form a borderline of energy.

This strong energy gradient causes the instability of the solutions passing by its coordinates. That justifies the concept of energy threshold evoked in our scenario of bifurcation.

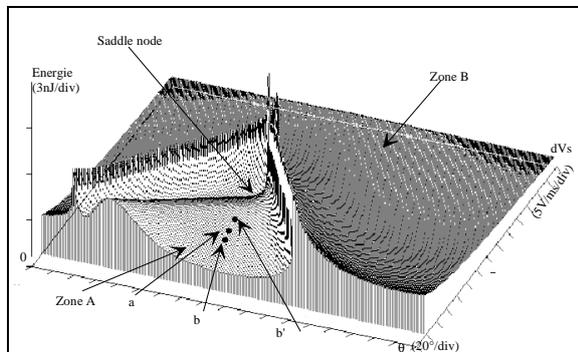


Fig. 11. Energy borderline. Injected energy by the NLC.

In zone A, point a indicates the initial conditions of the segment pertaining to $C_1(0,3)$. After the bifurcation, this segment is duplicated (point b and b'), and their difference in energy characterises the pumping effect.

The continuation of the increase in I_1 , causes the crossing of the borderline of energy of the points b and b', which amalgamates again. This new bifurcation leads to a division by 4 of f_1 (fig. 12).

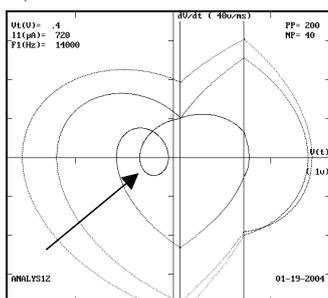


Fig. 12. Frequency bifurcation by 4.

We check on this figure the decentring of the inner loop, which corresponds to the role of ϕ , in the adjustment of the constraints.

VI. CONCLUSION

The constraint of congruence imposes in the space of evolution of the geometrical forms, which we clarified. In the case of the electric circuits, the choice of an assembly imposes its topography, and thus the advance of energy. The choice of the type of each dipole (resistance, coil, capacity), then of their values defines how energy will circulate, as well as the speed of the exchanges. These choices are as many superimposed constraints, to obtain the desired function.

Obtaining an oscillation imposes an energy balance no one on a cycle. The choice of the value of the components of an assembly determines the quantities of energy being able to be stored and/or dissipated. In the space of the phases, these constraints prohibited certain trajectories, and others select some.

The synchronization of two oscillations imposes the

constraint of congruence. The respect of the whole of the constraints is appreciably easier if there is an additional degree of freedom.

Thus, to satisfy the energy constraint and the constraint of congruence, the difference of phase between oscillations is this degree of freedom. We showed that for $\phi \neq 0$, the geometry of the singularities of congruence changes, but their topology remains.

The bifurcation scenario proposed, built an explanation of the reduction in the interval between two junctions when K increases. The detailed analysis of the oscillator of Van Der Pol., in forced mode, testifies for this physical example to the validity to this approach. A nearly constant energy threshold starts, by an identical mechanism, a continuation of bifurcation according to K. It seems to also explain why lowest congruence, appear more frequently than divisions of frequency of a very high nature.

From the point of view of the ordering of such a system, it is plausible to choose the order of the singularity of congruence and to determine the vector of order (K, ω_k) which will control the second oscillator.

It appears that the separation of the constraints offers another reading of the mechanisms of junction and their scheduling. It is an alternative to the approaches of the differential geometry, probably nearer to physics, which could facilitate its use.

Finally, it seems obvious to us that the study of the singularities of congruence of a nature higher than 1 should assemble the richness of the forms induce by this constraint. The duality synchronization bifurcation seems to justify the export of some of these results to the chaotic behaviour. Very generic properties will be can be exhibit, at least such is the objective of our next article.

VII. RÉFÉRENCES

- [1] M. Bennett, M.F. Schatz, H. Rockwood and K. Wiesenfeld, " Huygens' clocks ", 10.1098/rspa.2001.0888, pp. 563-579 .
- [2] R. Adler, "A study of locking phenomena in oscillators" Proceedings of the I.R.E. and waves electrons, June, 1946, pp351-357.
- [3] Van Der Pol, and Van Der Mark, "Frequency demultiplication " Nature 120, PP 363-364, 1927.
- [4] C. Knudsen, J. Sturis & J.S. Thomsen "Generic bifurcation structures of Arnol'd tongues in forced oscillators, Phys. Rev. A, 1991 Vol 44, N°6, 3503-3510.
- [5] MINORSKY « Nonlinear oscillations », Krieger Publishing Company, USA, 1983.
- [6] F. Kaiser and C. Eichwald "Bifurcation structure of a driven multi-limit-cycle Vand Der Pol oscillator", Int. jour. of bifurcations and chaos, Vol1, n°2, (1991) 485-491.
- [7] R. Tauleigne, " Une approche énergétique des bifurcations dans les résonateurs électrique non linéaires ", Thèse doctorat, CNAM, 1994.

Calculation of the largest Lyapunov exponent for a particular hybrid dynamical systems class

Céline Quémard

Index Terms—Chaos, hybrid dynamical systems, Lyapunov exponent, switchings, thermal application.

Extended abstract

Lyapunov exponents measure the exponential convergence or divergence rate of two trajectories with nearby initial conditions. A system of dimension N , $N \geq 1$, has N Lyapunov exponents and each measures the convergence or divergence rate according to one of system axes [1], [2], [3], [8].

Calculation of the largest Lyapunov exponent is the most interesting to realize. Indeed, a strictly positive value for this quantity generally implies chaos existence [1], [4], [9].

In this paper, we study a particular class of hybrid dynamical systems with autonomous switchings, these switchings being generated by a hysteresis phenomenon [5]. This class analysis presents an interest insofar as it admits a large number of industrial applications (in thermal, electronics, automotive...) and in any dimension. Moreover, despite its relatively simple mathematical model, complex behaviors can appear like period-doubling bifurcations [6], [7] and chaotic behaviors from dimension three systems.

It is to illustrate existence of chaotic behaviors that we will calculate the largest Lyapunov exponent considering a thermal application : a thermostat with an anticipative resistance controlling a convector located in the same room. The difficulty of this calculation occurs because of the existence of switchings that cause discontinuities in the system to solve.

In [3], a comparison of three methods :

- time series method using suites of points obtained by experimental measurements,
- difference method calculating two trajectories by numerical integration,
- variational method taking into account a small perturbation for an initial trajectory

is given and concludes that the variational method is the most suited to systems with discontinuities.

That is why, in this paper, first, we present this method giving the Müller [4] algorithm which permits calculate the largest Lyapunov exponent. Then, we apply all theoretical results to the thermostat with an anticipative resistance system and we give some numerical simulations to illustrate this study.

REFERENCES

- [1] G.-L. Baker and J.-P. Gollub, *Chaotic Dynamics: An introduction*, 2nd ed. Cambridge University Press, 1990.
- [2] T. M. Janaki and G. Rangarajan, *Computation of the Lyapunov spectrum for continuous-time dynamical systems and discrete maps*, *Physical Review*, E 60 (1999), p6614-6626.
- [3] Y. H. Lim and D. C. Hamill, *Problems of Computing Lyapunov Exponents in Power Electronics*, International Symposium on Circuits and Systems, Orlando FL 5 (1999), p297-301.
- [4] P. C. Müller, *Calculation of Lyapunov Exponents for Dynamic Systems with Discontinuities*, *Chaos, Solitons and Fractals* 5 (1995), p1671-1681.
- [5] C. Quémard, J.-C. Jolly and J.-L. Ferrier *Search for Cycles in Piecewise Linear Hybrid Dynamical Systems with Autonomous Switchings. Application to a Thermal Process*, IMACS'05 World Congress, Paris (France), 2005.
- [6] C. Quémard and J.-C. Jolly *Bifurcation doublement de période pour une classe particulière de systèmes dynamiques hybrides à commutations autonomes*, Congrès SMAI, Praz-sur-Arly (France), 2007.
- [7] C. Quémard *Analysis of a Class of Hybrid Dynamical Systems with Hysteresis Phenomenon*, NumAn (Numerical Analysis), Kalamata (Greece), 2007.
- [8] J. A. Yorke, T. D. Sauer and K. T. Alligood, *Chaos: An Introduction to Dynamical Systems*, Springer, 1997.
- [9] Z. Zhushabaliyev and E. Mosekilde, *Bifurcations and Chaos in Piecewise-Smooth Dynamical Systems*, *Nonlinear Science*, vol. 44, World Scientific, 2003.

C. Quémard is with the Department of Mathematics, ESIEE-Amiens, 14 quai de la Somme - BP10100, 80082 Amiens Cedex 2, France e-mail: quemard@esiee-amiens.fr.

Manuscript received March 13, 2009.

Forced synchronization of coupled hyperchaotic oscillators: Application to an authentication process

Laurent Laval

Abstract—This paper investigates the forced synchronization of a ring network of (controlled) hyperchaotic oscillators, to set up an authentication process for secure communication or filtered access purposes. In particular, we consider a set of three continuous-time, hyperchaotic circuits of Matsumoto-Chua-Kobayashi [1], with unidirectional, partial-state coupling¹ for synchronization purpose. Regarding to this context, one hyperchaotic (sub)system represents the “claimed identity” to be authenticated, while the two others can be viewed as locking/unlocking devices. The proposed authentication protocol then consists in performing a kind of tripartite *handshaking procedure*², by means of synchronizing some state trajectories of interest (the authentication being validated if the synchronization is achieved). Thus, after briefly describing the authentication process, this paper focuses on transmitted states between each pair of nodes, with regard to confidentiality preserving criteria and the synchronization objective. Then we address the (forced) synchronization problem, by considering both sliding mode controllers and endogenous control laws to drive each hyperchaotic subsystem. Finally, the proposed authentication process is illustrated through examples with right and wrong “claimed identity”.

Index Terms—Chaos synchronization, Hyperchaos, Chaotic networks, Authentication.

I. INTRODUCTION

Since the seminal paper of Nijmeijer and Mareels [2], synchronization of chaotic systems with regard to a control theory viewpoint, has received an increasing interest (see, for instance, [3][4][5][6] and references therein). Such an attention was indeed motivated by potential applications of chaos synchronization to many fields (e.g. [7][8]), with a particular interest in (secure) communication (e.g. [9][10]) and cryptography (e.g. [11][12][13]). Following a somewhat different purpose, this paper investigates the use of controlled synchronization of hyperchaotic systems as an authentication process. More precisely, we propose to consider the parameters and the mathematical structure of a hyperchaotic system as the signature of a “claimed identity” to be authenticated by means of an appropriate checking process. In this purpose, as suggested in [14], we investigate a kind of tripartite *handshaking procedure* based on the (controlled) synchronization properties of a ring network of unidirectionally coupled, hyperchaotic subsystems, including the claimed identity as one network node. With respect to such a context (involving some chosen state coupling conditions), the authentication is then validated if some selected state trajectories synchronize.

Manuscript received April 6, 2009; revised May 19, 2009.

L. Laval is with the Laboratoire ECS-ENSEA, Cergy-Pontoise, France.

¹i.e. restricted to only one or two transmitted state(s) per network node, while the M-C-K circuit is a fourth-order system.

²As in the context of secure communication protocols.

Such a controlled synchronization is nevertheless difficult to achieve due to:

- the hyperchaotic nature of each subsystem (as the high sensitivity of hyperchaotic systems to external signals intricate the control and, therefore, the forced synchronization),
- the selected (partial-state) coupling between each pair of nodes (as this selection mainly comes from confidentiality preserving criteria rather than facilities for synchronization),

and,

- the fact that each control law (within each network node) may induce, through the coupling, some destabilizing effects on the following connected node.

According to that, this paper is organized as follows. Section II briefly describes the authentication process. Section III introduces, from a control theory viewpoint, the synchronization problem intrinsic to that process. Section IV focuses on the network configuration, including the choice of transmitted states between each pair of nodes. Then, section V states the main results of this paper, in relation with the forced synchronization of the network. Finally, in section VI, the proposed authentication process is illustrated through examples with right and wrong “claimed identity”, before to conclude in section VII.

II. AUTHENTICATION PROCESS DESCRIPTION

First, let us recall that this paper considers the parameters and the intrinsic structure of a hyperchaotic system as the signature of a claimed identity to be authenticated. To perform such an authentication, we then propose to deal with the (controlled) synchronization properties of a ring network of hyperchaotic subsystems, including the “identity claimer” as one network node (see Figure 1).

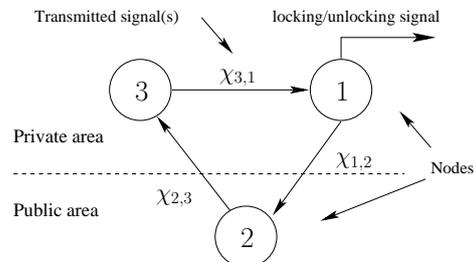


Fig. 1. Ring network of (controlled) hyperchaotic subsystems

This network is then composed of three nodes, where node 2 represents the “identity claimer”, while the two others (nodes 3

and 1) can be viewed as playing the role of locking/unlocking devices. Each node is composed of a continuous-time, hyperchaotic circuit whose behavior is driven by means of appropriate control laws acting with respect to received input and/or endogenous signals (see Figure 2).

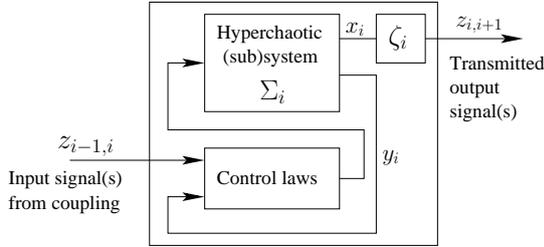


Fig. 2. Internal structure of a node N_i ($i = 1..3$)

For the authentication procedure, the three nodes are coupled, one-to-one, by chosen signals to perform a kind of tripartite "handshaking" (in the sense of a secure handshake protocol between communicating computers or electronic devices). The claimed identity is then validated if some state trajectories of interest synchronize. Contrarily, the authentication fails if these trajectories do not synchronize, as this implies that the system parameters and/or the intrinsic structure of node 2 hyperchaotic oscillator are/is wrong.

Remark 1: The consideration of a ring network (including the identity claimer) instead of only two nodes, is motivated by the objective of avoiding bidirectional communications and intruding the synchronization (to enhance, in a sense, the confidentiality and security features against possible eavesdropper).

III. PROBLEM STATEMENT

A. System description

From a control theory viewpoint, each free (i.e. uncontrolled) hyperchaotic subsystem Σ_i can be defined as an autonomous nonlinear system of the general form:

$$\Sigma_i \begin{cases} \dot{x}_i = f_i(x_i) & x_i(t_0) = x_{i0} \\ y_i = h_i(x_i) \end{cases} \quad (1)$$

where $x_i \in X_i \subseteq \mathbb{R}^n$ (for time $t \in (t_0, \infty)$), $x_{i0} \in \mathbb{R}^n$ is the initial state, and $t_0 \in \mathbb{R}$ is the initial time. $f_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a smooth vector field, $y_i \in \mathbb{R}^p$ is the output vector, and h_i (with $h_i : \mathbb{R}^n \rightarrow \mathbb{R}^p$) is a column vector of p scalar-valued smooth functions.

Remark 2: As we deal with continuous-time, hyperchaotic systems, we assume that $\dim(x_i) = n > 3$.

According to (1) and Figure 2, each node N_i can then be defined by:

$$N_i \begin{cases} \Sigma_i \begin{cases} \dot{x}_i = f_i(x_i) + \underbrace{g_i(x_i)u_i(y_i, z_{i-1,i})}_{U_i} \\ y_i = h_i(x_i) \end{cases} \\ z_{i,i+1} = \zeta_i(x_i) \\ z_{i-1,i} = \zeta_{i-1}(x_{i-1}) \end{cases} \quad x_i(t_0) = x_{i0} \quad (2)$$

where $z_{i,i+1} \in \mathbb{R}^{m_o}$ (resp. $z_{i-1,i} \in \mathbb{R}^{m_e}$) is the vector of transmitted signals for coupling between nodes N_i and N_{i+1} (resp. N_{i-1} and N_i), ζ_i (resp. ζ_{i-1}) is a column vector of m_o (resp. m_e) scalar-valued smooth functions. $g_i(x_i)$ (with $g_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$) is a smooth vector field, and $u_i(y_i, z_{i-1,i}) \in \mathbb{R}^q$ is a column vector of control laws based on output vector y_i and received signals vector $z_{i-1,i}$.

Remark 3: Here, as we consider partial-state coupling between consecutive nodes, thus:

$$\dim(z_{i,i+1}) = m_o < n \quad (3)$$

$$\dim(z_{i-1,i}) = m_e < n. \quad (4)$$

with, possibly, $m_e \neq m_o$.

Remark 4: As we consider a ring network of k nodes $N_{i=1..k}$, the subscript i is defined according to the following rules:

$$\begin{aligned} \text{if } i = 1 \text{ then } i - 1 = k \\ \text{if } i = k \text{ then } i + 1 = 1 \end{aligned} \quad (5)$$

Moreover, for convenience but without loss of generality, we consider, in this paper, the following assumptions:

Assumption A - 3.1: $y_i = C_i x_i$ with $C_i \in \mathbb{R}^{p \times n}$ (i.e. a linear state-to-output mapping).

Assumption A - 3.2: $f_i(\bullet)$, in (1), is globally Lipschitz and bounded.

Assumption A - 3.3: $z_{i-1,i} = \Gamma_i x_i$ with $\Gamma_i \in \mathbb{R}^{m_e \times n}$.

B. General problem statement

With respect to the connexion rules (5), we can define the synchronization error between two coupled nodes N_i and N_{i-1} , by:

$$e_{i,i-1} = x_i - x_{i-1} \quad (6)$$

The full-state synchronization can then be expressed as:

1) Case of exponential or asymptotical convergence:

$$\lim_{t \rightarrow \infty} \|x_i - x_{i-1}\| = 0 \quad (7)$$

2) Case of finite-time convergence:

$$\begin{aligned} \lim_{t \rightarrow T} \|x_i - x_{i-1}\| &= 0 \quad \text{with time } T \in (t_0, \infty) \\ \|x_i - x_{i-1}\| &= 0 \quad \forall t > T \end{aligned} \quad (8)$$

where $\|\bullet\|$ stands for the Euclidean norm.

In the sequel, we will also referred to as partial-state synchronization if, at least, one state of node N_i is synchronized with the corresponding state of node N_{i-1} and, only if, at least, one of the states does not synchronize.

Now, according to (2) and (6), we can define the synchronization error dynamics by:

$$\Sigma_{i,i-1}^e \begin{cases} \dot{e}_{i,i-1} = f_i(x_i) - f_{i-1}(x_{i-1}) + \underbrace{g_i(x_i)u_i(y_i, z_{i-1,i})}_{U_i} \\ \quad - \underbrace{g_{i-1}(x_{i-1})u_{i-1}(y_{i-1}, z_{i-2,i-1})}_{\Delta_i} \end{cases} \quad (9)$$

where the subscript $i-2$ follows the connexion rules given by (5).

Remark 5: With regard to (2), it is clear that transmitted signals vector $z_{i-1,i}$ (from node N_{i-1} to N_i) may directly or

indirectly contain some control components related to $u_{i-1}(\bullet)$. In other words, each control law (within each network node) may induce, through the coupling, some destabilizing effects on the following connected node. Moreover, regarding to (9), the forced synchronization of nodes N_{i-1} with N_{i-2} does intricate the synchronization of nodes N_i and N_{i-1} . Thus, the design of U_i has to result from a trade-off between acting for Σ_i and possible consequences on Σ_{i+1} . Moreover, control law U_i has to be robust enough to deal with Δ_i (coming from node N_{i-1}) as an external disturbance to be tackled.

Remark 6: As $z_{i-1,i}$ is directly related to x_{i-1} , then, by considering the synchronization error (6), control term U_i can be rewritten as:

$$U_i \equiv \bar{g}_i(e_{i,i-1})\bar{u}_i(y_i, z_{i-1,i}) + \bar{g}_i(x_i)\bar{u}_i(y_i, z_{i-1,i}) \quad (10)$$

with $\bar{g}_i(\bullet)$ and $\bar{u}_i(\bullet)$ of appropriate dimensions.

Regarding to (2), (6), (9), and connexion rules³ (5), the problem can be summarized by the following questions:

- 1) Is the system $\Sigma_{i,i-1}^e$ controllable in a neighborhood of an equilibrium point $e_{i,i-1}^o$, with, possibly, $e_{i,i-1}^o \equiv 0$?
- 2) In such a case, how to define U_i while considering remark 5 ?

Next sections investigate such questions, when dealing with the circuit of Matsumoto–Chua–Kobayashi [1] as subsystem $\Sigma_{i=1..3}$ (see Eq. (1)).

IV. NETWORK CONFIGURATION

A. The hyperchaotic system

As previously mentioned, this paper focuses on the fourth-order electronic circuit of Matsumoto–Chua–Kobayashi [1] (see Figure 3), as free hyperchaotic subsystem $\Sigma_{i=1..3}$ (see relation (1)).

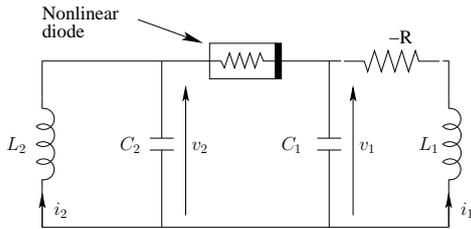


Fig. 3. The hyperchaotic circuit of Matsumoto–Chua–Kobayashi

This circuit can be defined by the following set of equations:

$$\Sigma \begin{cases} \frac{dv_1}{dt} = \frac{1}{C_1}(F(v) - i_1) \\ \frac{dv_2}{dt} = \frac{1}{C_2}(-F(v) - i_2) \\ \frac{di_1}{dt} = \frac{1}{L_1}(v_1 + Ri_1) \\ \frac{di_2}{dt} = \frac{1}{L_2}v_2 \end{cases} \quad (11)$$

³for subscript i .

where v_1 and v_2 represent the voltage across capacitors C_1 and C_2 respectively. i_1 and i_2 are the current through inductors L_1 and L_2 respectively. $F(v)$ is a piece-wise continuous function related to the (nonlinear) diode characteristics, and such that:

$$F(v) = m_0v + 0.5(m_1 - m_0)(|v + 1| - |v - 1|) \quad (12)$$

with $v = v_2 - v_1$.

By considering the following system parameters (see [1]): $C_1 = 0.5$, $C_2 = 0.05$, $L_1 = 1$ and $L_2 = 2/3$, $R = 1$, $m_0 = 3$ and $m_1 = -0.2$, this circuit then exhibits a typical hyperchaotic behavior, emphasized by computing the Lyapunov exponents (two of them are positive [1]).

Now, by posing $\alpha = \frac{1}{C_1}$, $\beta = \frac{1}{C_2}$, $\gamma = \frac{1}{L_1}$, and $\lambda = \frac{1}{L_2}$, system (11) under control can be expressed as:

$$\Sigma \begin{cases} \dot{v}_1 = \alpha(F(v_2 - v_1) - i_1) + U^{v_1} \\ \dot{v}_2 = \beta(-F(v_2 - v_1) - i_2) + U^{v_2} \\ \dot{i}_1 = \gamma(v_1 + Ri_1) + U^{i_1} \\ \dot{i}_2 = \lambda v_2 + U^{i_2} \end{cases} \quad (13)$$

where $U = [U^{v_1} \ U^{v_2} \ U^{i_1} \ U^{i_2}]^T$ is defined according to (2).

B. Transmitted signals

Now, let us deal with the coupling between two consecutive nodes (N_{i-1} and N_i), that we will referred to as transmitter and receiver respectively. Then, let us remain that when the transmitted signal (or vector) contains the nonlinear dynamics of the transmitter, this one is then globally weakly observable [15]. In such a case, there exists many possible observer design approaches which enable reconstructing the state of the transmitter. Therefore, the transmitted signals from node 1 to node 2 have to be chosen so as to disable or, at least, to seriously intricate the reconstruction of node 1 states (as part of the locking/unlocking device) by an unauthorized node 2 (such as an eavesdropper). Similarly, the transmitted signals from node 2 to node 3 have to be chosen carefully to intricate the reconstruction of node 2 states (that is the identity claimer) by a wrong authentication device (that is a falsified node 3). With regard to (13), such confidentiality preserving criteria then restrict the consideration to i_1 and/or i_2 state(s) as transmitted signal(s) for synchronization purposes (to avoid the direct transmission of $F(v_2 - v_1)$).

Thus, the question is: Is there a possible way to achieve a controlled (i.e. forced) synchronization of the ring network of hyperchaotic systems (13), with such a partial-state coupling ?

As an answer, we give some results in the next section.

V. MAIN RESULTS

First, let us start by considering a free hyperchaotic system (11) coupled with a controlled one (13), by means of transmitted signals restricted to currents i_1 and i_2 (coming from the free system).

Proposition 5.1: If two coupled systems Σ_i and Σ_j , defined respectively by (13) and (11), share the same (key) parameters $\alpha_i = \alpha_j = \alpha$, $\beta_i = \beta_j = \beta$, $\gamma_i = \gamma_j = \gamma$, $\lambda_i = \lambda_j = \lambda$, and $R_i = R_j = R$, then, with a coupling vector restricted to

states i_1 and i_2 , and for any initial conditions x_{i_0} and x_{j_0} , these systems fully synchronize by means of:

External driven sliding-mode control laws:

$$\begin{aligned} U^{i_1} &= K_1 \text{sign}(e_{i,j}^{i_1}) \\ U^{i_2} &= K_2 \text{sign}(e_{i,j}^{i_2}) \end{aligned} \quad (14)$$

and endogenous control laws:

$$\begin{aligned} U^{v_1} &= K_3 \times (e_{i,j}^{i_1} - U^{i_1} - R e_{i,j}^{i_1}) + \alpha e_{i,j}^{i_1} \\ U^{v_2} &= K_4 \times (e_{i,j}^{i_2} - U^{i_2}) + \beta e_{i,j}^{i_2} \end{aligned} \quad (15)$$

with $K_1, K_2, K_3, K_4 \in \mathbb{R}^- - \{0\}$, $e^{i_1} = i_{1\Sigma_i} - i_{1\Sigma_j}$ and $e^{i_2} = i_{2\Sigma_i} - i_{2\Sigma_j}$.

Proof: First, for readability, let us use the following notation: $x = v_1$, $y = v_2$, $z = i_1$, $w = i_2$, and $\tilde{f}_{i,i-1}(x_i, x_{i-1}) \equiv f_i(x_i) - f_{i-1}(x_{i-1})$, and replace the subscript j in Proposition (5.1) by $i-1$. Moreover, let us assume that the subsystems Σ_i and Σ_{i-1} have the same parameters: $\alpha_i = \alpha_{i-1} = \alpha$, $\beta_i = \beta_{i-1} = \beta$, $\gamma_i = \gamma_{i-1} = \gamma$, $\lambda_i = \lambda_{i-1} = \lambda$, and $R_i = R_{i-1} = R$.

Then, according to (9), (11) and (13), the synchronization error dynamics can be expressed as,

$$\dot{e}_{i,i-1}^x = \alpha \tilde{f}_{i,i-1}(x_i, x_{i-1}) - \alpha e_{i,i-1}^z + U^x \quad (16)$$

$$\dot{e}_{i,i-1}^y = -\beta \tilde{f}_{i,i-1}(x_i, x_{i-1}) - \beta e_{i,i-1}^w + U^y \quad (17)$$

$$\dot{e}_{i,i-1}^z = \gamma e_{i,i-1}^x + \gamma R e_{i,i-1}^z + U^z \quad (18)$$

$$\dot{e}_{i,i-1}^w = \lambda e_{i,i-1}^y + U^w \quad (19)$$

Regarding to (18) and (19), we can note that the synchronization errors $e_{i,i-1}^x$ and $e_{i,i-1}^y$ can be estimated from the (known) errors $e_{i,i-1}^z$ and $e_{i,i-1}^w$, their time-derivatives $\dot{e}_{i,i-1}^z$ and $\dot{e}_{i,i-1}^w$, and control laws U^z and U^w respectively.

Now, consider a Lyapunov function,

$$V = \underbrace{\frac{1}{2}(e_{i,i-1}^x)^2}_{V_1} + \underbrace{\frac{1}{2}(e_{i,i-1}^y)^2}_{V_2} + \underbrace{\frac{1}{2}(e_{i,i-1}^z)^2}_{V_3} + \underbrace{\frac{1}{2}(e_{i,i-1}^w)^2}_{V_4} > 0 \quad (20)$$

Its time-derivative is given by:

$$\dot{V} = \underbrace{\dot{e}_{i,i-1}^x e_{i,i-1}^x}_{V_1} + \underbrace{\dot{e}_{i,i-1}^y e_{i,i-1}^y}_{V_2} + \underbrace{\dot{e}_{i,i-1}^z e_{i,i-1}^z}_{V_3} + \underbrace{\dot{e}_{i,i-1}^w e_{i,i-1}^w}_{V_4} \quad (21)$$

with, explicitly,

$$\dot{V}_1 = \left[\alpha \tilde{f}_{i,i-1}(x_i, x_{i-1}) - \alpha e_{i,i-1}^z + U^x \right] e_{i,i-1}^x \quad (22)$$

$$\dot{V}_2 = \left[-\beta \tilde{f}_{i,i-1}(x_i, x_{i-1}) - \beta e_{i,i-1}^w + U^y \right] e_{i,i-1}^y \quad (23)$$

$$\dot{V}_3 = \left[\gamma e_{i,i-1}^x + \gamma R e_{i,i-1}^z + U^z \right] e_{i,i-1}^z \quad (24)$$

$$\dot{V}_4 = \left[\lambda e_{i,i-1}^y + U^w \right] e_{i,i-1}^w \quad (25)$$

Recalling that, from (14),

$$\begin{aligned} U^z &= K_1 \text{sign}(e_{i,i-1}^z) \\ U^w &= K_2 \text{sign}(e_{i,i-1}^w) \end{aligned} \quad (26)$$

and assuming that, during a sufficient time, $e_{i,i-1}^x$ and $e_{i,i-1}^y$ remain bounded by known upper bounds, then we can find some control gains $K_1 \in \mathbb{R}^{-*}$ and $K_2 \in \mathbb{R}^{-*}$, such

that \dot{V}_3 and \dot{V}_4 are negative. Therefore, from sliding-mode control theory (e.g. [16][17]), the synchronization errors $e_{i,i-1}^z$ and $e_{i,i-1}^w$ converge towards zero in finite-time T_z and T_w respectively.

Now, let us note that, after a finite-time $T \geq \max(T_z, T_w)$, (22) and (23) reduce respectively to:

$$\dot{V}_1 = \left[\alpha \tilde{f}_{i,i-1}(x_i, x_{i-1}) + U^x \right] e_{i,i-1}^x \quad (27)$$

$$\dot{V}_2 = \left[-\beta \tilde{f}_{i,i-1}(x_i, x_{i-1}) + U^y \right] e_{i,i-1}^y \quad (28)$$

Moreover, as $f_i(\bullet)$ and $f_{i-1}(\bullet)$ are globally Lipschitz (from assumption A-3.2) and assumed to be identical (in case of right claimed identity), thus, $\forall x_i \in X_i \subseteq \mathbb{R}^n, \forall x_{i-1} \in X_{i-1} \subseteq \mathbb{R}^n$

$$\|f_i(x_i) - f_{i-1}(x_{i-1})\| \leq \psi \|x_i - x_{i-1}\|, \quad \psi \geq 0 \quad (29)$$

meaning that $\tilde{f}_{i,i-1}(x_i, x_{i-1})$ is also globally Lipschitz.

Thus, assuming a known upper bound of $\|f_i(x_i) - f_{i-1}(x_{i-1})\|$ and regarding to (15), we can find some control gains $K_3 \in \mathbb{R}^{-*}$ and $K_4 \in \mathbb{R}^{-*}$ such that \dot{V}_1 and \dot{V}_2 , in (27) and (28), are negative. Therefore, $e_{i,i-1}^x$ and $e_{i,i-1}^y$ asymptotically converge towards 0. ■

Proposition 5.2: With respect to proposition 5.1, the ring network of nodes $N_{i=1..3}$, sharing the same hyperchaotic systems parameters, synchronize.

Sketch of proof When considering the ring network of controlled hyperchaotic systems, then, for two consecutive nodes, (16) (17) (18) and (19) become:

$$\dot{V}_1 = \left[\alpha \tilde{f}_{i,i-1}(x_i, x_{i-1}) - \alpha e_{i,i-1}^z - \Delta^x + U^x \right] e_{i,i-1}^x$$

$$\dot{V}_2 = \left[-\beta \tilde{f}_{i,i-1}(x_i, x_{i-1}) - \beta e_{i,i-1}^w - \Delta^y + U^y \right] e_{i,i-1}^y$$

$$\dot{V}_3 = \left[\gamma e_{i,i-1}^x + \gamma R e_{i,i-1}^z - \Delta^z + U^z \right] e_{i,i-1}^z$$

$$\dot{V}_4 = \left[\lambda e_{i,i-1}^y - \Delta^w + U^w \right] e_{i,i-1}^w$$

Assuming that $\|\Delta^x\|$, $\|\Delta^y\|$, $\|\Delta^z\|$, and $\|\Delta^w\|$ (related to control laws (14) and (15) remain bounded with known upper bounds, therefore, by considering such terms Δ^\bullet as disturbances, we can always find some control gains $K_{j=1..4}$ so that control laws (14) and (15) are robust enough to tackle these disturbances. In such a case, each node N_i can synchronize with its connected node N_{i-1} , leading finally the whole network to be synchronized. ■

Remark 7: Regarding to Proposition 5.1, if node 2 parameters or/and mathematical structure differs from those of nodes 1 and 3, then the global synchronization will fail, meaning that the claimed identity is a wrong one. Otherwise, the achieved synchronization can be viewed as a validated authentication.

VI. SIMULATION RESULTS

First, let us focus on a configuration restricted to two identical hyperchaotic subsystems Σ_1 and Σ_2 , with different initial conditions: $x_{10} = [0.2 - 0.1 - 0.01 - 0.1]^T$, $x_{20} = [0.3 - 0.15 - 0.015 - 0.15]^T$. Moreover, regarding to (14) and (15), let us consider the following control gains: $K_1 = -2$,

$K_2 = -2$, $K_3 = -10$ and $K_4 = -10$. We then obtain the simulation results of Figures 4 and 5,

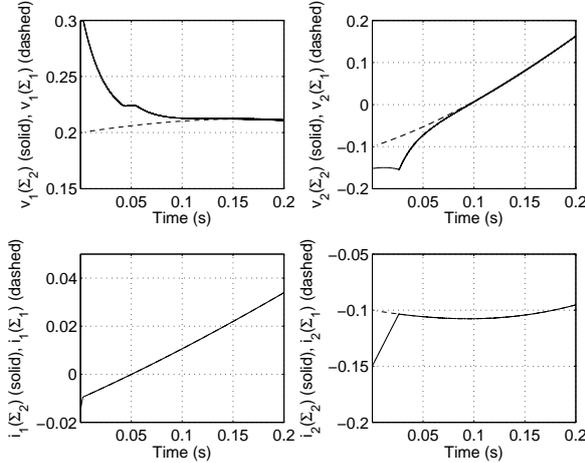


Fig. 4. Zoom on state trajectories of the two coupled, hyperchaotic systems under control

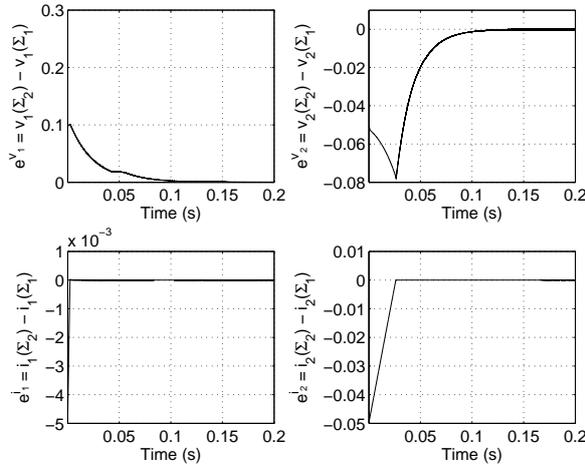


Fig. 5. Synchronization errors of the two coupled, hyperchaotic systems under control

According to Proposition 5.1, these figures then show a finite-convergence towards zero of the synchronization errors e^{i_1} and e^{i_2} , due to sliding mode control (14). Moreover, due to the endogenous control laws (15), the synchronization errors e^{v_1} and e^{v_2} asymptotically converge towards zero.

Now, by considering the whole network with partial-state coupling restricted to currents i_1 and i_2 , then with initial conditions $x_{10} = [0.4 \ -0.2 \ -0.02 \ -0.2]^T$, $x_{20} = [0.2 \ -0.1 \ 0.02 \ -0.1]^T$ and $x_{30} = [0.3 \ -0.15 \ -0.015 \ -0.15]^T$, and control gains:

	Σ_1	Σ_2	Σ_3
K_1	-0.2	-1	-0.2
K_2	-0.6	-1.4	-0.2
K_3	-3.2	-4	-4
K_4	-3.2	-4	-4

we obtain the simulation results of Figures 6–8, in case of a right claimed identity (corresponding to identical systems Σ_2 , Σ_3 and Σ_1).

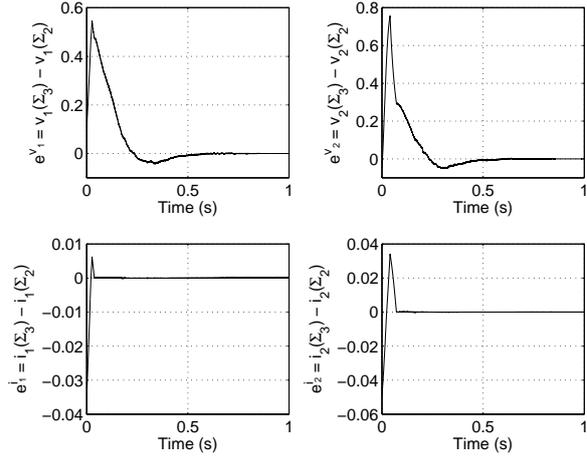


Fig. 6. Synchronization errors related to Σ_3 and Σ_2

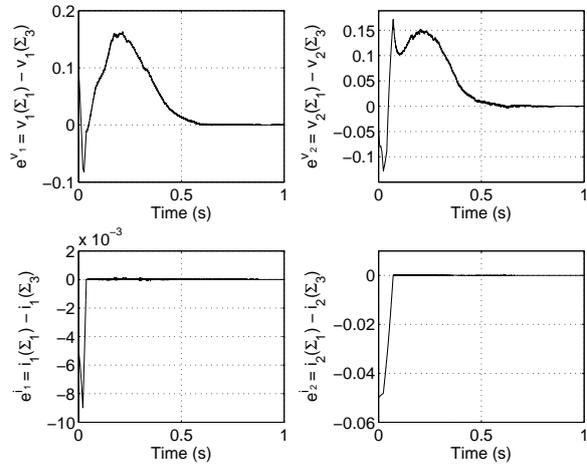


Fig. 7. Synchronization errors related to Σ_1 and Σ_3

According to Proposition 5.2 (and 5.1), these figures then show the finite-convergence or asymptotical convergence towards zero of all synchronization errors, meaning that node 2 can be viewed as a right claimed identity (by, for instance, simply inspection of synchronization errors between N_1 and N_3). Now, by slightly changing one node 2 parameter, that is to consider $1/C_1$ (of Σ_2) = 0.6 instead of 0.5, then the synchronization fails as shown through Figures 9–10, meaning that node 2 can be considered as a wrong claimed identity.

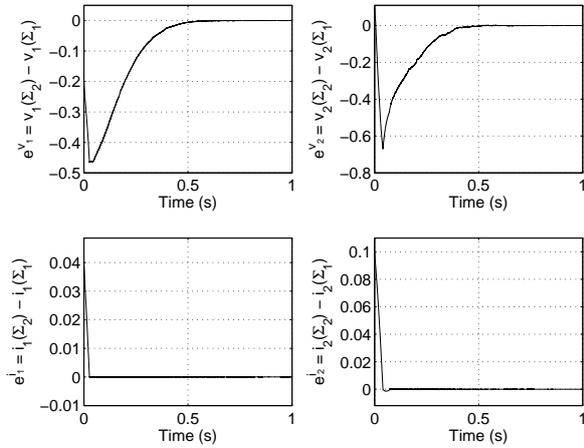


Fig. 8. Synchronization errors related to Σ_2 and Σ_1

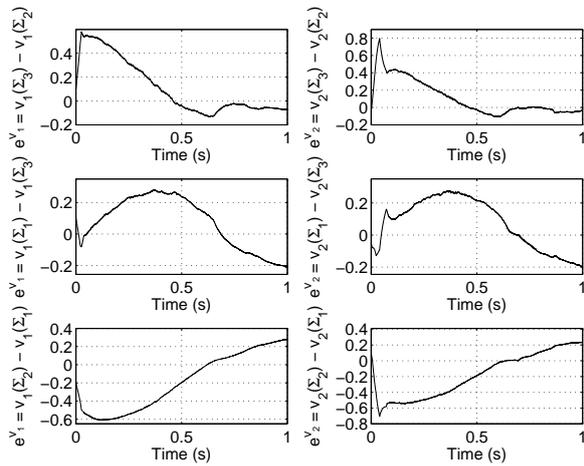


Fig. 9. Some synchronization errors (in case of wrong claimed identity)

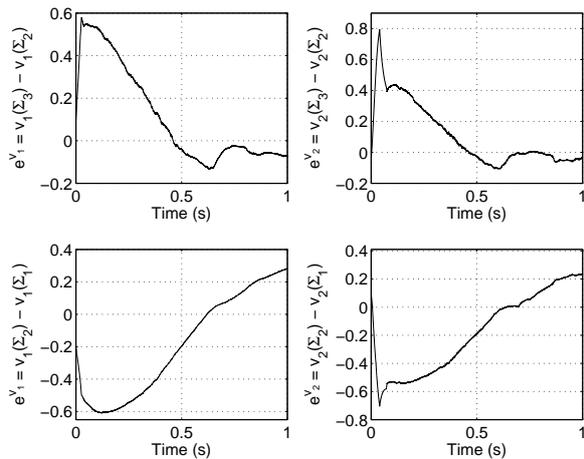


Fig. 10. Example of synchronization errors in case of wrong identity claimer (Σ_2)

VII. CONCLUSION

This paper has investigated an authentication process which considers the parameters and the mathematical structure of a hyperchaotic system as the signature of a claimed identity. With regard to this context, we have considered an authentication procedure based on the synchronization properties of a ring network of hyperchaotic systems (including the claimed identity as one network node). In particular, we have addressed the problem of designing appropriate control laws to achieve the full synchronization of the whole network, in case of right signature (and such that the synchronization fails in case of wrong signature). Future works will focus on enhancing the robustness and the convergence temporal properties of the whole set of synchronization errors, with regard to sliding-mode control theory.

REFERENCES

- [1] T. Matsumoto, L. O. Chua, and K. Kobayashi, "Hyperchaos: Laboratory experiment and numerical confirmations," *IEEE Trans. on Circuits and Systems*, vol. 33, no. 11, pp. 1143–1147, 1986.
- [2] H. Nijmeijer and I. M. Y. Mareels, "An observer look at synchronization," *Physica D*, vol. 154, pp. 219–228, 1997.
- [3] I. I. Bleckhman, A. L. Fradkov, H. Nijmeijer, and A. Y. Pogromsky, "On self-synchronization and controlled synchronization," *Syst. Cont. Lett.*, vol. 31, no. 5, pp. 299–306, 1997.
- [4] H. J. C. Huijberts, H. Nijmeijer, and R. Willems, "Regulation and controlled synchronization." Proceedings 37th IEEE Conference on Decision and Control, December 1998, pp. 3093–3098.
- [5] M. A. Aziz-Alaoui, "A survey on chaos synchronization." Tunisia: 12th IEEE-ICECS, December, 11-15 2005, pp. 523–527.
- [6] R. Femat and G. Solis-Perales, *Robust Synchronization of Chaotic Systems Via Feedback*, ser. Lecture Notes in Control and Information Sciences. Springer, 2008.
- [7] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Perseus Books Group, 2000.
- [8] —, *SYNC: The Emerging Science of Spontaneous Order*. Theia, March 2003.
- [9] T. Yang, "A survey of chaotic secure communications systems," vol. 2, no. 2, pp. 81–130, June 2004.
- [10] P. Stavroulakis, *Chaos Applications in Telecommunications*. Taylor & Francis, 2005.
- [11] G. Álvarez, F. Montoya, M. Romera, and P. G., "Chaotic cryptosystems," 1999, pp. 332–338.
- [12] F. Dachselt and W. Schwarz, "Chaos and cryptography," *IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 48, no. 12, pp. 1498–1509, December 2001.
- [13] L. Kocarev, "Chaos-based cryptography: A brief overview," *IEEE Circ. and Syst. Mag.*, vol. 48, pp. 6–21, March 2001.
- [14] L. Laval, "Authentication process based on the synchronization of chaotic oscillator networks," in *Proc. 6th International Symposium on Communication Systems, Networks and Digital Signal Processing (CNSDSP'08)*, Graz, Austria, July 2008, pp. 11–15.
- [15] R. Hermann and A. J. Krener, "Nonlinear controllability and observability," *IEEE Trans. on Automatic Control*, vol. 22, pp. 728–740, 1977.
- [16] V. I. Utkin, "Variable structure systems with sliding mode: A survey," *IEEE Trans. on Automatic Control*, vol. ACC-22, no. 2, pp. 212–222, 1977.
- [17] A. Sabanovic, L. Fridman, and S. K. Spurgeon, *Variable Structure Systems: From Principles to Implementation*. IEE Book Series, 2004.

Observers design for a new weakly coupled map function

Sébastien Hénaff*, Ina Taralova*, and René Lozi**

*IRCCyN UMR CNRS 6597, École Centrale Nantes, 1 rue de la Noë, BP 92 101, 44 321 Nantes Cedex 3, France

**Laboratoire J. A. Dieudonné, UMR CNRS 6621, Université de Nice Sophia-Antipolis, 06 108 Nice Cedex 02, France

Abstract—In some engineering applications, such as chaotic encryption, chaotic maps have to exhibit required statistical and spectral properties close to those of random signals. However, most of the papers dealing with synchronization and observer synthesis consider maps exhibiting poor statistical and spectral properties. Moreover, most of the time these properties, however essential for the chaotic encryption, are simply neglected. Unlike these papers, in our work we present the analysis of a new ultra weakly coupled maps system introduced by Lozi. The model is a deterministic one, but exhibits spectral properties (spectrum, correlation and autocorrelation) close to those of random signals, and successfully passed all the statistical tests for closeness to random signals (NIST). Two different observers have been designed. The convergence rate has been discussed in the case of affine maps, and the conditions to decrease the convergence rate by a factor of 16 have been presented, based on the locally linear behaviour of the weakly coupled map.

I. INTRODUCTION

CHAOS has recently received a growing interest in various fields of science and engineering, and in particular, in secure communications. Pecora and Carroll were the first who synchronised chaotic systems [1]. Several cryptographic schemes have been proposed since [2], [3] and can be classified in three main categories : chaotic masking, chaotic modulation and chaotic shift keying.

In the cryptographic application, the chaotic generator must exhibit appropriate features close to those of the pseudo-random generators. These adapted properties have been studied more precisely in [4], [5], [6].

Further researchers have then looked for finding appropriate systems testing different architectures : traditional chaotic maps (for example, the logistic map, the Hénon map, the generalised Hénon map) [7] [8], piece-wise linear map, cascaded map [9] or coupled map lattice. In order to evaluate the features of the system, statistical tests developed for random number generators (RNG) can also be applied to chaotic maps, in order to gather evidence that the map generates "good" chaotic signals, i.e. having a considerable degree of randomness. To address this particular problem, different statistical tests for the systematic evaluation of the randomness of cryptographic random number generators can be applied, among which the most popular NIST (National Institute of Standards and Technology) tests.

It appears that most of the maps classically used for chaotic encryption do not pass successfully these tests, and don't

exhibit the required features. However, most of the papers dealing with synchronisation and observer synthesis consider precisely these kinds of maps, highly inefficient in the context of chaotic encryption.

Unlike these models, Lozi [10] introduced in 2008 a new ultra weakly coupled maps system to generate pseudo-random signals which exhibits very good statistical properties. To use this system for secure communication, it must exhibit good spectral features and have to be observable. So the aim of this paper is to identify and to design an observer for the weakly coupled map system.

This paper is organised as follow : after briefly presenting the system under investigation, sections three and four present the issues on parameter identifiability and system observability. Sections five and six propose and compare two different observers. Finally, a concluding section ends the paper.

II. SYSTEM DEFINITION

The N-th order function F can be written as :

$$X(n+1) = F(X(n))$$

with $X(n) = (x_1(n), x_2(n), \dots, x_N(n))$

$$X(n+1) = F(X(n)) = A \Lambda(X(n))$$

where A is a $N \times N$ matrix defined by:

$$A = \begin{pmatrix} 1 - (N-1)\epsilon_1 & \epsilon_1 & \dots & \epsilon_1 \\ \epsilon_2 & 1 - (N-1)\epsilon_2 & \dots & \epsilon_2 \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_N & \epsilon_N & \dots & 1 - (N-1)\epsilon_N \end{pmatrix}$$

and Λ is the tent function applied to every the components of $X \in [-1; 1]^N$:

$$\Lambda(X(n)) = \begin{pmatrix} \Lambda(x_1(n)) \\ \Lambda(x_2(n)) \\ \vdots \\ \Lambda(x_N(n)) \end{pmatrix}$$

Since the function is piece-wise linear, it can be rewritten under a matrix form, by rewriting the tent function :

$$\Lambda(x) = \begin{cases} 2x + 1 & \text{if } x < 0 \\ -2x + 1 & \text{else} \end{cases}$$

or using the generic form :

$$\Lambda(x) = sx + 1$$

with :

$$s = \begin{cases} 2 & \text{if } x < 0 \\ -2 & \text{else} \end{cases}$$

For the second order, the general system F is then governed by :

$$\begin{pmatrix} x_1(n+1) \\ x_2(n+1) \end{pmatrix} = A_n \begin{pmatrix} x_1(n) \\ x_2(n) \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

where A_n is :

$$A_n = \begin{pmatrix} (1 - \epsilon_1)s_{10} & \epsilon_1 s_{20} \\ \epsilon_2 s_{10} & (1 - \epsilon_2)s_{20} \end{pmatrix}$$

The rest of the paper only consider the second order system.

III. IDENTIFIABILITY

The purpose of this section is to determine if the coder can generate two identical outputs from two different encryption keys. In terms of system theory, it means that the system generates two identical outputs for two different parameter combinations. If this is the case, the base of the varying parameters has to be modified, and the parameter redundancies removed. To do so, the two outputs have to be equalized and their impact on the parameters has to be investigated.

The presented study concerns the second order system without the scaling. Let consider two second order systems governed by the same law :

$$\begin{cases} x_1(n+1) = (1 - \epsilon_1)\Lambda(x_1(n)) + \epsilon_1\Lambda(x_2(n)) \\ x_2(n+1) = (1 - \epsilon_2)\Lambda(x_2(n)) + \epsilon_2\Lambda(x_1(n)) \\ y(n) = x_1(n) \end{cases}$$

$$\begin{cases} \hat{x}_1(n+1) = (1 - \hat{\epsilon}_1)\Lambda(\hat{x}_1(n)) + \hat{\epsilon}_1\Lambda(\hat{x}_2(n)) \\ \hat{x}_2(n+1) = (1 - \hat{\epsilon}_2)\Lambda(\hat{x}_2(n)) + \hat{\epsilon}_2\Lambda(\hat{x}_1(n)) \\ \hat{y}(n) = \hat{x}_1(n) \end{cases}$$

Considering the same outputs : $(\hat{y}(n))_n = (y(n))_n$, is it possible that the parameters would be different? The system is piece-wise linear, so let $s_{ij} \in \{-2; 2\}$ be defined by $\Lambda(x_i(n+j)) = 1 + s_{ij}$.

$$s_{ij} = \begin{cases} -2 & \text{if } x_i(n+j) > 0 \\ 2 & \text{else} \end{cases}$$

$$\hat{y}(n) = y(n) \Rightarrow \hat{x}_1(n) = x_1(n)$$

$$\begin{cases} \hat{y}(n) = y(n) \\ \hat{y}(n+1) = y(n+1) \end{cases} \Rightarrow (\hat{\epsilon}_1 - \epsilon_1)\Lambda(x_1(n)) = \hat{\epsilon}_1\Lambda(\hat{x}_2(n)) - \epsilon_1\Lambda(x_2(n))$$

$$\begin{cases} \hat{y}(n) = y(n) \\ \hat{y}(n+1) = y(n+1) \\ \hat{y}(n+2) = y(n+2) \end{cases} \Rightarrow \begin{aligned} & [(\hat{\epsilon}_1 - \epsilon_1)(1 - \epsilon_1)s_{11} - \hat{\epsilon}_1\hat{\epsilon}_2\hat{s}_{21} + \epsilon_1\epsilon_2s_{21} \\ & - (\hat{\epsilon}_1 - \epsilon_1)(1 - \hat{\epsilon}_2)\hat{s}_{21}]\Lambda(x_1(n)) \\ & = \epsilon_1[-(\hat{\epsilon}_1 - \epsilon_1)s_{11} - (1 - \epsilon_2)s_{21} \\ & + (1 - \hat{\epsilon}_2)\hat{s}_{21}]\Lambda(x_2(n)) \end{aligned}$$

Both x_1 and x_2 appear in the last. But $\{x_1; x_2\}$ is the state of the chaotic system, which has the property to visit the whole state space $[-1; 1]^2$. In other words, to a given parameter combination $\{\epsilon_1, \epsilon_2, \hat{\epsilon}_1, \hat{\epsilon}_2, s_{10}, s_{20}, s_{11}, s_{21}, \hat{s}_{10}, \hat{s}_{20}, \hat{s}_{11}, \hat{s}_{21}\}$ can be associated an infinity of states $\{x_1; x_2\}$. One can consider then the independent variables $\Lambda(x_1(n))$ et $\Lambda(x_2(n))$. In this case, one obtains the following system of equations :

$$\begin{cases} (\hat{\epsilon}_1 - \epsilon_1)(1 - \epsilon_1)s_{11} - \hat{\epsilon}_1\hat{\epsilon}_2\hat{s}_{21} + \epsilon_1\epsilon_2s_{21} \\ - (\hat{\epsilon}_1 - \epsilon_1)(1 - \hat{\epsilon}_2)\hat{s}_{21} = 0 \\ \epsilon_1[-(\hat{\epsilon}_1 - \epsilon_1)s_{11} - (1 - \epsilon_2)s_{21} + (1 - \hat{\epsilon}_2)\hat{s}_{21}] = 0 \end{cases}$$

One solution of the second equation is : $\epsilon_1 = 0$. ϵ_1 is one of the system parameters, and this solution corresponds to a decoupled system. Therefore, this particular case is to be excluded. One obtains then the new system of equations :

$$\begin{cases} (\hat{\epsilon}_1 - \epsilon_1)(1 - \epsilon_1)s_{11} - \hat{\epsilon}_1\hat{\epsilon}_2\hat{s}_{21} + \epsilon_1\epsilon_2s_{21} \\ - (\hat{\epsilon}_1 - \epsilon_1)(1 - \hat{\epsilon}_2)\hat{s}_{21} = 0 \\ -(\hat{\epsilon}_1 - \epsilon_1)s_{11} - (1 - \epsilon_2)s_{21} + (1 - \hat{\epsilon}_2)\hat{s}_{21} = 0 \end{cases}$$

The resolution leads to the following result:

$$\forall (s_{11}, s_{21}, \hat{s}_{21}) \in \{-2; 2\}^3, \begin{cases} s_{21} = \hat{s}_{21} \Rightarrow \{\hat{\epsilon}_1, \hat{\epsilon}_2\} = \{\epsilon_1, \epsilon_2\} \\ s_{11} = s_{21} = -\hat{s}_{21} \Rightarrow \epsilon_1 = 0 \text{ et } \hat{\epsilon}_2 + \epsilon_2 - \hat{\epsilon}_1 = 0 \\ -s_{11} = s_{21} = -\hat{s}_{21} \Rightarrow \epsilon_1 = 0 \text{ et } \hat{\epsilon}_2 + \epsilon_2 + \hat{\epsilon}_1 = 0 \end{cases}$$

Knowing that the solution $\epsilon_1 = 0$ is impossible, then the following conclusion can be drawn :

$$\begin{cases} \hat{y}(n) = y(n) \\ \hat{y}(n+1) = y(n+1) \\ \hat{y}(n+2) = y(n+2) \\ \epsilon_1 \neq 0 \end{cases} \Rightarrow \{\hat{\epsilon}_1, \hat{\epsilon}_2\} = \{\epsilon_1, \epsilon_2\}$$

Finally, there are no redundant parameters and the whole set of parameter combinations can be used as a set of encryption keys of the coder, there are no parameters different from the one used for the encryption which could allow to decrypt the message.

IV. OBSERVABILITY

An affine system can be written as :

$$\begin{cases} x(n+1) = F(x(n)) = A.x(n) + B \\ y(n) = Cx(n) \end{cases}$$

A second order affine system is observable if its observability matrix is a full-rank one :

$$O = \begin{pmatrix} C \\ CA \end{pmatrix}$$

Here, the system is piece-wise affine, therefore the observability matrix shall be different according to the region to which belongs the system state. It is equal to :

$$O = \begin{pmatrix} 1 & 0 \\ 2(1 - \epsilon_1)s_{10} & 2\epsilon_1s_{10} \end{pmatrix}$$

which is full-rank since $\epsilon_1 > 0$. Therefore, the system is observable.

V. LINEAR LUENBERGER OBSERVER

The system is piece-wise affine. Considering it as such, the present section identifies a piece-wise linear observer. The second order system can be rewritten using the affine form on the four domains where it is defined :

$$\begin{cases} x(n+1) = F(x(n)) = A.x(n) + B \\ y(n) = Cx(n) \end{cases}$$

$$\begin{cases} x(n+1) = \begin{pmatrix} (1-\epsilon_1)s_{10} & \epsilon_1 s_{20} \\ \epsilon_2 s_{10} & (1-\epsilon_2)s_{20} \end{pmatrix} x(n) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ y(n) = \begin{pmatrix} 1 & 0 \end{pmatrix} x(n) \end{cases}$$

The associated Luenberger system is :

$$\hat{x}(n+1) = \hat{A}\hat{x}(n) + B + K(C\hat{x}(n) - y(n))$$

K is a predefined gain such that the error $e(n)$ tends to zero. Let consider $\hat{x}(n)$ and $x(n)$ in the same region of definition. In this case, $\hat{A} = A$ and therefore,

$$e(n+1) = (A + KC)e(n)$$

One can identify the values of the gain K which cancel the eigenvalues of the matrix $(A + KC)$ as a function of the affine system model. In this case, since the matrix is of second order, $(A + KC)^2 = 0$ therefore if the system states x and its estimate \hat{x} belong to the same region of the state space twice consecutively, then the estimate shall synchronise with the original system.

Zero eigenvalues lead to the following solutions for the gain :

$$K = \begin{cases} \begin{pmatrix} 2(\epsilon_1 + \epsilon_2 - 2) \\ \frac{2}{\epsilon_1}(2\epsilon_2 - \epsilon_2 2 - \epsilon_1 \epsilon_2 - 1) \end{pmatrix} & \text{if } \hat{x}(n) \in [-1; 0]^2 \\ \begin{pmatrix} 2(\epsilon_2 - \epsilon_1) \\ \frac{2}{\epsilon_1}(2\epsilon_2 - \epsilon_2 2 + \epsilon_1 \epsilon_2 - 1) \end{pmatrix} & \text{if } \hat{x}(n) \in [0; 1] \times [-1; 0] \\ \begin{pmatrix} -2(\epsilon_2 - \epsilon_1) \\ -\frac{2}{\epsilon_1}(2\epsilon_2 - \epsilon_2 2 + \epsilon_1 \epsilon_2 - 1) \end{pmatrix} & \text{if } \hat{x}(n) \in [-1; 0] \times [0; 1] \\ \begin{pmatrix} -2(\epsilon_1 + \epsilon_2 - 2) \\ -\frac{2}{\epsilon_1}(2\epsilon_2 - \epsilon_2 2 - \epsilon_1 \epsilon_2 - 1) \end{pmatrix} & \text{if } \hat{x}(n) \in [0; 1]^2 \end{cases}$$

The zero eigenvalues assure the convergence in two iterations of the affine system if the system states remain in the same region of definition. Then the synchronisation may not take place for any states evolution.

The error of the linear system evolves following the equation :

$$e(n+1) = (A + KC)e(n)$$

Since the matrix $(A + KC)$ is nilpotent, if the system remains in the same domain of definition,

$$e(n+2) = (A + KC)^2 e(n) = 0$$

In reality, the system states have a probability of 1/4 to fall twice consecutively in the same domain of definition. Considering that both systems (the original one, and the observer) start from the same region, then statistically three iterations are necessary before the trajectories converge. When the system falls consecutively into two different configurations, the equation which governs the error becomes :

$$e(n+2) = (A_1 + K_1 C)(A_2 + K_2 C)e(n)$$

Let P_1, P_2 be two transformation matrices which triangularise respectively the matrices $(A_1 + K_1 C)$ and $(A_2 + K_2 C)$, and let D_1, D_2 be the two triangularised matrices. It comes :

$$e(n+2) = P_1 D_1 P_1^{-1} P_2 D_2 P_2^{-1} e(n)$$

As soon as $P_1 \neq P_2$, the error e does not cancel in two iterations.

Now, the proper bases of the matrices $(A + KC)$ are the same for the domains of definition $\hat{x}(n) \in [-1; 0]^2$ and $\hat{x}(n) \in [0; 1]^2$. On the other hand, the bases are the same for the domains of definition $\hat{x}(n) \in [0; 1] \times [-1; 0]$ and $\hat{x}(n) \in [-1; 0] \times [0; 1]$. In the exemple, since the matrices D_1 and D_2 have zero eigenvalues, if $P_1 = P_2$,

$$e(n+2) = P_1 D_1 D_2 P_2^{-1} e(n) = 0$$

Finally, considering that the two systems are in the same domains of definition, they have one chance out of two to synchronise.

Now, if one considers that the transition evolution of the two systems is independent in the domain of definition until they synchronise, they have statistically one chance out of sixteen to fall twice consecutively in the same domains of definition, which decreases the probability to synchronise at a given instant to 1/32.

Finally, two synchronisation strategies are possible : the classical one considers that the master system starts from any initial condition and follows the same law during the synchronisation. In this case, the slave system will synchronise - in average - after 32 iterations and it is governed by the equation :

$$\hat{x}(n+1) = F(\hat{x}(n)) + B + K(C\hat{x}(n) - y(n))$$

On the other hand, one can consider that the observer consists of several systems following different laws, each following its own law whatever the value of its state at the next iterates. A system can then be governed by the law :

$$S1 : \hat{x}(n+1) = \hat{A}_1 \hat{x}(n) + B + K_1 (C\hat{x}(n) - y(n))$$

where A_1 et K_1 are derived from the definition of the systems related to the desired domain of definition, $\hat{x}(n) \in [-1; 0]^2$ for instance. The observer systems have to cover the whole set of possible combinations of the state evolutions which allow to synchronise, i.e. four observers for a second order system. The advantage to use these systems lies in the fact that the probability that one of the forth systems synchronises with the original systems rises up to 1/2. Once synchronised, a classical observer can allow to follow the trajectory of the states of the original system.

If the classical use of a second order system leads to a synchronisation in 32 iterations in average, when the system order is increased, the synchronisation time increases exponentially. The simultaneous use of several observers allows to divide the time for synchronisation by 16 for a second order system. The drawback is that several observer systems have to run simultaneously.

VI. DEAD BIT OBSERVER

A second observer can be designed based on the inverse lag. It allows to identify the current states by considering the inverse function. For the second order system, the autonomous system is :

$$\begin{cases} x_1(n+1) = (1 - \epsilon_1)\Lambda(x_1(n)) + \epsilon_1\Lambda(x_2(n)) \\ x_2(n+1) = (1 - \epsilon_2)\Lambda(x_2(n)) + \epsilon_2\Lambda(x_1(n)) \\ y(n) = x_1(n) \end{cases}$$

With two measurements at the output y , it is possible to reconstruct the signal :

$$\begin{cases} \chi(n+1) = y(n) \\ \hat{z}_1(n) = y(n) \\ \hat{z}_2(n) = \frac{1-\epsilon_2}{\epsilon_1}y(n) + (\epsilon_2 - \frac{(1-\epsilon_2)(1-\epsilon_1)}{\epsilon_1})\Lambda(\chi(n)) \end{cases}$$

Finally, this reconstructor can identify the original state for all values, which is not the case of the first observer. However, this method can hardly be applied to greater order systems.

VII. CONCLUSION

Most of the papers devoted to observer synthesis considered maps with poor statistical and spectral properties. We present here the synthesis of efficient observers for the system of weakly coupled map which satisfied all statistical (NIST) and spectral analysis tests. Two different observers have been designed. The convergence rate has been discussed in the case of affine maps, and the conditions to decrease the convergence rate by a factor of 16 have been presented, based on the locally linear behaviour of the weakly coupled map. The design and analysis of higher order map observers is currently under investigation.

REFERENCES

- [1] T.L. Carroll and L.M. Pecora. Synchronizing chaotic circuits. *Circuits and Systems, IEEE Transactions on*, 38(4):453–456, Apr 1991.
- [2] L. Kocarev, J. Makraduli, and P. Amato. Public-key encryption based on chebyshev polynomials. *Circuits, systems, and signal processing*, 24(5):497–517, octobre 2005.
- [3] P. Fei, Q. Shui-Sheng, and L. Min. A secure digital signature algorithm based on elliptic curve and chaotic mappings. *Circuits, Systems, and Signal Processing*, 24(5):585–597, octobre 2005.
- [4] G. Álvarez and S. Li. Some basic cryptographic requirements for chaos-based cryptosystems. *International Journal of Bifurcation and Chaos*, 16(8):2129–2151, 2006.
- [5] M. Gotz, K. Kelber, and W. Schwarz. Discrete-time chaotic encryption systems. i. statistical design approach. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 44(10):963–970, Oct 1997.
- [6] NIST. *A Statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications*. 2001.
- [7] D.R. Frey. Chaotic digital encoding: an approach to secure communication. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, 40(10):660–666, Oct 1993.
- [8] H. Noura, S. Hénaff, I. Taralova, and S. El Assad. Efficient cascaded 1-d and 2-d chaotic generators. Second IFAC conference on analysis and control of chaotic systems, june 2009.
- [9] W.P. Tang and H.K. Kwan. Chaotic communications using nonlinear transform-pairs. In *International Symposium on Circuits and Systems*, volume 5 of *Circuits and Systems*, pages V–740–V–743. ISCAS, IEEE, May 2004.
- [10] R. Lozi. New enhanced chaotic number generators. *Indian journal of industrial and applied mathematics*, 1(1):1–23, janvier 2008.

Periodic Dynamical Systems in the Phase Space

Bruno ROSSETTO and Jean-Marc GINOUX

Toulon University Institute of Technology, BP 20132, 83957, LA GARDE Cedex (France)

rossetto@univ-tln.fr, ginoux@univ-tln.fr

Abstract—It is shown that the vector field of a multidimensional periodic autonomous dynamical systems confers to the trajectories some geometrical properties according to their location in the phase space. Different patterns can be identified, in relation with the Liapunov exponents of an associated periodic parameters linear equation (APPLE) defined at each point of the phase space, such as funneling, parametric resonance, period doubling, sensitivity to initial conditions. A method to compute the Liapunov exponents without integration is used. Then an associated constant coefficient equivalent system (ACCES) is defined and used to carry out the equation of an invariant manifold periodically crossed by the solutions.

The method is applied to bidimensional and tridimensional Chua equations with periodic coefficients.

I. PERIODIC AUTONOMOUS DYNAMICAL SYSTEMS

Consider the periodic autonomous dynamical system (PADS) defined in \mathbb{R}^n :

$$\dot{X}(t) = F(X(t), t) \quad (1)$$

where X is a n -dimensional vector of \mathbb{R}^n and the mapping $F(X(t), t) : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n, (X, t) \mapsto F$ is continuous and derivable with respect to t and to each component of X . The explicit time depending part of $F(X(t), t)$ is T -periodic with respect to t :

$$F(X, t + nT) = F(X, t) \quad \forall t \in \mathbb{R} \text{ and } n \in \mathbb{Z}. \quad (2)$$

The mapping $F(X(t), t)$ is supposed to be uniformly lipshitzian, $\forall t \in \mathbb{R}$ and $\forall X(t) \in \mathbb{R}^n$, with respect to each of the components of $X(t)$ so that a solution related to any Cauchy condition $X(t_0) = X_0$ exist and is unique.

II. ASSOCIATED PERIODIC PARAMETRIC LINEAR EQUATION

Let $\delta X(t)$ be spatial variations around $X(t)$. The locally associated periodic parametric linear equation (APPLE) is defined as

$$\delta \dot{X} = J_t(X) \delta X \quad (3)$$

where $J_t(X) : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^{n^2}, (X, t) \mapsto J_t(X)$ is the Jacobian matrix of the PADS in (X) :

$$J_t(x, y) = \frac{\partial F}{\partial X} \quad (4)$$

$J_t(x, y)$ is as well T -periodic with respect to t .

A. Liapunov's theory

According to the Liapunov theorem, the solution of APPLE is a linear combination of n modes:

$$\delta X(t) = e^{\mu_1 t} \chi_1(t) + e^{\mu_2 t} \chi_2(t) + \dots + e^{\mu_n t} \chi_n(t) \quad (5)$$

where $\mu_k \in \mathbb{C}, 1 \leq k \leq n$, are the Liapunov exponents and $\chi_k(t) : \mathbb{R} \rightarrow \mathbb{R}^n, t \mapsto \chi_k(t)$ are T-periodic mappings, continuous and derivable with respect to t. Of course, μ_k and $\chi_k(t)$ depend on X, the location of the phase space. The Liapunov coefficients are worked out using an algorithm which exponentially converges [5]. It is supposed in this work that, in some domain D of the phase space, μ_3, \dots, μ_n have a negative real part. Then, the corresponding modes vanishes and the APPLE has a *bimodal* solution related to μ_1 and μ_2 . Different patterns of the PADS can be identified, in relation with μ_1 and μ_2 , such as funneling, parametric resonance, period doubling, sensitivity to initial conditions.

B. Proposition 1: Liapunov exponents as eigenvalues of the Jacobian matrix defined on some time t_0 .

Let $\delta X(t)$ such a bimodal solution of APPLE related to the Liapunov exponent μ_1 and μ_2 :

$$\delta X(t) = e^{\mu_1 t} \chi_1(t) + e^{\mu_2 t} \chi_2(t) \quad (6)$$

For any point (x, y) of the phase plane, we consider the time t_0 , if it exists, such as the periodic part of the corresponding solutions follow the initial conditions

$$\dot{\chi}_1(t_0) = 0 \text{ and } \dot{\chi}_2(t_0) = 0 \quad (7)$$

Then each $\mu_k, 1 \leq k \leq 2$, is an eigenvalue of $J_{t_0}(X)$ and the related eigenvector is

$$\chi_k(t_0) = \chi_k(t_0 + nT) \quad (8)$$

In other words, the initial conditions of each $\chi_k(t)$ can be adjusted so as the following relationship holds

$$J_{t_0} \chi_k(t_0) = \mu_k \chi_k(t_0), 1 \leq k \leq 2 \quad (9)$$

Note that a necessary, but not sufficient, condition to satisfy the two conditions (9) is for $F(X(t), t)$ and then $J_t(X)$ to have at least two variable coefficients. They can be adjusted from the characteristic equation

$$\det(\mu_k I - J_{t_0}) = 0 \quad (10)$$

Proof. The derivative of the bimodal solution of APPLE can be written as

$$\delta \dot{X}(t) = \mu_1 e^{\mu_1 t} \chi_1(t) + \mu_2 e^{\mu_2 t} \chi_2(t) + e^{\mu_1 t} \dot{\chi}_1(t) + e^{\mu_2 t} \dot{\chi}_2(t)$$

Introducing these values in (3) and (6):

$$\begin{aligned} \mu_1 e^{\mu_1 t} \chi_1(t) + \mu_2 e^{\mu_2 t} \chi_2(t) + e^{\mu_1 t} \dot{\chi}_1(t) \\ + e^{\mu_2 t} \dot{\chi}_2(t) \\ = J_t e^{\mu_1 t} \chi_1(t) + J_t e^{\mu_2 t} \chi_2(t) \end{aligned}$$

When $t = t_0$, from the hypothesis (7), this relationship can be written as:

$$\begin{aligned} e^{\mu_1 t_0} (J_{t_0} \chi_1(t_0) - \mu_1 \chi_1(t_0)) \\ + e^{\mu_2 t_0} (J_{t_0} \chi_2(t_0) - \mu_2 \chi_2(t_0)) \\ = 0 \end{aligned}$$

This equation holds if and only if each $\mu_k, 1 \leq k \leq 2$, is an eigenvalue of J_{t_0} and $\chi_k(t_0)$ the related eigenvector.

III. ASSOCIATED CONSTANT COEFFICIENT EQUIVALENT SYSTEM

Let t_0 be the time value, if it exists, for which μ_k are eigenvalues of $J_{t_0}, 1 \leq k \leq 2$. Consider the non linear associated constant coefficients system (ACCES) where the time depending part of F is fixed by $t = t_0$:

$$\dot{Y}(t) = F(Y(t), t_0) \quad (11)$$

Some results concerning these dynamical systems are useful in this work, especially those concerning the invariant manifolds.

A. Invariant manifold of the ACCES

There are different ways to compute the invariant manifold $\phi(Y) = 0$ of such an autonomous dynamical system with constant coefficients. According [4], it can be worked out from a linear combination of the eigenvectors $\chi_1(t_0)$ and $\chi_2(t_0)$, which are tangent to the manifold. According the differential geometry method [1] [2], the invariant manifold is given by the equation

$$\det [\dot{Y} \mid \ddot{Y} \mid \ddot{\ddot{Y}} \mid \dots] = 0 \quad (12)$$

B. Proposition 2: trajectories envelop.

Let D be a domain of the phase space in which the Liapunov exponents μ_3, \dots, μ_n have a negative real part. Assume that there exists t_0 such that the relationship (9) holds. Then the manifold $\phi(X) = 0$ is the envelop of the trajectories.

Proof. When $t = t_0$, the initial PADS $\dot{X}(t) = F(X(t), t)$ and the ACCES $\dot{Y}(t) = F(Y(t), t_0)$ have the same velocity. Then, from the hypothesis, the trajectories of PADS are located on the manifold $\phi(X) = 0$. There, the solution tends towards two modes of the APPLE. From (6), the time derivative $\dot{X}(t)$ includes terms multiplying $\chi_1(t)$ and $\chi_2(t)$, which are parts of the velocity component on the manifold, and terms containing $\dot{\chi}_1(t)$ and $\dot{\chi}_2(t)$, that define the transversal component. On $\phi(X) = 0$, from (7), the transversal one is null. Then the trajectories are tangent to the manifold with a maximal or minimal amplitude.

A. Proposition 3: the manifold crossing problem.

Let t_0 be a value of the time defined in the propositions 1 and 2 corresponding to a maximal amplitude of the solutions, and t'_0 the time corresponding to the following minimum. For

any time t_1 , with $t_0 < t_1 < t'_0$, there are ν_k, η_k such that

$$J_{t_1} \eta_k = \nu_k \eta_k, \quad 1 \leq k \leq n \quad (13)$$

Let us suppose that X belongs to the domain D of the phase space as defined in § II A in which μ_3, \dots, μ_n have a negative real part. Let ν_k be the eigenvalue that tends to μ_k when t tends to t_0 and $\eta_k(t_0)$ the associated eigenvector, $1 \leq k \leq 2$. Let $\psi(X) = 0$ be the equation of the manifold constructed from the eigenvectors $\eta_1(t_0)$ and $\eta_2(t_0)$ in the same way that $\phi(X) = 0$ is constructed from $\chi_1(t_0)$ and $\chi_2(t_0)$.

Then the trajectories cross the manifold $\psi(X) = 0$ at times $t_1 + nT$.

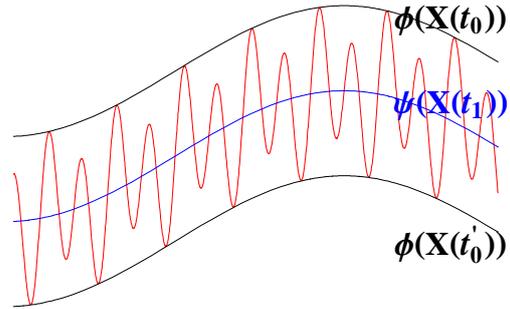


Fig.1. For a two-dimensional model, this drawing shows in some domain of the phase plane the envelop made of two curves: $\phi(X(t_0))$ tangent to the maxima, $\phi(X(t'_0))$ tangent to the minima, and the manifold $\psi(X(t_1))$ periodically crossed by the trajectories.

Proof. From the existence and uniqueness theorem, the manifold defined in this proposition is located between the layers of the manifold $\phi(X) = 0$.

IV. MONOMODAL CASES

In the case where μ_2, \dots, μ_n have a negative real part then the corresponding modes vanishes and the APPLE has a *monomodal* solution related to μ_1 . The theory is the same and simpler [6].

V. APPLICATION TO THE CHUA SYSTEM

A. Second order parametric Chua's system

Consider the periodic autonomous dynamical system defined in \mathbb{R}^2 :

$$\begin{cases} \dot{x} = f(x, t) - y \\ \dot{y} = 0.9 - x \end{cases}$$

where

$$f(x, t) = \begin{cases} [1 + a_1 \cos(\omega t)] x, & \text{if } |x| \leq 1 \\ [-3 + a_1 \cos(\omega t)] x + 4, & \text{if } x > 1 \\ [-3 + a_1 \cos(\omega t)] x - 4, & \text{if } x < -1 \end{cases}$$

For $a_1 = 0$, the dynamical system with constant coefficients has an unstable focus in $[0.9, 0]$ and a stable periodic solution. For $a_1 \neq 0$, the Jacobian matrix intervening in the APPLE is:

$$J_t = \begin{bmatrix} p + a_1 \cos(\omega t) & -1 \\ -1 & 0 \end{bmatrix}$$

with $p = 1$ if $|x| \leq 1$ and $p = -3$ if $|x| > 1$.

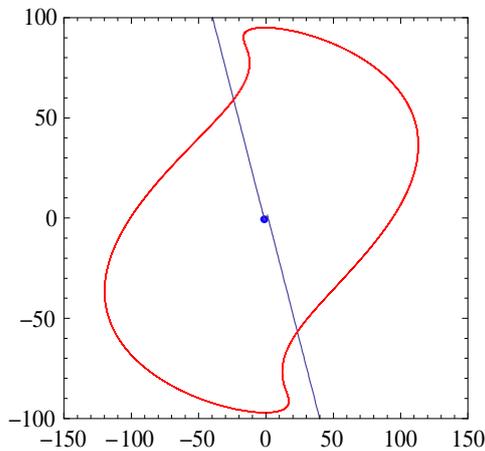


Fig.2. When $\omega = 1.25664$, $a_1 = 8$, the Liapunov exponents are : for $|x| > 1$, $\mu_1 = -0.0166557 + i 0.628319$, $\mu_2 = -2.98334 + i 0.628319$ and for $|x| \leq 1$, $\mu_1 = 1.62692 + i 0.628319$, $\mu_2 = -0.626917 + i 0.628319$. The imaginary part of the Liapunov exponent is exactly equal to $\frac{\omega}{2}$ (parametric resonance). The manifold (blue) is periodically crossed by the solution (red).

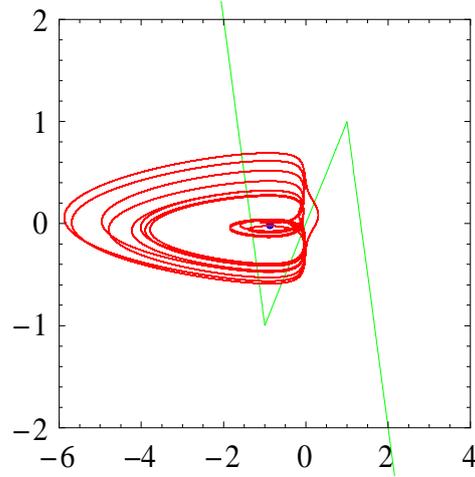


Fig.3. When $\omega = 6.34377$, $a_1 = 6$, the Liapunov exponents are : for $|x| > 1$, $\mu_1 = -2.37471$, $\mu_2 = -0.625293$ (stable zone) and for $|x| \leq 1$, $\mu_1 = 0.5 - i 1.1631$, $\mu_2 = -0.626917 + i 1.1631$ (unstable zone). This picture shows a eight period solution on the route towards chaos (red) and $f(x, \frac{\pi}{2})$ (green).

B. Third order parametric Chua system

Consider the Chua system defined in \mathbb{R}^3 :

$$\begin{cases} \dot{x} = \alpha(t)[y - x - f(x)] \\ \dot{y} = x - y + z \\ \dot{z} = -\beta(t)y \end{cases}$$

with

$$f(x) = bx + \frac{1}{2}(a - b)(|x + 1| - |x - 1|)$$

and $\alpha(t), \beta(t) : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto \alpha(t)$ and $\beta(t)$, are T-periodic. It is assumed that these variable coefficients have Fourier series development.. The Jacobian matrix is:

$$J_t = \begin{bmatrix} -\alpha(t)\left(\frac{df(x)}{dx} + 1\right) & \alpha & 0 \\ 1 & -1 & 1 \\ 0 & -\beta(t) & 0 \end{bmatrix}$$

$$\frac{df(x)}{dx} = \begin{cases} a & \text{if } |x| \leq 1 \\ b & \text{if } |x| > 1 \end{cases}$$

$$\begin{aligned} \text{Det}(J_{t_0}) &= -\alpha(t_0) \beta(t_0) \left(\frac{df(x)}{dx} + 1 \right) \\ &= \mu_1 \mu_2 \mu_3 \end{aligned}$$

$$\begin{aligned} \text{Tr}(J_{t_0}) &= -\alpha(t_0) \left(\frac{df(x)}{dx} + 1 \right) - 1 \\ &= \mu_1 + \mu_2 + \mu_3 \end{aligned}$$

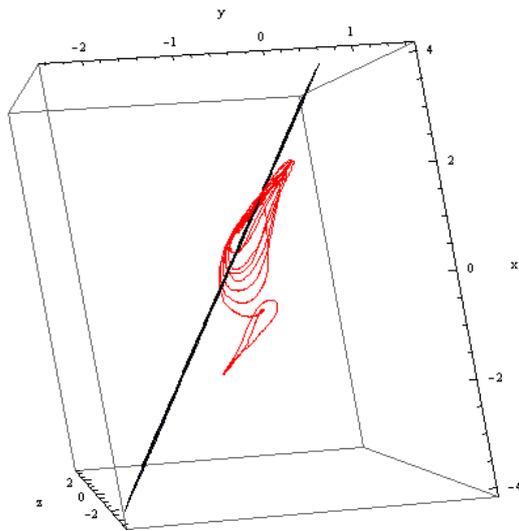


Fig.4. Third order Chua system with $a = -\frac{8}{7}$, $b = -\frac{5}{7}$, $\alpha = 16.6$ and $\beta(t) = 52[1 + .901060001 \cos(2\pi t)]$. The Liapunov exponents are $\mu_1 = -3.4518$, $\mu_2 = -0.133916$ and $\mu_3 = 4.81767$. The monomodal solution (red) periodically crosses the manifold.

REFERENCES

- [1] Ginoux J.-M. [2009] *Differential Geometry Applied to Dynamical Systems* (World Scientific Series on Nonlinear Science, series A)
- [2] Ginoux J.-M., Rossetto B. & Chua L. [2008] "Slow Invariant Manifolds as Curvature of the Flow of Dynamical Systems", *Int. J. Bifurcation & Chaos* **18** (11), 3409-3430.
- [3] Hill G. W. [1886] "On the Part of the Motion of the Lunar Perigee which is a Function of the Mean Motions of the Sun and the Moon", *Acta Mathematica* **8** (1), 1-36.
- [4] Rossetto B., Lenzini T., Ramdani S. & Suchey G. [1998] "Slow-Fast autonomous Dynamical Systems", *Int. J. of Bifurcation and Chaos* **8** (11), 2135-2145.
- [5] Rossetto B. [1986] "Détermination des exposants de Floquet-Liapounov de l'équation de Hill d'ordre n. Applications", *Japan Journal of Applied Mathematics* **3** (1), 93-128.
- [6] Rossetto B., Zhang Y. [2009] "Two-Dimensional Dynamical Systems with Periodic Coefficients", accepted for publication in *Int. J. Bifurcation & Chaos*.

Network Synchronization of Unified Chaotic Systems in Master-Slave Coupling

Cruz-Hernández C., López-Gutiérrez R. M., Inzunza-González E., and Cardoza-Avendaño L.

Abstract—In this paper, we use Generalized Hamiltonian forms to synchronize two unified chaotic systems. Synchronization is obtained in master-slave configuration with the slave being given by a state observer. In addition, this approach is applied to synchronize a complex dynamical network with chaotic nodes connected in chain via output signals. An illustrative example is given with a network with three chaotic unified systems, that is a single master node with two slaves (state observers).

Index Terms—Synchronization, Complex Dynamical Networks, Chaotic Systems, Observers, Generalized Hamiltonian Systems

I. INTRODUCTION

SYNCHRONIZATION of two coupled chaotic systems has received great attention from mathematicians, physicists, biologists, control engineers, etc. during the last decades, see e.g. [1]-[10]. This interest has been greatly motivated by the possibility of encrypted information transmission by using a chaotic carrier, see e.g. [11]-[19].

On the other hand, a complex dynamical network can be defined as a set interconnected of nodes, each node is considered like basic element with behavior depending of the nature of the network. This class of structures, has been observed in physics, biology, computer sciences, economy, chemistry, engineering, social sciences, see e.g. [20]-[23].

Recently synchronization in complex dynamical networks have received a great deal of attention from the scientific community. Particularly interesting is the case where the connected nodes have chaotic dynamics. Network synchronization is supposed to have interesting applications in different fields, see e.g. [24]-[30].

The aim of this paper is to study synchronization of two unified chaotic systems [31]. This objective is achieved by synchronizing the unified chaotic systems via Generalized Hamiltonian forms and observer design [5]. In addition, we show that the proposed approach is indeed suitable to synchronize complex dynamical networks composed by coupled chaotic nodes (unified systems) in master-slave architecture.

The remainder of this paper is organized as follows: In Section II, synchronization via Generalized Hamiltonian forms and observer design is provided. In Section III, the model of the unified chaotic system is described. Section IV presents the synchronization of two unified chaotic systems. In Section V, analytical conditions for chaos synchronization are given. While, synchronization in a chain of multiple unified chaotic

systems in master-slave topology is shown. In particular, synchronization is achieved among a single master node and two slave nodes. The paper is concluded with some remarks in Section VII.

II. SYNCHRONIZATION VIA GENERALIZED HAMILTONIAN FORMS AND OBSERVERS

Consider the following dynamical system

$$\dot{x} = f(x), \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a nonlinear function.

In [5] is reported how the dynamical system (1) can be written in the following *Generalized Hamiltonian canonical form*,

$$\dot{x} = \mathcal{J}(x) \frac{\partial H}{\partial x} + \mathcal{S}(x) \frac{\partial H}{\partial x} + \mathcal{F}(x), \quad x \in \mathbb{R}^n, \quad (2)$$

$H(x)$ denotes a smooth *energy function* which is globally positive definite in \mathbb{R}^n . The *gradient vector* of H , denoted by $\partial H / \partial x$, is assumed to exist everywhere. We use *quadratic energy function* $H(x) = (1/2)x^T M x$ with M being a constant, symmetric positive definite matrix. In such case, $\partial H / \partial x = Mx$. The matrices, $\mathcal{J}(x)$ and $\mathcal{S}(x)$ satisfy, for all $x \in \mathbb{R}^n$, the properties: $\mathcal{J}(x) + \mathcal{J}^T(x) = 0$ and $\mathcal{S}(x) = \mathcal{S}^T(x)$. The vector field $\mathcal{J}(x)\partial H / \partial x$ exhibits the conservative part of the system and it is also referred to as the workless part, or work-less forces of the system; and $\mathcal{S}(x)$ depicting the working or nonconservative part of the system. For certain systems, $\mathcal{S}(x)$ is negative definite or negative semidefinite. Thus, the vector field is considered as the dissipative part of the system. If, on the other hand, $\mathcal{S}(x)$ is positive definite, positive semidefinite, or indefinite, it clearly represents, respectively, the global, semi-global, and local destabilizing part of the system. In the last case, we can always (although nonuniquely) decompose such an indefinite symmetric matrix into the sum of a symmetric negative semidefinite matrix $\mathcal{R}(x)$ and a symmetric positive semidefinite matrix $\mathcal{N}(x)$. Finally, $\mathcal{F}(x)$ represents a locally destabilizing vector field.

In the context of observer design, we consider a *special class* of Generalized Hamiltonian forms with linear output map $y(t)$, given by

$$\begin{aligned} \dot{x} &= \mathcal{J}(y) \frac{\partial H}{\partial x} + (\mathcal{I} + \mathcal{S}) \frac{\partial H}{\partial x} + \mathcal{F}(y), \quad x \in \mathbb{R}^n, \quad (3) \\ y &= \mathcal{C} \frac{\partial H}{\partial x}, \quad y \in \mathbb{R}^m, \end{aligned}$$

Cruz-Hernández C. is with Electronic and Telecommunications Department, Scientific Research and Advanced Studies of Ensenada (CICESE), Ensenada B.C., 22860 México, e-mail: ccruz@cicese.mx

López Gutiérrez R.M., Inzunza-González E., Cardoza-Avendaño L. are with Engineering Faculty, Baja California Autonomous University, México

where \mathcal{S} is a constant symmetric matrix, not necessarily of definite sign. The matrix \mathcal{I} is a constant skew symmetric matrix, and \mathcal{C} is a constant matrix.

We denote the *estimate* of the state $x(t)$ by $\xi(t)$, and consider the Hamiltonian energy function $H(\xi)$ to be the particularization of H in terms of $\xi(t)$. Similarly, we denote by $\eta(t)$ the estimated output, computed in terms of the estimated state $\xi(t)$. The gradient vector $\partial H(\xi)/\partial \xi$ is, naturally, of the form $\mathcal{M}\xi$ with \mathcal{M} being a constant, symmetric positive definite matrix.

A *nonlinear state observer* for the special class of Generalized Hamiltonian form (3) is given by

$$\begin{aligned}\dot{\xi} &= \mathcal{J}(y) \frac{\partial H}{\partial \xi} + (\mathcal{I} + \mathcal{S}) \frac{\partial H}{\partial \xi} + \mathcal{F}(y) + K(y - \eta), \quad (4) \\ \eta &= \mathcal{C} \frac{\partial H}{\partial \xi}, \quad \eta \in \mathbb{R}^m,\end{aligned}$$

with $\xi \in \mathbb{R}^n$ and K is the *observer gain*.

The *state estimation error*, defined as $e(t) = x(t) - \xi(t)$ and the output estimation error, defined as $e_y(t) = y(t) - \eta(t)$, are governed by

$$\begin{aligned}\dot{e} &= \mathcal{J}(y) \frac{\partial H}{\partial e} + (\mathcal{I} + \mathcal{S} - KC) \frac{\partial H}{\partial e}, \quad e \in \mathbb{R}^n, \quad (5) \\ e_y &= \mathcal{C} \frac{\partial H}{\partial e}, \quad e_y \in \mathbb{R}^m,\end{aligned}$$

where the vector $\partial H/\partial e$ actually stands, with some abuse of notation, for the gradient vector of the *modified energy function*, $\partial H(e)/\partial e = \partial H/\partial x - \partial H/\partial \xi = \mathcal{M}(x - \xi) = \mathcal{M}e$. We set, when needed, $\mathcal{I} + \mathcal{S} = \mathcal{W}$.

Definition 1 (Chaotic synchronization) [10] *The slave system (4) (nonlinear state observer) synchronizes with the chaotic master system in the special class of Generalized Hamiltonian form (3), if*

$$\lim_{t \rightarrow \infty} \|x(t) - \xi(t)\| = 0, \quad (6)$$

no matter which initial conditions $x(0)$ and $\xi(0)$ have. Where the state estimation error $e(t) = x(t) - \xi(t)$ represents the synchronization error.

Theorem 1 [5] *The state $x(t)$ of the nonlinear system (3) can be globally, exponentially, asymptotically estimated by the state $\xi(t)$ of an observer of the form (4), if the pair of matrices $(\mathcal{C}, \mathcal{W})$, or the pair $(\mathcal{C}, \mathcal{S})$, is either observable or, at least, detectable.*

A necessary and sufficient condition for global asymptotic stability to zero of the estimation error (5) is given by the following theorem.

Theorem 2 [5] *The state $x(t)$ of the nonlinear system (3) can be globally, exponentially, asymptotically estimated, by the state $\xi(t)$ of the observer (4) if and only if, there exists a constant matrix K such that the symmetric matrix*

$$\begin{aligned}[\mathcal{W} - KC] + [\mathcal{W} - KC]^T &= [\mathcal{S} - KC] + [\mathcal{S} - KC]^T \\ &= 2 \left[\mathcal{S} - \frac{1}{2}(KC + C^T K^T) \right]\end{aligned}$$

is negative definite.

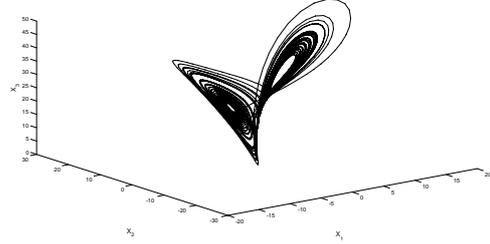


Fig. 1. Lorenz attractor projected onto the (x_1, x_2, x_3) -space.

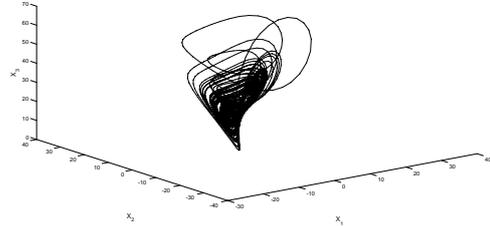


Fig. 2. Chen attractor projected onto the (x_1, x_2, x_3) -space.

III. UNIFIED CHAOTIC SYSTEM

Consider the unified chaotic system [31], described by

$$\begin{aligned}\dot{x}_1 &= (25\alpha + 10)(x_2 - x_1), \quad (7) \\ \dot{x}_2 &= (28 - 35\alpha)x_1 - x_1x_3 + (29\alpha - 1)x_2, \\ \dot{x}_3 &= x_1x_2 - \left(\frac{\alpha + 8}{3}\right)x_3,\end{aligned}$$

where the parameter $\alpha \in [0, 1]$, for the whole interval the unified system is always chaotic. Obviously, when $\alpha = 0$ the system (7) is the original Lorenz system [32]. While for $\alpha = 1$ the system (7) is the original Chen system [33]. For $\alpha = 4/5$ the system (7) corresponds to the critical (Lü) system [34]. In fact, the system (7) bridges the gap between the Lorenz and Chen systems [31].

By using the initial conditions $x(0) = (0.1, 0.1, 0.01)$, Figs. 1, 2, and 3 show the chaotic attractors of Lorenz, Chen, and Lü projected onto (x_1, x_2, x_3) -space.

IV. SYNCHRONIZATION OF TWO UNIFIED SYSTEMS

In this section, we synchronize two chaotic unified systems (7) in master-slave configuration (see Fig. 4(a)), via Generalized Hamiltonian forms and observer design proposed in [5]. Firstly, we rewrite the unified system (7) in Hamiltonian form as the master system and design a state observer for (7) like the slave system according to Fig. 4(b), as follows. Taking as Hamiltonian energy function to

$$H(x) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \quad (8)$$

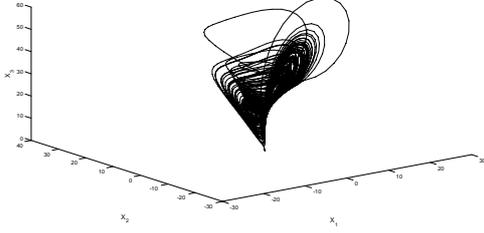


Fig. 3. Lü attractor projected onto the (x_1, x_2, x_3) -space.

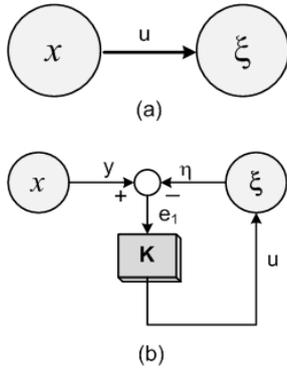


Fig. 4. Master-slave synchronization scheme.

and gradient vector as

$$\frac{\partial H}{\partial x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \quad (9)$$

The unified system (7) in Hamiltonian form according to Eq. (3) (as *master* system) is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 30\alpha - 9 & 0 \\ -(30\alpha - 9) & 0 & -x_1 \\ 0 & x_1 & 0 \end{bmatrix} \frac{\partial H}{\partial x} + \begin{bmatrix} -(25\alpha + 10) & -5\alpha + 19 & 0 \\ -5\alpha + 19 & 29\alpha - 1 & 0 \\ 0 & 0 & -\frac{(\alpha+8)}{3} \end{bmatrix} \frac{\partial H}{\partial x}. \quad (10)$$

The output signal to be transmitted to slave system is $y = [1 \ 0 \ 0] \partial H / \partial x = x_1$. The matrices \mathcal{C} , \mathcal{S} and \mathcal{I} , are given by

$$\mathcal{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},$$

$$\mathcal{S} = \begin{bmatrix} -(25\alpha + 10) & -5\alpha + 19 & 0 \\ -5\alpha + 19 & 29\alpha - 1 & 0 \\ 0 & 0 & -\frac{(\alpha+8)}{3} \end{bmatrix},$$

$$\mathcal{I} = \begin{bmatrix} 0 & 30\alpha - 9 & 0 \\ -(30\alpha - 9) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The pair of matrices $(\mathcal{C}, \mathcal{S})$ constitutes a pair detectable, but non observable. Even though the addition of the matrix \mathcal{I} to \mathcal{S} does not improve the lack of observability, the pair $(\mathcal{C}, \mathcal{W}) = (\mathcal{C}, \mathcal{S} + \mathcal{I})$ remains detectable. Then, according to Theorem 1, it is possible to design an observer for system (7). The nonlinear state observer (as *slave* system) according to Eq. (4) is designed as

$$\begin{bmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \\ \dot{\xi}_3 \end{bmatrix} = \begin{bmatrix} 0 & 30\alpha - 9 & 0 \\ -(30\alpha - 9) & 0 & -y \\ 0 & y & 0 \end{bmatrix} \frac{\partial H}{\partial \xi} + \begin{bmatrix} -(25\alpha + 10) & -5\alpha + 19 & 0 \\ -5\alpha + 19 & 29\alpha - 1 & 0 \\ 0 & 0 & -\frac{(\alpha+8)}{3} \end{bmatrix} \frac{\partial H}{\partial \xi} + \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} e_1, \quad (11)$$

$$\eta = \xi_1,$$

where the synchronization error is defined as $e_1(t) = y(t) - \eta(t)$. From (10) and (11) the synchronization error dynamics is governed by

$$\begin{bmatrix} \dot{e}_1 \\ \dot{e}_2 \\ \dot{e}_3 \end{bmatrix} = \begin{bmatrix} 0 & 30\alpha - 9 & 0 \\ -(30\alpha - 9) & 0 & -y \\ 0 & y & 0 \end{bmatrix} \frac{\partial H}{\partial e} + \begin{bmatrix} -(25\alpha + 10) & -5\alpha + 19 & 0 \\ -5\alpha + 19 & 29\alpha - 1 & 0 \\ 0 & 0 & \frac{(\alpha+8)}{3} \end{bmatrix} \frac{\partial H}{\partial e}. \quad (12)$$

V. STABILITY CONDITIONS

In this section, we examine the stability of the synchronization error (12) between master (10) and slave (11) systems. Invoking Theorem 2, which guarantees global asymptotic stability to zero of $e(t)$. In particular, for the synchronization of two unified systems, the matrix $2[\mathcal{S} - (\frac{1}{2})(\mathcal{K}\mathcal{C} - \mathcal{C}^T\mathcal{K}^T)]$ is given by

$$\begin{bmatrix} -50\alpha - 20 - 2k_1 & -10\alpha + 38 - k_2 & -k_3 \\ -10\alpha + 38 - k_2 & 58\alpha - 2 & 0 \\ -k_3 & 0 & -\frac{(2\alpha+16)}{3} \end{bmatrix} \quad (13)$$

by applying the Sylvester's criterion -which provides a test for negative definite of a matrix- thus, we have that the matrix (13) will be a negative definite matrix, if we choose k_1 , k_2 , and k_3 such that the following conditions are satisfied:

$$k_1 > -25\alpha - 10, \quad (14)$$

$$a < 0,$$

$$k_3^2 > \frac{\left(\frac{-2\alpha-16}{3}\right)(58\alpha-2)b-c}{(58\alpha-2)},$$

where $a = k_2^2 - 76k_2 + 3000\alpha^2 + 500\alpha + 1484 + 116\alpha k_1 + 4k_1$, $b = (-50\alpha - 20 - 2k_1)(58\alpha - 2)$, and $c = (-10\alpha + 38 - k_2)^2$. We have selected $k_1 = 11$, $k_2 = 50$, and $k_3 = 32$ and initial conditions $x(0) = (0.1, 0.1, 0.1)$ and $\xi(0) = (0, 0, 0)$. Fig. 5 shows the synchronization between master (10) and slave (11) systems.

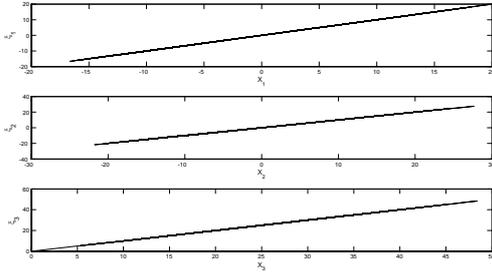


Fig. 5. Synchronization between master (10) and slave (11) systems.

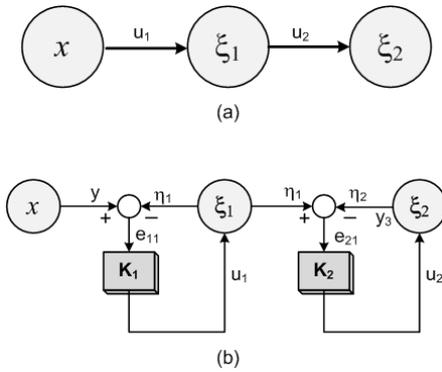


Fig. 6. Chain of nodes in master-slave coupling configuration.

VI. SYNCHRONIZATION OF UNIFIED SYSTEMS IN A NETWORK WITH CHAIN TOPOLOGY

Now, we show multiple synchronization of unified chaotic systems (7) like coupled nodes through output signal, by constituting a dynamical network with chain topology. In particular, for simplicity and illustration purpose only, we consider a dynamical network with three nodes in master-slave coupling configuration (see Fig. 6(a)); a single master node with two slave unified nodes (7), via Generalized Hamiltonian forms and observer design (see Fig. 6(b)). Firstly, we rewrite the unified system (7) for the master node and we design two observers connected in chain as slaves 1 and 2 in the following.

Consider the same master system (10) and two slaves 1 and 2 given by

$$\begin{aligned} \begin{bmatrix} \dot{\xi}_{11} \\ \dot{\xi}_{12} \\ \dot{\xi}_{13} \end{bmatrix} &= \begin{bmatrix} 0 & 30\alpha - 9 & 0 \\ -(30\alpha - 9) & 0 & -y \\ 0 & y & 0 \end{bmatrix} \frac{\partial H}{\partial \xi_1} \quad (15) \\ &+ \begin{bmatrix} -(25\alpha + 10) & -5\alpha + 19 & 0 \\ -5\alpha + 19 & 29\alpha - 1 & 0 \\ 0 & 0 & -\frac{(\alpha+8)}{3} \end{bmatrix} \frac{\partial H}{\partial \xi_1} \\ &+ \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} e_{11}, \\ \eta_1 &= \xi_{11}, \end{aligned}$$

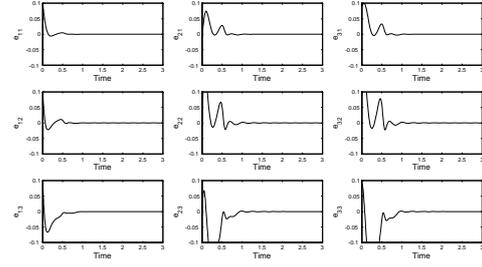


Fig. 7. Synchronization error trajectories for Lorenz system.

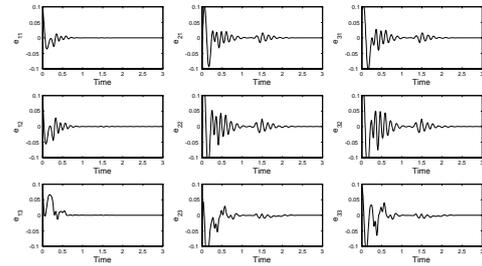


Fig. 8. Synchronization error trajectories for Chen system.

and the second observer (slave 2) as

$$\begin{aligned} \begin{bmatrix} \dot{\xi}_{21} \\ \dot{\xi}_{22} \\ \dot{\xi}_{23} \end{bmatrix} &= \begin{bmatrix} 0 & 30\alpha - 9 & 0 \\ -(30\alpha - 9) & 0 & -y \\ 0 & y & 0 \end{bmatrix} \frac{\partial H}{\partial \xi_2} \quad (16) \\ &+ \begin{bmatrix} -(25\alpha + 10) & -5\alpha + 19 & 0 \\ -5\alpha + 19 & 29\alpha - 1 & 0 \\ 0 & 0 & -\frac{(\alpha+8)}{3} \end{bmatrix} \frac{\partial H}{\partial \xi_2} \\ &+ \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} e_{21}, \\ \eta_2 &= \xi_{21}. \end{aligned}$$

where $\xi_1 = (\xi_{11}, \xi_{12}, \xi_{13})$ and $\xi_2 = (\xi_{21}, \xi_{22}, \xi_{23})$ are the state vectors of two slave chaotic nodes, respectively; and $e_{11} = y - \eta_1$ and $e_{21} = \eta_1 - \eta_2$. In the following numerical simulations, we have selected the initial conditions: $x(0) = (0.1, 0.1, 0.1)$, $\xi_1 = (0, 0, 0)$, and $\xi_2 = (0.01, 0.01, 0.01)$ and observer gains $K_1 = K_2 = (11, 50, 32)$. Fig. 7 shows the synchronization error trajectories among master node (10), slave node 1 (15), and slave node 2 (16), when Lorenz system ($\alpha = 0$) is considered like node (7). While Figs. 8 and 7 illustrate similar cases for Chen ($\alpha = 1$) and Lü ($\alpha = 4/5$), respectively.

VII. CONCLUSION

In this paper, we have firstly presented, synchronization between two coupled unified chaotic systems in master-slave

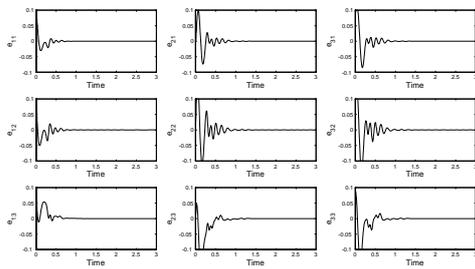


Fig. 9. Synchronization error trajectories for Lü system.

topology. Chaotic synchronization was achieved by using Generalized Hamiltonian forms and observer design proposed in [5]. Subsequently, this methodology was used to synchronize a chain of three coupled unified systems, that is a single master node coupled with two slave nodes via output signals and being given by two state observers. Nevertheless, this approach can be easily extended to network synchronization of N chaotic nodes (unified systems) coupled in master-slave configuration.

This result is particularly interesting given its potential application in communication network systems, where this connection is required. In addition, the approach can be implemented on experimental set-up [18], [19], [30].

VIII. ACKNOWLEDGMENT

The authors would like to thank to CONACYT, México by the support under Research Grants Nos. J49593-Y and P50051-Y.

REFERENCES

[1] L. M. Pecora and T.L. Carroll, "Synchronization in chaotic systems," *Phys. Rev. Lett.* 64, 821-824, 1990.

[2] Special Issue on Chaos synchronization and control: Theory and applications, *IEEE Trans. Circuits Syst. I*, 44, 1997.

[3] Special Issue on Control and synchronization of chaos, *Int. J. Bifurc. Chaos*, 10, 2000.

[4] C. Cruz-Hernández and H. Nijmeijer, "Synchronization through filtering," *Int. J. Bifurc. Chaos*, 10(4), 763-775 (2000). Synchronization through extended Kalman filtering. In: Nijmeijer H, Fossen TI, editors. *New trends in nonlinear observer design*. Lecture notes in control and information sciences, 244 London: Springer; 469-490, 1999.

[5] H. Sira-Ramírez and C. Cruz-Hernández, "Synchronization of chaotic systems: a generalized Hamiltonian systems approach," *Int. J. Bifurc. Chaos*, 11(5), 1381-1395 (2001). And in: Proceedings of the American Control Conference, Chicago, USA, 769-773, 2000.

[6] D. López-Mancilla and C. Cruz-Hernández, "Output synchronization of chaotic systems: model-matching approach with application to secure communication," *Nonlinear Dynamics and Systems Theory*, 5(2), 141-15, 2005.

[7] H. Nijmeijer and I. M. Y. Mareels, "An observer looks at synchronization," *IEEE Trans. Circuits Syst. I*, 44(10), 882-890, 1997.

[8] S. Boccaletti, J. Kurths, G. Osipov, D.L. Valladares, and C. Zhou, "The synchronization of chaotic systems," *Physics Reports* (2002) 336:1.

[9] A.C.J. Luo, "A theory for synchronization of dynamical systems," *Commun. Nonlinear Sci. Numer. Simulat.* (2008), doi:10.1016/j.cnsns.2008.07.002.

[10] *Advances in chaotic dynamics with applications*, Cambridge Scientific Publishers Ltd., Vol. 4, 2009.

[11] K. M. Cuomo, A.V. Oppenheim, and S. H. Strogatz, "Synchronization of Lorenz-based chaotic circuits with applications to communications," *IEEE Trans. Circuits Syst. I*, 40, 626-633, 1993.

[12] T. Yang and L.O. Chua, "Secure communication via chaotic parameter modulation," *IEEE Trans. Circuits Syst. I* 43(9) (1996) 817-819.

[13] J. Lu, X. Wu, and J. Lü, "Synchronization of a unified chaotic system and the application in secure communication," *Physics Letters A* 305 (2002) 365-370.

[14] C. Cruz-Hernández, "Synchronization of time-delay Chua's oscillator with application to secure communication," *Nonlinear Dynamics and Systems Theory* 4(1) (2004) 1-13.

[15] C. Posadas-Castillo, R.M. López-Gutiérrez, and C. Cruz-Hernández, "Synchronization of chaotic solid-state Nd:YAG lasers: application to secure communication," *Commun. Nonlinear Sci. Numer. Simul.* 13(8) (2008) 1655-1667.

[16] R.M. López-Gutiérrez, C. Posadas-Castillo, D. López-Mancilla, and C. Cruz-Hernández, "Communicating via robust synchronization of chaotic lasers," *Chaos, Solitons and Fractals* (2009), doi:10.1016/j.chaos.2008.11.019.

[17] L. Gámez-Guzmán, C. Cruz-Hernández, R.M. López-Gutiérrez, and E.E. García-Guerrero, "Synchronization of Chua's circuits with multi-scroll attractors: Application to communication," *Commun. Nonlinear Sci. Numer. Simulat.* 14 (2009) 2765-2775.

[18] C. Posadas-Castillo, C. Cruz-Hernández, and R. Núñez, "Experimental realization of binary signals transmission based on synchronized Lorenz circuits," *Journal of Applied Research Technology*, 2, 127-137, 2004.

[19] C. Cruz-Hernández, D. López-Mancilla, V. García, H. Serrano, and R. Núñez, "Experimental realization of binary signal transmission using chaos," *J. Circ. Syst. Comput.*, 14, 453-468, 2005.

[20] Y. Bar-Yam, *Dynamics of Complex Systems*, Addison-Wiley, USA, 1997.

[21] In search of a theory of complexity, *Chaos Solitons & Fractals* 34 (1) (2007) (Special Issue).

[22] S.C. Manrubia, A.S. Mikhailov, and D.H. Zanette, *Emergence of Dynamical Order Synchronization Phenomena in Complex Systems*, Lecture Notes in Complex Systems, vol. 2, World Scientific, Singapore, 2004.

[23] C.W.Wu, *Synchronization in complex networks of nonlinear dynamical systems*, World Scientific, Singapore, 2007.

[24] A. Yu Pogromsky and H. Nijmeijer, "Cooperative oscillatory behavior of mutually coupled dynamical systems," *IEEE Trans. Circ. Syst. I* 48(2) (2001) 152-162.

[25] X.F. Wang and G. Chen, "Synchronization in small-world dynamical networks," *Int. J. Bifurc. Chaos* 12(1) (2002) 187-192.

[26] X.F. Wang, "Complex networks: topology, dynamics and synchronization," *Int. J. Bifurc. Chaos* 12(5) (2002) 885-916.

[27] C. Posadas-Castillo, C. Cruz-Hernández, and R.M. López-Gutiérrez, "Synchronization in arrays of chaotic neural networks," in: P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, W. Pedrycz (Eds.), *Foundations of Fuzzy Logic and Soft Computing*, Lecture Notes in Artificial Intelligence, vol. 4529, Springer-Verlag, Berlin Heidelberg, 2007.

[28] C. Posadas-Castillo, C. Cruz-Hernández, and D. López-Mancilla, "Synchronization of chaotic neural networks: a generalized Hamiltonian systems approach," in: O. Castillo, P. Melin, J. Kacprzyk, W. Pedrycz (Eds.), *Hybrid Intelligent Systems Analysis and Design, Studies in Fuzziness and Soft Computing*, No. 208, Springer-Verlag, Berlin Heidelberg, 2007.

[29] C. Posadas-Castillo, C. Cruz-Hernández, and R.M. López-Gutiérrez, "Synchronization of chaotic neural networks with delay in irregular networks," *Applied Mathematics and Computation* 205 (2008) 487-496.

[30] C. Posadas-Castillo, C. Cruz-Hernández, and R.M. López-Gutiérrez, "Experimental realization of synchronization in complex networks with Chua's circuits like nodes," *Chaos, Solitons & Fractals*, in press doi:10.1016/j.chaos.2007.09.076.

[31] J. Lü, G. Chen, D. Cheng, and S. Celikovskiy, "Bridge the gap between the Lorenz system and the Chen system," *Int. J. Bifurc. Chaos*, 12 (2002) 2917-2926.

[32] E.N. Lorenz, "Deterministic non-periodic flows," *J. Atmos Sci.* 20 (1963) 130-141.

[33] G. Chen and T. Ueta, "Yet another chaotic attractor," *Int. J. Bifurc. Chaos* 9 (1999) 1465-1466.

[34] J. Lü and G. Chen, "A new chaotic attractor coined," *Int. J. Bifurc. Chaos* 12(3) (2002) 659-661.

DIFFERENTIAL GEOMETRY APPLIED TO DYNAMICAL SYSTEMS

Jean-Marc GINOUX & Bruno ROSSETTO

Laboratoire P.R.O.T.E.E.,
 Université du Sud Toulon - Var
 Institut Universitaire Technologique de Toulon,
 B.P. 20132, 83957, La Garde Cedex France
 e-mail: ginoux@univ-tln.fr
 http://ginoux.univ-tln.fr/

Abstract— *This work aims to present a new approach recently developed in a book [12] and called Flow Curvature Method that applies Differential Geometry to Dynamical Systems. Hence, for a trajectory curve, an integral of any n-dimensional dynamical system as a curve in Euclidean n-space, the curvature of the trajectory - or the flow - may be analytically computed. Then, the location of the points where the curvature of the flow vanishes defines a manifold called flow curvature manifold. Such a manifold being defined from the time derivatives of the velocity vector field, contains information about the dynamics of the system, hence identifying the main features of the system such as fixed points and their stability, local bifurcations of codimension one, center manifold equation, normal forms, linear invariant manifolds (straight lines, planes, hyperplanes).*

In the case of singularly perturbed systems or slow-fast dynamical systems, the flow curvature manifold directly provides the slow invariant manifold analytical equation associated with such systems. Also, starting from the flow curvature manifold, the corresponding dynamical system may be found again and thus, the inverse problem is solved.

Flow Curvature Method may be applied to any n-dimensional dynamical system autonomous or non-autonomous such as Van der Pol Model; FitzHugh-Nagumo Model; Pikovskii-Rabinovich-Traktengerts Model; Rikitake Model; Chua's Model; Lorenz Model. This article aims to present application of this method to the determination of the slow invariant manifold of a non-autonomous dynamical system: the forced Van der Pol model.

I. INTRODUCTION

Dynamical systems consisting of *nonlinear* differential equations are generally not integrable. In his famous memoirs: *Sur les courbes définies par une équation différentielle*, Poincaré [19], [22] faced to

this problem proposed to study *trajectory curves* properties in the *phase space*.

“... any differential equation can be written as:

$$\frac{dx_1}{dt} = X_1, \quad \frac{dx_2}{dt} = X_2, \quad \dots, \quad \frac{dx_n}{dt} = X_n$$

where X are integer polynomials.

If t is considered as the time, these equations will define the motion of a variable point in a space of dimension n .”

– Poincaré (1885, p. 168) –

Let's consider the following system of differential equations defined in a compact E included in \mathbb{R} as:

$$\frac{d\vec{X}}{dt} = \vec{\mathfrak{S}}(\vec{X}) \quad (1)$$

with

$$\vec{X} = [x_1, x_2, \dots, x_n]^t \in E \subset \mathbb{R}^n$$

and

$$\vec{\mathfrak{S}}(\vec{X}) = [f_1(\vec{X}), f_2(\vec{X}), \dots, f_n(\vec{X})]^t \in E \subset \mathbb{R}^n$$

The vector $\vec{\mathfrak{S}}(\vec{X})$ defines a velocity vector field in E whose components f_i which are supposed to be continuous and infinitely differentiable with respect to all x_i and t , i.e. are C^∞ functions in E and with values included in \mathbb{R} , satisfy the assumptions of the Cauchy-Lipschitz theorem. For more details, see for example [2]. A solution of this system is a *trajectory curve* $\vec{X}(t)$ tangent¹ to $\vec{\mathfrak{S}}$ whose values define the *states* of the *dynamical system* described by the Eq. (1).

¹Except at the *fixed points*.

Thus, *trajectory curves* integral of dynamical systems (1) regarded as n -dimensional *curves*, possess local metrics properties, namely *curvatures* which can be analytically² deduced from the so-called Frénet formulas [9]. For low dimensions two and three the concept of *curvatures* may be simply exemplified. A three-dimensional³ curve for example has two *curvatures*: *curvature* and *torsion* which are also known as *first* and *second curvature*. *Curvature*⁴ measures, so to speak, the deviation of the curve from a straight line in the neighborhood of any of its points. While the *torsion*⁵ measures, roughly speaking, the magnitude and sense of deviation of the curve from the *osculating plane*⁶ in the neighborhood of the corresponding point of the curve, or, in other words, the rate of change of the *osculating plane*. Physically, a three-dimensional curve may be obtained from a straight line by bending (*curvature*) and twisting (*torsion*). For high dimensions greater than three, say n , a n -dimensional curve has $(n - 1)$ *curvatures* which may be computed while using the Gram-Schmidt orthogonalization process [13] and provides the Frénet formulas [9] for a n -dimensional curve.

In [10] it has been established that the location of the point where the *curvature of the flow*, i.e. the *curvature of the trajectory curves* integral of any *slow-fast dynamical systems* of low dimensions two and three vanishes directly provides the *slow invariant manifold* analytical equation associated to such dynamical systems. So, in this work the new approach proposed by Ginoux *et al.* [10] is generalized to high-dimensional dynamical systems.

In a book recently published [12] efficiency of the *Flow Curvature Method* has been extensively exemplified. One of the main applications of the *Flow Curvature Method* presented in the next section establishes that *curvature of the flow*, i.e. *curvature of trajectory curves* of any n -dimensional dynamical system directly provides its *slow manifold* analytical equation the *invariance* of which is proved according to *Darboux Theorem*.

²Since only time derivatives of the *trajectory curves* are involved in the *curvature* formulas.

³A two-dimensional curve, i.e. a plane *curve* has a *torsion* vanishing identically.

⁴The notion of *curvature* of a plane curve first appears in the work of Apollonius of Perga.

⁵The name *torsion* is due to L.I. Valle, *Trait de Gomtrie Descriptive*.

⁶The *osculating plane* is defined as the plane spanned by the instantaneous velocity and acceleration vectors.

II. SLOW INVARIANT MANIFOLD ANALYTICAL EQUATION

The concept of *invariant manifolds* plays a very important role in the stability and structure of dynamical systems and especially for *slow-fast dynamical systems* or *singularly perturbed systems*. Since the beginning of the twentieth century it has been subject to a wide range of seminal research. The classical geometric theory developed originally by Andronov [1], Tikhonov [25] and Levinson [15] stated that *singularly perturbed systems* possess *invariant manifolds* on which trajectories evolve slowly and toward which nearby orbits contract exponentially in time (either forward and backward) in the normal directions. These manifolds have been called asymptotically stable (or unstable) *slow manifolds*. Then, Fenichel [5], [8] theory for the persistence of normally hyperbolic invariant manifolds enabled to establish the local invariance of *slow manifolds* that possess both expanding and contracting directions and which were labeled *slow invariant manifolds*.

Thus, various methods have been developed in order to determine the *slow invariant manifold* analytical equation associated to *singularly perturbed systems*. The essential works of Wasow [26], Cole [3], O'Malley [17], [18] and Fenichel [5], [8] to name but a few, gave rise to the so-called *Geometric Singular Perturbation Theory* and the problem for finding the *slow invariant manifold* analytical equation turned into a regular perturbation problem in which one generally expected, according to O'Malley (1974 p. 78, 1991 p. 21) the asymptotic validity of such expansion to breakdown.

So, the main result of this work established in the next section is that *curvature of the flow*, i.e. *curvature of trajectory curves* of any n -dimensional dynamical system directly provides its *slow manifold* analytical equation the *invariance* of which is established according to *Darboux Theorem*. Since it uses neither eigenvectors nor asymptotic expansions but simply involves time derivatives of the velocity vector field, it constitutes a general method simplifying and improving the *slow invariant manifold* analytical equation determination of high-dimensional dynamical systems.

A. *Slow manifold of high-dimensional dynamical systems*

In the framework of *Differential Geometry trajectory curves* $\vec{X}(t)$ integral of n -dimensional dynamical systems (1) satisfying the assumptions of the

Cauchy-Lipschitz theorem may be regarded as n -dimensional *smooth curves*, i.e. *smooth curves* in Euclidean n -space *parametrized in terms of time*.

Proposition II.1: *The location of the points where the curvature of the flow, i.e. the curvature of the trajectory curves of any n -dimensional dynamical system vanishes directly provides its $(n - 1)$ -dimensional slow invariant manifold analytical equation which reads:*

$$\begin{aligned} \phi(\vec{X}) &= \dot{\vec{X}} \cdot \left(\ddot{\vec{X}} \wedge \ddot{\vec{X}} \wedge \dots \wedge \overset{(n)}{\vec{X}} \right) \\ &= \det \left(\dot{\vec{X}}, \ddot{\vec{X}}, \ddot{\vec{X}}, \dots, \overset{(n)}{\vec{X}} \right) = 0 \end{aligned} \quad (2)$$

where $\overset{(n)}{\vec{X}}$ represents the time derivatives of \vec{X} .

Proof: Cf. Ginoux *et al.* [11]; Ginoux [12] ■

B. Darboux invariance theorem

According to Schlomiuk [23], [24] and Llibre *et al.* [16] it seems that in his memoir entitled: *Sur les équations différentielles algébriques du premier ordre et du premier degré*, Gaston Darboux (1878, p. 71) has been the first to define the concept of *invariant manifold*. Let's consider a n -dimensional dynamical system (1) describing "the motion of a variable point in a space of dimension n ." Let $\vec{X} = [x_1, x_2, \dots, x_n]^t$ be the coordinates of this point and $\vec{V} = [\dot{x}_1, \dot{x}_2, \dots, \dot{x}_n]^t$ its velocity vector.

Proposition II.2: *Consider the manifold defined by $\phi(\vec{X}) = 0$ where ϕ is a C^1 in an open set U is invariant with respect to the flow of (1) if there exists a C^1 function denoted $K(\vec{X})$ and called cofactor which satisfies:*

$$L_{\vec{V}}\phi(\vec{X}) = K(\vec{X})\phi(\vec{X}) \quad (3)$$

for all $\vec{X} \in U$ and with the Lie derivative operator defined as:

$$L_{\vec{V}}\phi = \vec{V} \cdot \vec{\nabla}\phi = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} \dot{x}_i = \frac{d\phi}{dt}.$$

In the following *invariance* of the *slow manifold* will be established according to what will be referred as *Darboux Invariance Theorem*.

Proof: Cf. Ginoux *et al.* [11]; Ginoux [12] ■

III. FORCED VAN DER POL MODEL

In this section it will be shown that *Flow Curvature Method* may applied to non-autonomous dynamical systems. As an example let's consider the forced Van der Pol model [14] which may be written as:

$$\vec{V} \begin{pmatrix} \varepsilon \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \vec{\mathfrak{F}} \begin{pmatrix} f(x, y, z) \\ g(x, y, z) \\ h(x, y, z) \end{pmatrix} = \begin{pmatrix} x + y - \frac{x^3}{3} \\ -x + a \sin(2\pi z) \\ \omega \end{pmatrix}$$

A suitable variable changes may transform this *non-autonomous* system into a *slow-fast autonomous* one which reads:

$$\begin{aligned} \vec{V} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} &= \vec{\mathfrak{F}} \begin{pmatrix} f_1(x_1, x_2, x_3, x_4) \\ f_2(x_1, x_2, x_3, x_4) \\ f_3(x_1, x_2, x_3, x_4) \\ f_4(x_1, x_2, x_3, x_4) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\varepsilon} \left(x_1 + x_2 - \frac{x_1^3}{3} \right) \\ -x_1 + ax_3 \\ \Omega x_4 \\ -\Omega x_3 \end{pmatrix} \end{aligned} \quad (4)$$

where $\varepsilon = 0.002$, $a = 1.8$, $\omega = 1.342043$ and $\Omega = 2\pi\omega$. Although this transformation increases the dimension of the system the *Flow Curvature Method* enables, according to Prop. II.1, to directly compute the *slow manifold analytical equation* associated with system (4) the equation of which reads:

$$\phi(\vec{X}) = \vec{V} \cdot (\vec{\gamma} \wedge \dot{\vec{\gamma}} \wedge \ddot{\vec{\gamma}} \wedge \overset{(4)}{\vec{X}}) = 0 \quad (5)$$

Then, it may be stated that in the vicinity of the *flow curvature manifold* both *flow curvature manifold* and its Lie derivative are merged. Thus, according to Darboux Invariance Theorem and Prop. II.2 the *slow manifold* of forced Van der Pol model is locally invariant (Cf. Fig. 1).

IV. DISCUSSION

In this work a new approach which consists in applying *Differential Geometry to Dynamical Systems* and called *Flow Curvature Method* has been partially presented. By considering the *trajectory curve*, integral of any n -dimensional dynamical system, as a *curve* in Euclidean n -space, the *curvature* of the *trajectory curve*, i.e. *curvature of the flow* has been analytically computed enabling thus to define a manifold called: *flow curvature manifold*. Since such man-

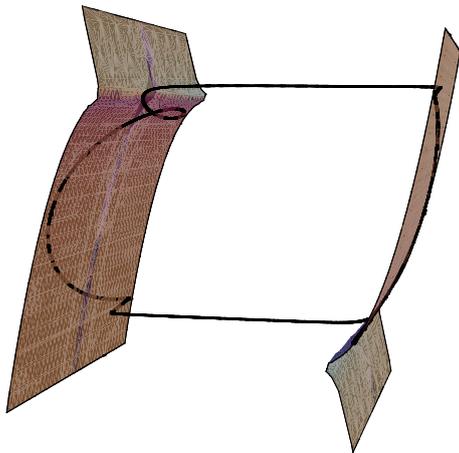


Fig. 1. Forced Van der Pol model *slow invariant manifold* in (x_1, x_2, x_3) phase space

ifold only involves the time derivatives of the velocity vector field and so, contains information about the *dynamics* of the *system*, it enables to find again the main features of the *dynamical system* studied. Thus, *Flow Curvature Method* enables to analytically compute: *fixed points stability, invariant sets, center manifold approximation, normal forms, local bifurcations, slow invariant manifold and integrability* of any *n-dimensional dynamical systems* but also to “detect” *linear invariant manifolds* of any *n-dimensional dynamical systems* which may be used to build *first integrals* of these systems.

Then, according to [12] *Flow Curvature Method* may be applied to any *n-dimensional autonomous dynamical systems singularly perturbed or non-singularly perturbed*, i.e. *slow-fast autonomous dynamical systems* such as Lorenz, Rikitake, (PRT) models, . . . , or any *n-dimensional non-autonomous dynamical systems singularly perturbed or non-singularly perturbed* such as forced Van der Pol model as exemplified in this work.

REFERENCES

[1] A. A. Andronov, S. E. Khaikin & A. A. Vitt. *Theory of oscillators*, I, Moscow, 1937. (Engl. transl., Princeton Univ. Press, Princeton, N. J., 1949).

[2] E. A. Coddington & N. Levinson. *Theory of Ordinary Differential Equations*, Mac Graw Hill, New York, 1955.

[3] J. D. Cole. *Perturbation Methods in Applied Mathematics*, Blaisdell, Waltham, MA, 1968.

[4] G. Darboux. Sur les équations différentielles algébriques du premier ordre et du premier degré, *Bull. Sci. Math.*, Sr. 2(2), pp. 60–96, pp. 123–143, pp. 151–200, 1878.

[5] N. Fenichel. Persistence and Smoothness of Invariant Manifolds for Flows, *Ind. Univ. Math. J.* **21**, pp. 193–225, 1971.

[6] N. Fenichel. Asymptotic stability with rate conditions, *Ind. Univ. Math. J.* **23**, pp. 1109–1137, 1974.

[7] N. Fenichel. Asymptotic stability with rate conditions II, *Ind. Univ. Math. J.* **26**, pp. 81–93, 1977.

[8] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations, *J. Diff. Eq.* **31**, pp. 53–98, 1979.

[9] F. Frénet. *Sur les courbes à double courbure*, Thèse Toulouse, 1847. Résumé dans *J. de Math.*, 17, 1847.

[10] J. M. Ginoux & B. Rossetto. Differential Geometry and Mechanics Applications to Chaotic Dynamical Systems, *Int. J. Bifurcation and Chaos* **4**, Vol. 16, pp. 887–910, 2006.

[11] J. M. Ginoux, B. Rossetto & L. O. Chua. Slow Invariant Manifolds as Curvature of the flow of Dynamical Systems, *Int. J. Bifurcation and Chaos* **11**, Vol. 18, pp. 3409–3430, 2008.

[12] J. M. Ginoux. *Differential Geometry Applied to Dynamical Systems*, World Scientific Series on Nonlinear Science, Series A, vol. 66, World Scientific, Singapore, 2009.

[13] H. Gluck. Higher curvatures of curves in Euclidean space, *American Math Monthly* **73**, pp. 699–704, 1966.

[14] J. Guckenheimer, K. Hoffman & W. Weckesser. The forced van der Pol equation I: the slow flow and its bifurcations, *SIAM J. App. Dyn. Sys.* **2**, pp. 1–35, 2002.

[15] N. Levinson. Perturbations of discontinuous solutions of non-linear Systems of differential equations, *Acta Mathematica* **82**, pp. 71–106, 1950.

[16] J. Llibre & J. C. Medrado. On the invariant hyperplanes for d-dimensional polynomial vector fields, *J. Phys. A: Math. Theor.* **40**, pp. 8385–8391, 2007.

[17] R. E. O’Malley. *Introduction to Singular Perturbations*, Academic Press, New York, 1974.

[18] R. E. O’Malley. *Singular Perturbation Methods for Ordinary Differential Equations*, Springer-Verlag, New York, 1991.

[19] H. Poincaré. Sur les courbes définies par une équation différentielle, *J. de Math. Pures et Appl.*, Série III **7**, pp. 375–422, 1881.

[20] H. Poincaré. Sur les courbes définies par une équation différentielle, *J. de Math Pures Appl.*, Série III **8**, pp. 251–296, 1882.

[21] H. Poincaré. Sur les courbes définies par une équation différentielle, *J. de Math. Pures et Appl.*, Série IV **1**, pp. 167–244, 1885.

[22] H. Poincaré. Sur les courbes définies par une équation différentielle, *J. de Math. Pures et Appl.*, Série IV **2**, pp. 151–217, 1886.

[23] D. Schlomiuk. Elementary first integrals of differential equations and invariant algebraic curves, *Expositiones Mathematicae* **11**, pp. 433–454, 1993.

[24] D. Schlomiuk. Algebraic and geometric aspects of the theory of planar polynomial vector fields, in *Bifurcations and Periodic Orbits of Vector Fields*, D. Schlomiuk Editor, NATO Advanced Study Institutes Series, Series C: Mathematical and Physical Sciences, Kluwer Academic Publishers, Vol. 408, pp. 429–467, 1993.

[25] A. Tikhonov. On the dependence of solutions of differential equations on a small parameter, *Mat. Sb. N.S.* **22** (2), pp. 193–204 (In Russian), 1948.

[26] W. R. Wasow. *Asymptotic Expansions for Ordinary Differential Equations*, Wiley-Interscience, New York, 1965.

High gain observer based synchronization for a class of time-delay chaotic systems. Application to secure communications.

Estelle Cherrier, Mohammed M'Saad

Abstract—This work investigates high gain observer design to synchronize a time-delay chaotic system. It is shown that the underlying class of nonlinear systems can be put into the canonical observable form, and thus high gain observer design framework can be extended to chaotic synchronization problem. Our approach is motivated by its simplicity of implementation: the observer gain synthesis relies on the explicit resolution of a time-invariant algebraic Lyapunov equation, which leads to a single parameter design. The proposed synchronization scheme is validated in a real-time experimental setup, based on Analog/Digital dSpace electronic device. At the end of the paper an information transmission process is provided, based on the previous synchronization scheme.

Index Terms—Chaos synchronization, high-gain observer, time-delay system

I. INTRODUCTION

If state estimation of linear systems has been widely treated through the last four decades, the nonlinear case, which concerns most of physical processes, remains however an open and very active research field. Among the recent applications of nonlinear state estimation theory, chaotic synchronization represents a pregnant issue, even if the words "chaos" and "synchronization" themselves have seemed incompatible for a long time. Indeed, on the one hand, the word "synchronization" come from the Greek roots $\sigma\upsilon\gamma$ (*syn*), which means "with", and $\chi\rho\nu\nu\omicron\varsigma$ (*chronos*), which means "time". Hence we can give a first definition of synchronization notion: it characterizes two systems having the same behavior at the same time. In fact, synchronization effects have been observed since the XVIIth century, when the Dutch mathematician Huygens noticed the synchronization of two pendulum clocks placed against the same wall. Consequently, synchronization was reserved to periodic systems (two signals were said synchronized if their periods were identical). On the other hand, among nonlinear systems, chaotic systems are characterized by a very complex behavior, asymptotically aperiodic. *A priori*, the nature of chaotic systems would seem to challenge the notion of synchronization. No further attention was paid to this issue, until 1983, and the work of Yamada and Fujisaka [1]. They noticed that, by coupling oscillators which on their own evolved chaotically, it was possible under certain hypotheses to force them to evolve in an identical manner. This happened even if the two

systems did not start with the same initial conditions. Despite this breakthrough, the subject of chaotic synchronization seemed to have no obvious applications until 1990. In their pioneering paper [2], Pecora and Carroll gave necessary and sufficient conditions under which two chaotic systems would synchronize. They also indicated that by using chaotic synchronization it might be possible to communicate in a secure way, by using the chaotic signal as a mask, used to hide the information-bearing message. This promising application gave rise to a huge number of papers concerned with chaotic synchronization. For general surveys on this subject, the reader is referred to the references [3], [4], [5]. Then synchronization has become a state estimation issue. The papers [6], [7] have shown that it is possible to estimate chaotic systems states, using nonlinear control theory. Indeed, the chaotic transmitter belongs to the wide class of nonlinear dynamical systems, whereas the receiver can be viewed as a nonlinear observer of the transmitter system. Furthermore, nonlinear estimation theory can be used to design a receiver which synchronizes with the driving system. This nonlinear control point of view brings many approaches to the receiver conception problem, and the underlying synchronization analysis problem. Among the huge amount of references on this subject, we can quote [8], which builds an observer-based synchronization scheme, guaranteeing an exponential synchronization. A generalization to a larger class of nonlinearities is proposed in [9]. [10] details a particular observer design, whose gain can be expressed in function of the desired convergence speed. But other approaches can also be found in the tremendous literature. For instance, the synchronization problem is addressed as a chaos suppression issue in [11]. [12] has established a synchronization criterion based on a linear feedback control, applied to Chua's circuit. [13] deals with a reduced-order observer-based exponential synchronization scheme, while [14] considers synchronization as a control problem. A comparison between different synchronization schemes, applied to well-known chaotic systems is performed in [15]. Sliding mode observers theory and an integral observer are used for synchronization purposes, respectively in [16] and [17]. [18] deals with synchronization of a class of time-delay chaotic systems, and proposes a phase-modulation based transmission scheme. More recently, a new family of chaotic systems has been exhibited, relying on the multimodel framework, and a dedicated synchronization process is detailed in [19]. Some adaptive unknown input observer have been proposed, for

The authors are with GREYC UMR 6072 CNRS, Boulevard du Maréchal Juin, 14050 Caen France.
E-mail: estelle.cherrier@greyc.ensicaen.fr

example in [20] or [21]. The former develops an adaptive unknown input observer for a chaotic transmitter whose linear part is affected by a time-delay, while the latter is also concerned with a robust approach to cope with parametric uncertainties and external disturbances. A new transmitter is dealt with in [22], called unified chaotic system: when a parameter is varied, the chaotic attractor is topologically equivalent to a Lorenz attractor, or a Chen or a Lü one. Most of the mentioned papers address a chaotic synchronization problem and propose an application to secure transmissions, but rarely with a security analysis or a precise exhibition of what the secret key is. This point will be discussed at the end of our paper.

In many papers that can be read in the literature, as in several aforementioned papers, the observer gain design relies on the resolution of a Linear Matricial Inequality (LMI), thanks to numerical convex optimization algorithms, provided that conservative assumptions are fulfilled. What we propose in this paper is a chaotic synchronization scheme using high-gain observer framework, extending our recent results detailed in [23]. In this latter paper, a high gain observer was proposed in the presence of one (or more) variable and known delay. The exponential convergence of the observer relies on the resulting solution of an algebraic Lyapunov equation and leads to an explicit expression of the observer gain.

The layout of this paper is as follows. Section II presents high gain observer design for a class of nonlinear time-delay chaotic systems with a synchronization purpose. The obtained results are applied in section III to information transmission, and tested both in simulation and through real-time experimental setup, based on Analog/Digital dSpace electronic device.

Notations: throughout this paper, $x_\tau(t)$ stands for $x(t - \tau)$.

II. HIGH-GAIN OBSERVER BASED SYNCHRONIZATION

This section presents a new observer based synchronization scheme, relying on high gain design framework, for a class of nonlinear time-delay systems.

A. Time-delay chaotic transmitter

It is claimed in some papers dealing with cryptanalysis (see [24] for example) that hyperchaotic systems are well suited for security purpose when used in synchronization and communication schemes. Besides, the presence of a delay in the dynamics of a nonlinear systems leads to an hyperchaotic behavior, this has been detailed in ref. [25]. Therefore we consider the following class of time-delay chaotic systems:

$$\dot{x}(t) = Ax(t) + F(x(t)) + H(x_\tau(t)) \quad (1)$$

with

$$A = \begin{pmatrix} -\alpha & \alpha & 0 \\ 1 & -1 & 1 \\ 0 & -\beta & -\gamma \end{pmatrix} \quad (2)$$

$$F(x(t)) = \begin{pmatrix} -\alpha\delta \tanh(x_1(t)) \\ 0 \\ 0 \end{pmatrix} \quad (3)$$

$$H(x_\tau(t)) = \begin{pmatrix} 0 \\ 0 \\ \varepsilon \sin(\sigma x_1(t - \tau)) \end{pmatrix} \quad (4)$$

Fig. 1 shows the bifurcation diagram of system (1) when the parameter σ is varied. The reader is referred to ref. [18] for a thorough study of this chaotic transmitter. Once the transmitter

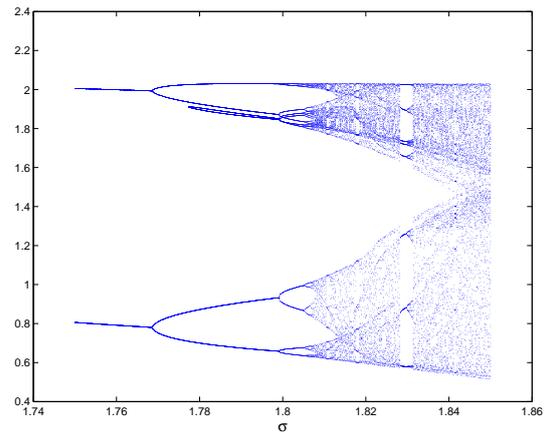


Fig. 1. Bifurcation diagram

has been chosen, we address in next subsection the dedicated receiver design, using high gain observer framework.

B. High gain observer synthesis

Since the pioneering paper of Gauthier *et al.* [26], which presents a high gain observer for a class of nonlinear systems called uniformly observable for all inputs, the general high gain framework has been extended to larger classes of nonlinear systems [27] (MIMO systems), [23] (time-delay systems), as well as larger problems (including state estimation), such as adaptive observers [28], to mention just a few.

We present in this paper an extension of the results established in reference [23] about high gain observer design in the presence of one (or more) variable and known delay. This class of high gain observers has, to the authors knowledge, not yet been applied to time-delay chaotic synchronization. The advantage of this approach principally remains in its simplicity of implementation, in the sense that the observer gain is obtained from the resolution of an algebraic Lyapunov equation, and can be given explicitly.

The main results of reference [23] are now briefly summed up.

Consider the following class of nonlinear systems [23]:

$$\begin{cases} \dot{x}(t) = Ax(t) + g(u(t), u_\tau(t), x(t), x_\tau(t)) \\ y(t) = Cx(t) \end{cases} \quad (5)$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}$, $u \in \mathbb{R}^m$, are respectively the state, the (scalar) output and the input of system (5).

A is the *anti-shift* matrix:

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix} \quad (6)$$

and the matrix C is defined by:

$$C = (1 \ 0 \ \dots \ 0) \quad (7)$$

The components of the nonlinear function $g: \mathbb{R}^{m+2n} \rightarrow \mathbb{R}^n$ are noted g_i , $i = 1, n$ and each one of them has a triangular structure w.r.t. x and x_τ , i.e. :

$$g_i(u, u_\tau, x, x_\tau) = g_i(u, u_\tau, x_1, \dots, x_i, x_{\tau,1}, \dots, x_{\tau,i}) \quad (8)$$

We introduce two matrices Δ_θ and S , which belong to the general high gain framework, as follows :

$$\Delta_\theta = \text{diag} \left[1 \quad \frac{1}{\theta} \quad \dots \quad \frac{1}{\theta^{n-1}} \right] \quad (9)$$

where θ is a strictly positive real number ;

S is the unique solution of the algebraic Lyapunov equation below:

$$S + A^T S + SA - C^T C = 0 \quad (10)$$

As in most of works dealing with high gain synthesis, we make the following assumption (cf. [27]) :

- (H1) The function g is global Lipschitz w.r.t. x and x_τ , uniformly in u .

Consider the following candidate observer:

$$\begin{cases} \dot{\hat{x}}(t) = A\hat{x}(t) + g(u(t), u_\tau(t), \hat{x}(t), \hat{x}_\tau(t)) \\ -\theta\Delta^{-1}S^{-1}C^T C(\hat{x}(t) - x(t)) \end{cases} \quad (11)$$

We give the main theorem of ref. [23] ensuring the convergence of observer (11):

Theorem 1:

Under hypothesis (H1), there exists $\theta_0 > 0$ such that for all $\theta > \theta_0$, system (11) is an exponential observer for system (5).

Now we will show how the chaotic transmitter (1) can be put into the canonical form (5)-(9) by an appropriate coordinate change:

$$\begin{cases} \dot{z}(t) = Az(t) + g(u(t), u_\tau(t), z(t), z_\tau(t)) \\ y(t) = Cz(t) \end{cases} \quad (12)$$

with A and C respectively defined by (6) and (7), and g of the form (8).

The appropriate coordinate change is given by [26]:

$$z(t) = \phi(x(t)) = \begin{pmatrix} x_1(t) \\ L_g x_1(t) \\ L_g^2 x_1(t) \end{pmatrix} \quad (13)$$

where $L_g f$ stands for the Lie derivative operator. If we note ϕ_i , $i = 1, 3$ the three components of ϕ , we obtain:

$$\begin{aligned} \phi_1(x(t)) &= x_1(t) \\ \phi_2(x(t)) &= -\alpha x_1(t) + \alpha x_2(t) - \alpha \delta \tanh(x_1(t)) \\ \phi_3(x(t)) &= \alpha(\alpha + 1)x_1(t) + \alpha^2 \delta(1 + \delta) \tanh(x_1(t)) \\ &\quad + \alpha^2 \delta \tanh(x_1(t))^2 (-x_1(t) + x_2(t)) \\ &\quad - \alpha^2 \delta^2 \tanh(x_1(t))^3 \\ &\quad - \alpha(\alpha + 1 + \alpha \delta)x_2(t) + \alpha x_3(t) \end{aligned} \quad (14)$$

Then following the results of [23], one can explicitly compute the observer gain for the canonical system (12):

$$K_z = \theta \Delta_\theta S^{-1} C^T$$

Once this is achieved, one has to find the expression of the observer gain in the original coordinates, which can be expressed as:

$$K = \left(\frac{\partial \phi}{\partial x} \right)^{-1} K_z$$

where $\left(\frac{\partial \phi}{\partial x} \right)$ stands for the Jacobian matrix of function ϕ .

It has been shown in [27] that only the diagonal terms of this Jacobian matrix are necessary, the other terms being controlled. It is also worth noticing the property below [26]:

$$S^{-1} C^T = (C_n^1 \ C_n^2 \ \dots \ C_n^m)$$

where $C_n^p = \frac{n!}{p!(n-p)!}$.

To conclude this section, we have proposed a new synchronization scheme, based on high gain observer framework, which has been recalled, for a time-delay chaotic transmitter.

III. REAL-TIME APPLICATION AND SECURE TRANSMISSION

The aim of this section is twofold. First we illustrate the effectiveness of the proposed synchronization scheme in simulations using Matlab, then in real-time experimental setup, based on Analog/Digital dSpace electronic device. Finally this synchronization process will be included in a complete communication system.

A. Real-time synchronization

We recall the model of the chosen transmitter, and the numerical values of its parameters:

$$\begin{cases} \dot{x}_1(t) = -\alpha x_1(t) + \alpha x_2(t) - \alpha \delta \tanh(x_1(t)) \\ \dot{x}_2(t) = x_1(t) - x_2(t) + x_3(t) \\ \dot{x}_3(t) = -\beta x_2(t) - \gamma x_3(t) + \varepsilon \sin(\sigma x_{1\tau}(t)) \end{cases} \quad (15)$$

with

α	β	γ	δ	ε	σ	τ
9	14	5	-1	10	10 ²	0.1

TABLE I
PARAMETERS OF SYSTEM (15)

For the simulation, we choose a fourth-order Runge-Kutta integration solver, with a constant step fixed to 1 ms.

The following initial conditions have been fixed for the transmitter and the receiver:

$$\begin{cases} x(t) = \hat{x}(t) = (0 \ 0 \ 0)^T \text{ for } t \in [-\tau, 0] \\ x(0) = (0.1 \ 0.1 \ 0.1)^T \\ \hat{x}(0) = (-0.1 \ -0.1 \ -0.1)^T \end{cases} \quad (16)$$

The value of the tuning parameter θ has been set to 10.

A comparison between the transmitter states and the receiver state is depicted in figure 2 and shows that identical synchronization is achieved after a few seconds. This synchronization time can be shortened by using larger values for θ . However this tuning must be made carefully, since the larger θ is, the less robust (to additive noise on the transmitted signal $y(t)$) the observer is.

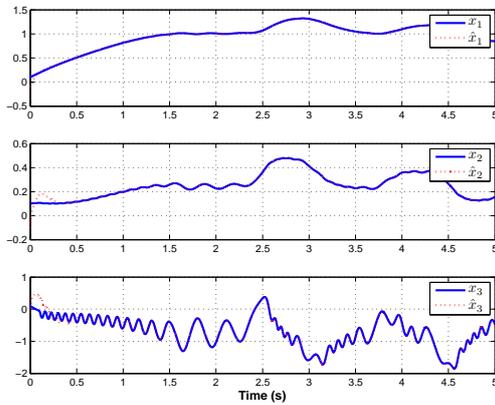


Fig. 2. Synchronization of the transmitter states and the receiver states

Now experimental results are performed on two calculators (Transmitter / Receiver) communicating through Analog/Digital dSpace electronic devices. At the first calculator, the Matlab-Simulink software simulates the chaotic model and transmits the output signal $y(t)$ through the dspace card (using a coaxial cable) to the receiver. At the receiver, the second calculator uses the proposed high gain observer based approach for synchronization. Fig. 3 shows the experimental results. It can be noticed that experimenting real transmission conditions inevitably lead to some degradations of the performances: while the first state x_1 seems exactly recovered, some unaccuracies appear during the synchronization of the second and the third states. These problems have been taken into account and are under study.

B. Application to information transmission

One proposes to integrate the previous high gain observer based synchronization scheme into a complete communication process. The information transmission is performed using the two-channel principle, as in [18]: a first signal (corresponding

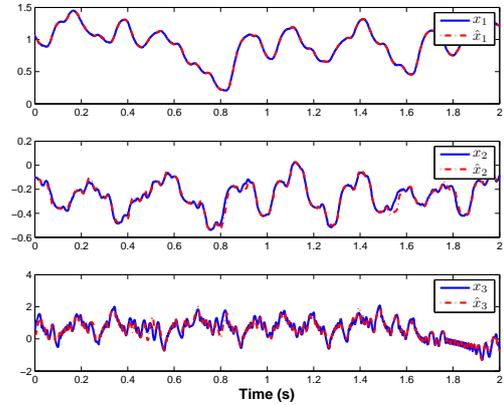


Fig. 3. Real-time synchronization

to $y(t)$ defined in (5)) is sent to the receiver, for synchronization purpose only. No information about the message is contained in this signal. Then, once synchronization is achieved at the receiver end, a second signal $y_2(t)$ containing the information (corresponding to an encryption of the message) is sent. To be able to decrypt the information, the receiver must possess the secret key, given by the transmitter. This point has been discussed in [18], where it has been shown that the parameter σ of the transmitter (5) can play the role of the secret key. In this case, we are dealing with a symmetric cryptosystem, since the same key is used to encrypt and decrypt the information. For lack of place, the security of the proposed communication scheme will not be longer discussed here, it would deserve an entire paper.

We give now the expression of the second signal $y_2(t)$ which is used to conceal the information, noted $u(t)$:

$$y_2(t) = x_3(t - T_u u(t)) \quad (17)$$

where we suppose without restriction that $u(t) \in [0, 1]$ and T_u is chosen equal to the fixed integration step.

Then the decryption formula is given by (see [18] for a detailed proof):

$$\hat{u}(t) = \frac{\hat{x}_3(t) - y_2(t)}{T_u \dot{\hat{x}}_3(t)} \quad (18)$$

where $\hat{u}(t)$ stands for the deciphered message.

Fig. 4 shows the effectiveness of the proposed cryptosystem when the following message is chosen: $u(t) = 0.5(1 + \sin(2\pi f_o t))$ with $f_o = 0.2\text{Hz}$.

Since the obtained results within the experimental setup were not totally satisfying, we decide not to make a real-time transmission trial. We prefer to perform a deeper study of high gain observer based synchronization. This paper is the first step in our approach.

IV. CONCLUSION

In this paper we addressed a chaotic synchronization problem. We propose a specific solution for a class of time-delay

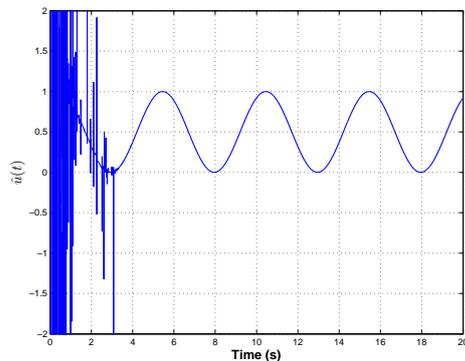


Fig. 4. Decrypted message $u(t) = 0.5(1 + \sin(2\pi f_o t))$

hyperchaotic transmitters, by designing a high-gain observer as a receiver. We first showed that the considered transmitter belongs to the class of uniformly observable nonlinear systems which is dealt with in the high gain framework. Then we detailed the conception of the receiver, whose efficiency has been tested not only in simulation using Matlab, but also in real-time experiment, using dSpace Analog/Digital device. At the end of the paper, the proposed synchronization scheme has been used to design a two-channel communication scheme based on chaotic phase modulation. This paper represents a first step in using high gain techniques for chaotic synchronization purpose. Further real-time experimentations of chaotic cryptosystems are under study.

REFERENCES

- [1] T. Yamada and H. Fujisaka, "Stability theory of synchronized motion in coupled-oscillator systems," *Prog. Theor. Phys.*, vol. 69, no. 1, pp. 32–47, 1983.
- [2] L. Pecora and T. Carroll, "Synchronization in chaotic systems," *Phys. Rev. Lett.*, vol. 64, no. 8, pp. 821–824, 1990.
- [3] L. Pecora, T. Carroll, G. Johnson, and D. Mar, "Fundamentals of synchronization in chaotic systems, concepts and applications," *Chaos*, vol. 7, no. 4, pp. 520–543, 1997.
- [4] S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. Zhou, "The synchronization of chaotic systems," *Physics Reports*, vol. 366, pp. 1–101, 2002.
- [5] M. Aziz-Alaoui, "Synchronization of chaos," *Encyclopedia of Mathematical Physics*, vol. 5, pp. 213–226, 2006.
- [6] O. Morgül and E. Solak, "Observer based synchronization of chaotic systems," *Physical Review E*, vol. 54, no. 5, pp. 4803–4811, 1996.
- [7] H. Nijmeijer and I. Mareels, "An observer looks at synchronization," *IEEE Trans. Circuit Syst. I*, vol. 44, no. 10, pp. 882–890, 1997.
- [8] T.-L. Liao and N.-S. Huang, "An observer-based approach for chaotic synchronization with applications to secure communications," *IEEE Trans. Circuit Syst. I*, vol. 46, no. 9, pp. 1144–1150, 1999.
- [9] M. Boutayeb, M. Darouach, and H. Rafaralahy, "Generalized state-space observers for chaotic synchronization and secure communication," *IEEE Trans. Circuit Syst. I*, vol. 49, no. 3, pp. 345–349, 2002.
- [10] J. Alvarez-Ramirez, H. Puebla, and I. Cervantes, "Convergence rate of observer-based approach for chaotic synchronization," *Phys. Lett. A*, vol. 289, pp. 193–198, 2001.
- [11] R. Femat, R. Jauregui-Ortiz, and G. Solís-Perales, "A chaos-based communication scheme via robust asymptotic feedback," *IEEE Trans. Circuit Syst. I*, vol. 48, no. 10, pp. 1161–1169, 2001.
- [12] G.-P. Jiang and W. Zheng, "An LMI criterion for linear-state-feedback based chaos synchronization of a class of chaotic systems," *Chaos, Solitons and Fractals*, vol. 26, pp. 437–443, 2005.
- [13] M. Feki, "Observer-based chaotic synchronization in the presence of unknown inputs," *Chaos, Solitons and Fractals*, vol. 15, pp. 831–840, 2003.
- [14] S. Čelikovský and G. Chen, "Secure synchronization of a class of chaotic systems from a nonlinear observer approach," *IEEE Trans. Automatic Control*, vol. 50, no. 1, pp. 76–82, 2005.
- [15] M. Haeri and B. Khademan, "Comparison between different synchronization methods of identical chaotic systems," *Chaos, Solitons and Fractals*, vol. 29, no. 4, pp. 1002–1022, 2006.
- [16] M. Chen, D. Zhou, and Y. Shang, "A new observer-based synchronization scheme for private communication," *Chaos, Solitons and Fractals*, vol. 24, pp. 1025–1030, 2005.
- [17] G. Jiang, W. Zheng, W. Tang, and G. Chen, "Integral-observer-based chaos synchronization," *IEEE Trans. Circuit Syst. II*, vol. 53, no. 2, pp. 110–114, 2006.
- [18] E. Cherrier, M. Boutayeb, and J. Ragot, "Observers based synchronization and input recovery for a class of nonlinear systems," *IEEE Trans. Circuit Syst. I*, vol. 53, no. 9, 2006.
- [19] E. Cherrier, M. Boutayeb, J. Ragot, and M. Aziz-Alaoui, "Observer-based exponential synchronization of chaotic multimodels," in *Proceedings of the European Control Conference, Kos, Greece*, 2007.
- [20] D. Li, Z. Wang, J. Zhou, J. Fang, and J. Ni, "A note on chaotic synchronization of time-delay secure communication systems," *Chaos, Solitons and Fractals*, vol. 38, no. 4, pp. 1217 – 1224, 2008.
- [21] S. Bowong and J. Tewa, "Unknown inputs' adaptive observer for a class of chaotic systems with uncertainties," *Mathematical and Computer Modelling*, vol. 48, no. 11-12, pp. 1826 – 1839, 2008.
- [22] J. Grzybowski, M. Rafikov, and J. Balthazar, "Synchronization of the unified chaotic system and application in secure communication," *Communications in Nonlinear Science and Numerical Simulation*, vol. 14, no. 6, pp. 2793 – 2806, 2009.
- [23] E. Cherrier, A. Sboui, M. Farza, and M. M'Saad, "Observateurs grand gain pour une classe de systmes non linaires retard," in *Proceedings of the IEEE Conférence Internationale Francophone d'Automatique, Bucarest, Roumania*, 2008.
- [24] G. Alvarez and S. Li, "Some basic cryptographic requirements for chaos-based cryptosystems," *Int. J. Bifurc. Chaos*, vol. 16, no. 8, pp. 2129–2151, 2006.
- [25] J. Farmer, "Chaotic attractors of an infinite-dimensional dynamical system," *Physica D*, vol. 4, pp. 366–393, 1982.
- [26] J. Gauthier, H. Hammouri, and S. Othman, "A simple observer for nonlinear systems: Application to bioreactors," vol. 37, no. 6, pp. 875–880, 1992.
- [27] M. Farza, M. M'Saad, and L. Rossignol, "Observer design for a class of MIMO nonlinear systems," *Automatica*, vol. 40, pp. 135–143, 2004.
- [28] M. Farza, M. M'Saad, T. Maatoug, and M. Kammoun, "A set of adaptive observers for a class of mimo nonlinear systems," in *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference, Seville, Spain*, 2005.

Observer based approach for synchronization in complex dynamical networks.

M. Boutayeb, A. Zemouche and J. J. Slotine

Abstract:

In this note, we investigate a state observer based approach for synchronization in complex dynamical networks. The later are in the general form with multi-dimensional links. Synchronization in the dynamical network is expressed in terms of asymptotic stability conditions deduced from the contraction theory. We show, thanks to the differential mean value theorem, that design and computation of the observer's gain may be performed through Linear Matrix Inequalities (convex problem).

Introduction

During the last decade, tremendous research activities were devoted to study complex dynamical networks which were introduced to model many aspects of real-world systems. Indeed, interesting mathematical models such as small-world models or scale-free models [1]-[2]-[4]-[11] were introduced recently to describe various dynamical systems in different area; without being exhaustive let us mention some of them, such as: Biological, Internet or Electrical Power network systems. The problem of synchronization become, therefore, a central question in particular to characterize behaviours of the dynamical networks. For doing so, several research works were recently devoted to analyze synchronization in complex network systems in different configurations such as connected neural networks with or without delays, synchronization of uncertain dynamical networks using robust impulsive techniques or synchronization in switching dynamical networks [5]-[6]-[7]-[8]-[12]-[13]-[17]-[18]. It should be noticed that most of the obtained results assume the complete knowledge of the state vector of each node, unfortunately, this condition is rarely satisfied. To overcome this obstacle, an interesting strategy consists in employing state observers based approach to assure synchronization when only one-dimensional links or few state components are available.

In the case of two chaotic master-slave systems, observers based approach for synchronization was largely investigated during the last two decades [3]-[9]-[10] with extensions to unknown input recovery [20]-[21]-[22]. Unfortunately, very few works were developed to deal with synchronization in complex dynamical networks. In [19], Jiang et al. have proposed a simple and useful observer based approach for synchronization in complex dynamical networks when only one dimensional links.

In this contribution we consider the problem of synchronization in complex dynamical networks from a state observer point of view. The systems considered here are in the general form with multi-dimensional

links not necessarily with the same measurement matrix. The stability analysis is investigated through the contraction theory [14]-[15]-[16] where the synthesis of the observer's gain may be deduced from a non conservative Linear Matrix Inequality (LMI) condition [24].

1 - Problem formulation and existing results

In this section we provide first some definitions and contributions to assure synchronization for a class of complex dynamical networks, after, we introduce a recent result on observers based approach for synchronization [19]. We should notice that we consider, here, systems that are described by a general form where the connections are not symmetric, in the sense that the information sent from node i to node j is multi-dimensional and may not be the same.

A - Definition

Consider a real world complex network with N nodes [4]-[5]-[6]-[7]-[8] :

$$\dot{x}_i(t) = f(x_i(t)) + \sum_{j=1}^N c_{ij} A x_j(t) \quad 1$$

where $x_i(t) \in R^n$, c_{ij} and $A \in R^{n \times n}$ represent the state vector of node i , the coupling strength between node i and node j and, a constant inner coupling matrix between nodes respectively.

The coupling configuration matrix of the network $C = (c_{ij})_{N \times N}$ is defined as follows :

$$\begin{aligned} c_{ij} &\neq 0 && \text{If there is a connection} \\ &&& \text{between node } i \text{ and node } j && 2 \\ c_{ij} &= 0 && \text{Otherwise} && 3 \end{aligned}$$

$$\text{with } c_{ii} = - \sum_{\substack{j=1 \\ j \neq i}}^N c_{ij} \quad 4$$

based on this configuration we may write:

$$\dot{s}(t) = f(s(t)) = f(s(t)) + \sum_{j=1}^N c_{ij} A s(t) \quad 5$$

which allows to deduce :

$$\dot{s}(t) - \dot{x}_i(t) = f(s(t)) - f(x_i(t)) + \sum_{j=1}^N c_{ij} A (s(t) - x_j(t)) \quad 6$$

Now let us introduce a general definition of network synchronization [].

M. Boutayeb and A. Zemouche are with Nancy University – CRAN UMR 7039 – France

J. J. Slotine is with MIT – NSL – Cambridge - USA

Definition-1: Let $x_i(t, t_0, x_0)$ denotes a solution for the differential equation:

$$\dot{x}_i(t) = f(x_i(t)) + g_i(x(t)) \quad i = 1, \dots, N \quad 7$$

with $x_0 = (x_1^0, \dots, x_N^0)$ and

$$x(t) = (x_1(t), \dots, x_N(t))$$

assume that $f(\cdot)$ and $g(\cdot)$ are continuously differentiable on an open subset D and there exists a nonempty $D_0 \subseteq D$ with $x_i^0 \in D_0$ such that $x_i(t) \in D_0$ and :

$$\lim_{t \rightarrow \infty} \|x_i(t, t_0, x_0) - x_j(t, t_0, x_0)\| = 0 \quad \text{for all } i, j \quad 8$$

therefore (8) is said to realize synchronization.

Now, let us go back to the system (7) with the notation: $\varepsilon_i(t) = s(t) - x_i(t)$, we deduce (using (5)) that:

$$\dot{\varepsilon}_i(t) = f(s(t)) - f(s(t) - \varepsilon_i(t)) + \sum_{j=1}^N c_{ij} A \varepsilon_j(t) \quad 9$$

$$\text{or } \dot{\varepsilon}(t) = F(t, s(t), \varepsilon(t)), \quad \varepsilon(t) = (\varepsilon_1(t) \dots \varepsilon_N(t)) \quad 10$$

using the fact that the coupling configuration matrix C has interesting properties (and assumed to be irreducible with real eigenvalues λ_k [23]) i.e. :

$$\lambda_1 = 0 > \lambda_2 \geq \dots \geq \lambda_N \quad 11$$

we obtain, through a linear and non singular transformation, an equivalent system (locally around the trajectory $s(t)$):

$$\dot{\delta}(t) = (Df(s(t)) + \lambda_k A) \delta(t), \quad \lambda_k = 2, \dots, N \quad 12$$

where $Df(s(t))$ is the Jacobien matrix of $f(s(t))$ around the trajectory $s(t)$. We notice that, as $\lambda_1 = 0$, $x_1(t) = s(t)$ is considered as the reference trajectory.

We can state now the main result for exponential synchronization [8] :

Lemma 1: Assume that the Jacobian matrix $DF(t, x(t))$ is bounded and Lipschitz (Locally) on an open subset Ω , uniformly in t , and C is diagonalized; then if $(A + A^T)$ is a positive semi definite matrix and the measure of matrix $\mu(Df(s(t)) + \lambda_2 A) < \alpha < 0$ then the dynamical complex network is exponentially stable (synchronized).

The proof is straightforward if we take a Lyapunov function as:

$$V(t) = \frac{1}{2} \delta^T(t) \cdot \delta(t) \quad 13$$

With

$$\begin{aligned} \dot{V}(t) = & \delta^T(t) \left(\frac{(Df(s(t)) + \lambda_2 A)^T + (Df(s(t)) + \lambda_2 A)}{2} \right) \delta(t) \\ & + (\lambda_k - \lambda_2) \delta^T(t) \left(\frac{A^T + A}{2} \right) \delta(t) \end{aligned}$$

$$\leq \delta^T(t) \left(\frac{(Df(s(t)) + \lambda_2 A)^T + (Df(s(t)) + \lambda_2 A)}{2} \right) \delta(t) \quad \text{for}$$

all k 14

B - Observers based approach for synchronization

To our knowledge there are very few results on state observers for synchronization, hereafter we summarize the main idea of a recent result on this subject. Indeed, instead of the complete knowledge of the coupling state vector x_j in the dynamical model (1), the latter is into the form :

$$\dot{x}_i(t) = f(x_i(t)) + \sum_{j=1}^N c_{ij} L y_j(t) \quad 15$$

where $y_j(t)$ is the output of node j with:

$$y_j(t) = H x_j \quad 16$$

H is a row vector $\in \mathbb{R}^{1 \times n}$

The problem is therefore how to determine the gain vector so that synchronization of the dynamical network is assured. Based on the Lemma 1 and the fact that $L.H$ is a constant matrix (which may represent the matrix A in (1)), the authors deduced an LMI condition to determine L for synchronization.

C - Problem formulation

In this note we consider a more general class of dynamical complex network where the connection between node i and node j is a multi-dimensional link, each node j provides only a part of the state vector x_j i.e. $y_j(t) = H_j x_j$, $H_j \in \mathbb{R}^{p_j \times n}$ with $p_j \leq n$. The dynamical model becomes:

$$\dot{x}_i(t) = f(x_i(t)) + \sum_{j=1}^N c_{ij} L_j y_j(t) \quad 17$$

Since the matrices $L_j H_j$ for $j = 1, \dots, N$ are not necessarily the same, the obtained results in [19] for synchronization can not be applied. To deal with problem, there are two options :

- The first one consists in determining matrices L_j under the constraint $L_j H_j = L_i H_i$ for all i, j . so that synchronization is assured.

- The second solution is more general and consists in determining matrices L_j in the sense of definition 1 i.e.

$$\begin{aligned} \lim_{t \rightarrow \infty} \|x_j(t) - x_i(t)\| &= 0 \quad \text{and not necessarily} \\ \lim_{t \rightarrow \infty} \|x_j(t) - s(t)\| &= 0 \quad 18 \end{aligned}$$

In this note we focus on the first option using simple and useful tools such as contraction theory for the stability analysis and the differential mean value theorem to determine matrices L_j through an LMI condition.

2 - Main results

A - Preliminary transformations

First of all, let us provide a general solution to the constraints $L_j H_j = L_i H_i$ which will be introduced into the global stability condition. Indeed, we may write :

$$L_j H_j = L_i H_i \quad \text{for } i, j = 1, \dots, N \quad 19$$

Into the equivalent form:

$$\begin{pmatrix} -H_1^T & H_2^T & 0 & \cdot & 0 \\ 0 & -H_2^T & H_3^T & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & -H_{N-1}^T & H_N^T \end{pmatrix} \begin{pmatrix} L_1^T \\ L_2^T \\ L_3^T \\ \cdot \\ L_N^T \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ 0 \end{pmatrix} \quad 20$$

or $H.L = 0$
with

$$H = \begin{pmatrix} -H_1^T & H_2^T & 0 & \cdot & 0 \\ 0 & -H_2^T & H_3^T & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & -H_{N-1}^T & H_N^T \end{pmatrix}, L = \begin{pmatrix} L_1^T \\ L_2^T \\ L_3^T \\ \cdot \\ L_N^T \end{pmatrix}$$

solutions of the system may be parameterized as follows:

$$L = (H^+ H - I_p) Z, \quad p = \sum_1^N p_i \quad 21$$

where H^+ is the pseudo-inverse matrix of H ($HH^+H = H$ and $H^+HH^+ = H^+$) and Z is an arbitrary matrix of dimension $p.n$ that parameterizes all solutions of the system. Indeed, it is obvious that : $H.L = 0$ for all matrices Z .

The solution (21) is therefore introduced into the stability condition to assure synchronization of the complex network.

Under the equality constraint, the state error vector of the complex network with respect to the arbitrary matrix Z may be written into the following form [19]:

$$\begin{pmatrix} \dot{\varepsilon}_1(t) & \dots & \dot{\varepsilon}_N(t) \end{pmatrix} = Df(s(t)) \begin{pmatrix} \varepsilon_1(t) & \dots & \varepsilon_N(t) \end{pmatrix} + \bar{Z}\bar{H} \begin{pmatrix} \varepsilon_1(t) & \dots & \varepsilon_N(t) \end{pmatrix} C^T \quad 22$$

$$\text{with } \bar{Z} = Z^T, \bar{H} = \begin{pmatrix} H^T H^{+T} - I_p \\ \cdot \\ H_N \end{pmatrix} \text{ and}$$

$$C^T = \begin{pmatrix} c_{11} & \dots & c_{N1} \\ \cdot & \dots & \cdot \\ c_{1N} & \dots & c_{NN} \end{pmatrix} \quad 23$$

or equivalently, through a non singular transformation :

$$\begin{pmatrix} \dot{\varepsilon}_1(t) & \dots & \dot{\varepsilon}_N(t) \end{pmatrix} \Phi = Df(s(t)) \begin{pmatrix} \varepsilon_1(t) & \dots & \varepsilon_N(t) \end{pmatrix} \Phi + \bar{Z}\bar{H} \begin{pmatrix} \varepsilon_1(t) & \dots & \varepsilon_N(t) \end{pmatrix} C^T \Phi \quad 24$$

$$\text{with } C^T \Phi = \Phi \Gamma, \quad \Gamma = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$$

Therefore we obtain a decoupled dynamical complex network in the same form as in [19]:

$$\dot{\delta}_i(t) = (Df(s(t)) + \lambda_i \bar{Z}\bar{H}) \delta_i(t)$$

$$\text{and } \begin{pmatrix} \delta_1(t) & \dots & \delta_N(t) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(t) & \dots & \varepsilon_N(t) \end{pmatrix} \Phi \quad 25$$

where \bar{Z} should be determined so that synchronization is assured.

B - Stability analysis

In the following, we investigate convergence of $\delta_i(t)$ to zero with the goal to introduce non conservative conditions. Indeed, let us provide first a general stability condition, based on the contraction theory approach [16]-[25].

Theorem 1:

Consider the system (25), if there exist a matrix Z and a contraction metric $M(x)$ (i.e. an $n.n$ symmetric matrix that is uniformly positive definite) such that the symmetric matrix

$$\begin{pmatrix} Df(s(t)) + \lambda_i \bar{Z}\bar{H} \end{pmatrix}^T M(x) + M(x) \begin{pmatrix} Df(s(t)) + \lambda_i \bar{Z}\bar{H} \end{pmatrix} + \dot{M}(x)$$

is uniformly negative definite for all $i=2, \dots, N$ then all trajectories $\delta_i(t)$ goes to zero.

The proof is straightforward (for details see [16]).

It should be noticed that the contraction metric $M(x)$ may be constructed using sum of squares (SOS) programming. In [25], the authors show how to obtain less conservative stability conditions for polynomial metrics $M(x)$.

In the following we will provide an LMI condition to deduce the arbitrary matrix Z so that $\begin{pmatrix} Df(s(t)) + \lambda_i \bar{Z}\bar{H} \end{pmatrix}^T M(x) + M(x) \begin{pmatrix} Df(s(t)) + \lambda_i \bar{Z}\bar{H} \end{pmatrix} + \dot{M}(x)$ is uniformly negative definite when $M(x)$ is a constant positive definite matrix M .

Theorem 2:

Under the assumption that the Jacobian matrix $Df(s(t))$ is a bounded matrix (i.e. each component has a lower and upper bound \underline{s}_{ij} and \bar{s}_{ij} respectively), if there exist matrices $P = P^T > 0$ and R of appropriate dimensions such that the following LMIs' are satisfied:

$$F^T(\underline{s}_{ij}, \underline{s}_{ij})P + PF(\underline{s}_{ij}, \underline{s}_{ij}) + \lambda_k (\bar{H}^T R^T + R\bar{H}) < 0 \quad 26$$

for all $k = 2, \dots, N$
then the state error vector $\delta_i(t)$ goes asymptotically to zero.

On the other hand, when (26) is feasible the gain matrix is given by $\bar{Z} = P^{-1}R$ and $L = (H^+H - I_p)\bar{Z}^T$

For the lack of space here, the proof is omitted here but may be deduced along the recent work in [24].

We may also show that the asymptotic stability condition (26) is non conservative in the sense that (26) allows very large Lipschitz constant and in the same time we didn't need that the Jacobian matrix to be into the form $Df(s(t)) = A + g(s(t))$ where (A, \bar{H}) is observable (comparisons with standard results are given in [24])

Conclusion

In this work we provide a constructive approach to assure synchronization of complex dynamical networks. The latter is in the general form using multi-dimensional links and not necessarily with the same measurement matrix. In order to cope with matrix inequalities under equality constraints, we provide first a preliminary transformation to parameterize all solutions of the equality constraints. After, thanks to the differential mean value theorem, a non conservative LMI condition is provided to assure synchronization of the dynamical network.

References

- [1] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.
- [2] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.
- [3] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Phys. Rev. Lett.*, vol. 64, no. 8, pp. 821–824, 1990.
- [4] X. Wang and G. Chen, "Complex network: Small-world, scale-free, and beyond," *IEEE Circuits Syst. Mag.*, vol. 3, no. 2, pp. 6–20, Feb. 2003.
- [5] —, "Synchronization in scale-free dynamical networks: Robustness and fragility," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 1, pp. 54–62, Jan. 2002.
- [6] —, "Synchronization in small-world dynamical networks," *Int. J. Bifurc. Chaos*, vol. 12, no. 1, pp. 187–192, 2002.
- [7] J. Lu, X. Yu, and G. Chen, "Chaos synchronization of general complex dynamical networks," *Physica A*, vol. 334, no. 1–2, pp. 281–302, 2004.
- [8] J. Lu, X. Yu, G. Chen, and D. Cheng, "Characterizing the synchronizability of small-world dynamical networks," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 4, pp. 787–796, Apr. 2004.
- [9] G. Grassi and S. Mascolo, "Nonlinear observer design to synchronize hyperchaotic systems via a scalar signal," *IEEE Trans. Circuits Syst. I*, vol. 44, no. 10, pp. 1011–1014, Oct. 1997.
- [10] Ö. Morgul and E. Solak, "On the synchronization of the chaotic systems by using state observers," *Int. J. Bifurc. Chaos*, vol. 7, no. 6, pp. 1307–1322, 1997.
- [11] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [12] Z. Li and G. Chen, "Global synchronization and asymptotic stability of complex dynamical networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 1, pp. 28–33, Jan. 2006.
- [13] J. Zhou, J. Lu, and J. Lu, "Adaptive synchronization of an uncertain complex dynamical network," *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 652–656, Apr. 2006.
- [14] W. Wang, J.-J. Slotine. Fast Computation with Neural Oscillators. *Neurocomputing*, 63, 2005.
- [15] W. Wang, J.-J. Slotine. On Partial Contraction Analysis for Coupled Nonlinear Oscillators. *Biological Cybernetics*, 92 (1):38–53, 2005.
- [16] W. Lohmiller, J.-J. Slotine. On Contraction Analysis for Nonlinear Systems. *Automatica*, 34 (6):671–682, 1998.
- [17] S.H. Strogatz, *Nonlinear Dynamics and Chaos: with applications to physics, biology, chemistry, and engineering*, Addison-Wesley Pub., Reading, MA, 1994.
- [18] Q. Zhang, Junan Lu, J. Lu, K. Tse, Adaptive Feedback Synchronization of a General Complex Dynamical Network With Delayed Nodes *IEEE TCAS—II*, VOL. 55, NO. 2, pp. 183-187, 2008.
- [19] G.P. Jiang, W.K.S. Tang and G. Chen. A State-Observer-Based Approach for Synchronization in Complex Dynamical Networks. *IEEE TCAS—I*, VOL. 53, NO. 12, pp.2739-2745, 2006.
- [20] M. Boutayeb, "Synchronization and input recovery in digital nonlinear systems," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 51, no. 8, pp. 393–399, Aug. 2004.
- [21] M. Boutayeb, M. Darouach, and H. Rafaralahy, "Generalized state space observers for chaotic synchronization and secure communication," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no.3, pp. 345–349, Mar. 2002.
- [22] T. L. Liao and N. S. Huang, "An observer-based approach for chaotic synchronization with applications to secure communications," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 46, no. 9, pp.1144–1150, Sep. 1999.
- [23] C. W. Wu, "Synchronization in an array of linearly coupled dynamical systems" *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 42, pp.430–447, Aug. 1995.
- [24] Zemouche A., M. Boutayeb, G. I. Bara. Observers for a Class of Lipschitz Systems with Extension to H-infinity Performance Analysis. *Systems & Control Letters*, vol. 57, no1, 2008.
- [25] Aylward E.M., Parrilo P. A., J.J. Slotine. Stability and robustness of non linear systems via contraction metrics and SOS programming . *Automatica*, vol. 44, pp. 2163-2170, 2008.

LPV Approach for the Stabilization of a Class of Dynamical Systems

Ali ZEMOUCHE and Mohamed BOUTAYEB

Abstract—In this note, we investigate the problem of \mathcal{H}_∞ observer-based state feedback controllers for a class of Lipschitz nonlinear systems. A simple systematic and useful synthesis method is proposed. Thanks to the use of the Differential Mean Value Theorem (DMVT), separation results are obtained. The synthesis conditions are given in term of Linear Matrix Inequalities (LMIs), easily tractable by convex optimization algorithms. Based on some mathematical tools such as Barbalat's lemma and convergence of series, an asymptotic convergence towards zero is provided in the \mathcal{H}_∞ free context. A numerical example is given in order to show the performances of the proposed approach.

Index Terms—Nonlinear observers, observer-based control, \mathcal{H}_∞ analysis, LMI approach, the Differential Mean Value Theorem (DMVT).

I. INTRODUCTION

Tremendous research activities on the problem of implementation and observers based control for linear and nonlinear dynamic systems show a growing interest in control theory area during the last decade. Thus, compliance with specific performances requires a careful choice of controller. To do this, the ideal strategy would be to have completely states and inputs of the system. However, for intrinsic reasons to the system or for high costs of the sensors implementation, the measurement of the state is often only partial. Moreover, the measurements may be affected by noises. That is why the controllers proposed in the literature are often based on state observers. Indeed, the observers incorporate a filter that estimates the entire state vector, or the unmeasured part only. It is one of the main reasons for which the observers design problem is extensively investigated in the recent literature [1], [2], [3], [4], [5], [6]. However, very less approaches concern the observer-based control problem for nonlinear systems. In the linear case, the problem is easy to investigate thanks to the separation principle. Nevertheless, the main difficulty for nonlinear systems lies in the fact that the separation principle is not always true.

Consequently, obtaining synthesis conditions in term of LMIs is a difficult task because of bilinear couplings between certain variables resulting from the stability analysis. To overcome this obstacle, LMI conditions under an equality constraint for a class of nonlinear discrete time systems are established [7]. But, this equality constraint is difficult to be solved for certain systems, in particular those with a single input.

A. ZEMOUCHE is with the Centre de Recherche en Automatique de Nancy, CRAN UMR 7039 CNRS, Nancy-Universit, 54400 Cosnes et Romain, France, (e-mail: Ali.Zemouche@iut-longwy.uhp-nancy.fr)

M. BOUTAYEB is with the Centre de Recherche en Automatique de Nancy, CRAN UMR 7039 CNRS, Nancy-Universit, 54400 Cosnes et Romain, France, (e-mail: Mohamed.Boutayeb@iut-longwy.uhp-nancy.fr)

Till now, the problem of observer-based state feedback controllers for nonlinear systems remains an open research subject. Obtaining a systematic observer-based control method for nonlinear systems under LMI conditions becomes a difficult challenge.

It is worth to notice that several research activities have been paid toward the study of the problem of observer-based control for linear systems and many applications are provided for the nonlinear case. We refer the reader to [8], [9], [10], [11], [12], just to mention some works. Note also that separation results are proposed in [13] and [14] for a class of nonlinear systems using a sufficiently fast high-gain observer. An extension, by the same authors, is given in [15], where their proposed state feedback controller renders a certain compact set positively invariant and asymptotically attractive. On the other hand, for all these works recent advances on stability analysis, using contraction theory, may be investigated [16], [17].

This paper deals with the problem of \mathcal{H}_∞ observer-based state feedback controllers for a class of Lipschitz nonlinear systems. A simple systematic synthesis method is proposed. Thanks to the use of the DMVT, we obtain separation results for a class of nonlinear systems. The synthesis conditions are given in term of linear matrix inequalities, easily tractable by convex optimization algorithms. It should be noticed that this work concerns the continuous-time and the discrete-time systems. Indeed, the main contribution of this work may be summarized into two points. The first one lies in the fact that a unified observer based control approach is established for both continuous and discrete time systems. The second point concerns the stability conditions which are expressed in terms of LMIs with \mathcal{H}_∞ performances.

This note is arranged as follows. In section II, we state the problem formulation that we consider in this study. In section III, we introduce some separation results and sufficient synthesis conditions. A numerical example is given in section IV for the discrete-time case, in order to illustrate the results. We end this note by a conclusion in section V.

Notations : The following notations will be used throughout this paper.

- $\|\cdot\|$ is the usual Euclidean norm;
- $(*)$ is used for the blocks induced by symmetry;
- A^T represents the transposed matrix of A ;
- I_r represents the identity matrix of dimension r ;
- for a square matrix S , $S > 0$ ($S < 0$) means that this matrix is positive definite (negative definite);
- the set $Co(x, y) = \{\lambda x + (1 - \lambda)y, 0 \leq \lambda \leq 1\}$ is the

- convex hull of $\{x, y\}$;
- $e_s(i) = \underbrace{(0, \dots, 0, \overset{i \text{ th}}{1}, 0, \dots, 0)}_{s \text{ components}}^T \in \mathbb{R}^s, s \geq 1$ is a vector of the canonical basis of \mathbb{R}^s ;
 - The notation $\|x\|_{\mathcal{L}_2^s} = \left(\int_0^\infty \|x(s)\|^2 ds \right)^{\frac{1}{2}}$ represents the \mathcal{L}_2^s norm of $x(t) \in \mathbb{R}^s$. Also, the ℓ_2 norm of $x(k)$ is defined by $\|x\|_{\ell_2^s} = \left(\sum_{k=0}^\infty \|x(k)\|^2 \right)^{\frac{1}{2}}$. Note that $\|\cdot\|_{\mathcal{L}_2^s}$ is used for the continuous-time case and $\|\cdot\|_{\ell_2^s}$ is used for the discrete-time case. The sets \mathcal{L}_2^s and ℓ_2^s are defined as

$$\mathcal{L}_2^s = \left\{ x \in \mathbb{R}^s \mid \|x\|_{\mathcal{L}_2^s} < +\infty \right\},$$

$$\ell_2^s = \left\{ x \in \mathbb{R}^s \mid \|x\|_{\ell_2^s} < +\infty \right\}.$$

II. PROBLEM FORMULATION

In this section, we introduce the class of nonlinear systems investigated in this paper. Since, we propose a unified method for both the continuous-time case and the discrete-time case, then we consider the class of systems described under the following unified form :

$$\sigma_x = A_x x + A_u u + B f(x) + E_\omega \omega \quad (1a)$$

$$y = C x + D_\omega \omega \quad (1b)$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ is the input vector and $y \in \mathbb{R}^p$ is the output of the system. $\omega \in \mathcal{L}_2^s$ (or $\omega \in \ell_2^s$) is the vector of disturbances. The matrices $A_x \in \mathbb{R}^{n \times n}$, $A_u \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{n \times q}$ and $C \in \mathbb{R}^{p \times n}$ are constant.

The notation σ_x is defined as follows :

$$\sigma_x(t) = \begin{cases} \dot{x}(t), & t \in \mathbb{R}_+ & \text{for the continuous-time case} \\ x(t+1), & t \in \mathbb{N} & \text{for the discrete-time case} \end{cases} \quad (2)$$

Assume that the function f satisfies the following condition :

$$a_{ij} \leq \frac{\partial f_i}{\partial x_j}(z) \leq b_{ij}, \quad \forall z \in \mathbb{R}^n \quad (3)$$

Without loss of generality, we assume also that $f(0) = 0$.

Remark 2.1: Note that condition (3) means that the differentiable function f is globally Lipschitz. In addition, (3) is not restrictive. Indeed, many nonlinearities can be regarded as Lipschitz, at least locally. For instance, the sinusoidal functions usually encountered in many physical processes are globally Lipschitz. On the other hand, even functions like x^3 (or all polynomial terms on x) can be considered as Lipschitz, provided that the state x is bounded (which is the case of many physical systems). Indeed, if the nonlinearity of the system is not globally Lipschitz on the whole \mathbb{R}^s , then if the state x is bounded, there exists always a positively invariant set Ω for the system. Hence, it suffices to extend the nonlinear term of the dynamics of the system from Ω to a globally Lipschitz function on \mathbb{R}^s , such that the trajectories of the system and those of the extending system starting from Ω are identical [18].

A structure of the dynamic observer-based control for the system (1) is given as follows :

$$\sigma_{\hat{x}} = A_x \hat{x} + A_u u + B f(\hat{x}) + L(y - C \hat{x}) \quad (4a)$$

$$u = -K \hat{x} \quad (4b)$$

where \hat{x} is the estimation of x . $L \in \mathbb{R}^{n \times p}$ is the observer gain and $K \in \mathbb{R}^{m \times n}$ is the control gain.

The objective is to determine the matrices K and L such that the system (1) becomes globally robustly asymptotically stable under the action of the observer-based linear static feedback

$$u = -K \hat{x}.$$

Using (4), we obtain the following dynamics :

$$\sigma_x = (A_x - A_u K)x + B f(x) + A_u K \varepsilon + E_\omega \omega \quad (5a)$$

$$\sigma_\varepsilon = (A_x - LC)\varepsilon + B(f(x) - f(\hat{x})) + (E_\omega - LD_\omega)\omega \quad (5b)$$

where $\varepsilon = x - \hat{x}$ is the estimation error. The system (5) can be rewritten under the unified form :

$$\sigma_{[x^T \quad \varepsilon^T]^T} = \begin{bmatrix} (A_x - A_u K) & A_u K \\ 0 & (A_x - LC) \end{bmatrix} \begin{bmatrix} x \\ \varepsilon \end{bmatrix} + \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} f(x) \\ f(x) - f(\hat{x}) \end{bmatrix} + \begin{bmatrix} E_\omega \\ E_\omega - LD_\omega \end{bmatrix} \omega. \quad (6)$$

Applying the DMVT [4], we have the existence of $z_i \in Co(0, x)$ and $\bar{z}_i \in Co(x, \hat{x})$, for $i = 1, \dots, q$, such that :

$$f(x) = \left(\sum_{i,j=1}^{q,n} H_{ij} \frac{\partial f_i}{\partial x_j}(z_i) \right) x \quad (7)$$

and

$$f(x) - f(\hat{x}) = \left(\sum_{i,j=1}^{q,n} H_{ij} \frac{\partial f_i}{\partial x_j}(\bar{z}_i) \right) \varepsilon \quad (8)$$

where

$$H_{ij} = e_q(i) e_n^T(j).$$

For simplicity of the presentation, we introduce some definitions as follows :

First, define the function

$$h : \mathbb{R}^{q \times n} \rightarrow \mathbb{R}^{q \times n} \quad (9)$$

$$\zeta \mapsto h(\zeta) = \left(h_{ij}(\zeta_i) \right)_{ij}$$

where

$$h_{ij}(\zeta_i) = \frac{\partial f_i}{\partial x_j}(\zeta_i), \quad \zeta = \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_q \end{bmatrix}, \quad \zeta_i \in \mathbb{R}^n$$

Define also the affine matrix function

$$A(h(\zeta)) = A_x + B \sum_{i,j=1}^{q,n} H_{ij} h_{ij}(\zeta_i) \quad (10)$$

According to (7), (8), (9), (10), the dynamics (6) can be rewritten under the form :

$$\sigma_\xi = \begin{bmatrix} \mathcal{A}(h(z)) - A_u K & A_u K \\ 0 & \mathcal{A}(h(\bar{z})) - LC \end{bmatrix} \xi + \begin{bmatrix} E_\omega \\ E_\omega - LD_\omega \end{bmatrix} \omega \quad (11)$$

where

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_q \end{bmatrix}, \quad \bar{z} = \begin{bmatrix} \bar{z}_1 \\ \vdots \\ \bar{z}_q \end{bmatrix}, \quad \xi = \begin{bmatrix} x \\ \varepsilon \end{bmatrix}$$

Condition (3) implies that the parameter vector $h(\cdot)$ belongs to the following convex bounded set :

$$\mathcal{H}_{q,n} = \left\{ h(\eta) \in \mathbb{R}^{q \times n} : a_{ij} \leq h_{ij}(\eta_i) \leq b_{ij}, \right. \\ \left. \forall \eta = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_q \end{bmatrix} \in \mathbb{R}^{q^n}, \eta_i \in \mathbb{R}^n \right\} \quad (12)$$

of which the set of vertices is given by :

$$\mathcal{V}_{\mathcal{H}_{q,n}} = \left\{ \Phi \in \mathbb{R}^{q \times n} : \Phi_{ij} \in \{a_{ij}, b_{ij}\} \right\}. \quad (13)$$

III. SYNTHESIS METHOD : SEPARATION RESULTS

We establish, in this section, some separation results for the class of nonlinear systems (1). The goal is to find conditions which ensure the robust asymptotic stability of the system (11) under the action of the observer-based linear static feedback (4b). On the other word, the problem consists to design the matrices L and K so that

$$\|\xi\|_{\mathcal{L}_2^{2n}} \leq \lambda \|\omega\|_{\mathcal{L}_2^s}, \text{ for } \xi(0) = 0. \quad (14)$$

where λ is the disturbance attenuation level to be determined. In the discrete-time case, we make ℓ_2 instead of \mathcal{L}_2 . The objective is to develop sufficient conditions which give the gains L and K separately. To do this, we propose, first, to study the \mathcal{H}_∞ asymptotic stability of the estimation error ε . This latter is then used to show the \mathcal{H}_∞ asymptotic stability of the state x around zero. Indeed, the system (11) may be decomposed as follows :

$$\sigma_x = \begin{bmatrix} \mathcal{A}(h(z)) - A_u K \\ A_u \end{bmatrix} \omega + \begin{bmatrix} E_\omega \\ E_\omega - LD_\omega \end{bmatrix} \omega \quad (15a)$$

$$\sigma_\varepsilon = \begin{bmatrix} \mathcal{A}(h(\bar{z})) - LC \\ E_\omega - LD_\omega \end{bmatrix} \omega \quad (15b)$$

where

$$\bar{\omega} = \begin{bmatrix} K\varepsilon \\ \omega \end{bmatrix}$$

Instead of (14), we propose to show independently the following :

$$\|\varepsilon\|_{\mathcal{L}_2^{2n}} \leq \lambda_1 \|\omega\|_{\mathcal{L}_2^s}, \text{ for } \varepsilon(0) = 0 \quad (16)$$

and

$$\|x\|_{\mathcal{L}_2^{2n}} \leq \lambda_2 \|\bar{\omega}\|_{\mathcal{L}_2^{m+s}}, \text{ for } x(0) = 0 \quad (17)$$

which lead to (14) with

$$\lambda = \sqrt{\lambda_1^2 + \lambda_2^2 (1 + \lambda_1^2 \lambda_{\max}(K^T K))}$$

where $\lambda_{\max}(K^T K)$ represents the maximum eigenvalue of the square matrix $K^T K$.

Remark 3.1: Here, we give some remarks on the definition of $\bar{\omega}$ and the \mathcal{H}_∞ criteria (14).

- In (15a), we have $\bar{\omega} \in \mathcal{L}_2^{s+m}$. Indeed, from (16) we have $\varepsilon \in \mathcal{L}_2^n$ which leads to $K\varepsilon \in \mathcal{L}_2^m$, hence $\bar{\omega} \in \mathcal{L}_2^{s+m}$.
- When $\xi(0) \neq 0$, the \mathcal{H}_∞ criteria (14) must be replaced by the following one :

$$\|\xi\|_{\mathcal{L}_2^{2n}} \leq \sqrt{\lambda^2 \|\omega\|_{\mathcal{L}_2^s}^2 + \gamma \|\xi(0)\|^2} \quad (18)$$

where γ is a positive constant to be determined.

Note that the \mathcal{H}_∞ problem (16)-(17) may be reduced to find Lyapunov functions V and W so that (see [4] for more details)

$$\dot{V} + \varepsilon^T \varepsilon - \lambda_1^2 \omega^T \omega < 0 \quad (19)$$

and

$$\dot{W} + x^T x - \lambda_2^2 \bar{\omega}^T \bar{\omega} < 0 \quad (20)$$

For the discrete-time case, we must use $\Delta V = V_{k+1} - V_k$ and $\Delta W = W_{k+1} - W_k$ instead of \dot{V} and \dot{W} , respectively. The following theorem provides some LMI conditions under which (19)-(20) hold.

Theorem 3.2: The \mathcal{H}_∞ criteria (14) is fulfilled under the action of the observer-based linear static feedback (4b) if there exist two positive scalar μ, ν and matrices $P = P^T > 0$, $S = S^T > 0$, X and Y of adequate dimensions so that the following LMI conditions are feasible :

- **The continuous-time case :** $\forall \Phi \in \mathcal{V}_{\mathcal{H}_{q,n}}$

1) $\min(\mu)$ subject to

$$\begin{bmatrix} \mathbb{M}(P, Y, \Phi) & PE_\omega - Y^T D_\omega \\ (*) & -\mu I_s \end{bmatrix} < 0 \quad (21)$$

where

$$\mathbb{M}(P, Y, \Phi) = A^T(\Phi)P - C^T Y + PA(\Phi) - Y^T C + I_n$$

2) $\min(\nu)$ subject to

$$\begin{bmatrix} \mathbb{N}(S, X, \Phi) & [A_u \ E_\omega] & S \\ (*) & -\nu I_{m+s} & 0 \\ (*) & (*) & -I_n \end{bmatrix} < 0 \quad (22)$$

where

$$\mathbb{N}(S, X, \Phi) = SA^T(\Phi) - XA_u^T + A(\Phi)S - A_u X^T$$

- **The discrete-time case :** $\forall \Phi \in \mathcal{V}_{\mathcal{H}_{q,n}}$

1) $\min(\mu)$ subject to

$$\begin{bmatrix} -P + I_n & 0 & \mathcal{A}^T(\Phi)P - C^TY \\ (\star) & -\mu I_s & E_\omega^T P - D_\omega^T Y \\ (\star) & (\star) & -P \end{bmatrix} < 0 \quad (23)$$

2) $\min(\nu)$ subject to

$$\begin{bmatrix} -S & 0 & \mathbb{L}(S, X, \Phi) & S \\ (\star) & -\nu I_{m+s} & [A_u \ E_\omega]^T & 0 \\ (\star) & (\star) & -S & 0 \\ (\star) & (\star) & (\star) & -I_n \end{bmatrix} < 0 \quad (24)$$

where

$$\mathbb{L}(S, X, \Phi) = SA^T(\Phi) - XA_u^T$$

The controller and the observer gains are computed, respectively by :

$$K = X^T S^{-1} \quad \text{and} \quad L = P^{-1} Y^T \quad (25)$$

and the disturbance attenuation level that we obtain by our approach is

$$\lambda^* = \sqrt{\mu^* + \nu^* (1 + \mu^* \lambda_{\max}(K^T K))} \quad (26)$$

where

$$\mu^* = \min(\mu), \quad \nu^* = \min(\nu)$$

Proof: For the lack of space, we give only the proof in the continuous-time case. We can use the same methodology to show the result for the discrete-time case.

Now, consider the Lyapunov function candidate $V = \varepsilon^T P \varepsilon$ for the system (15b) and $W = x^T S^{-1} x$ for (15a). After computing \dot{V} and \dot{W} , we obtain

$$\begin{aligned} V_\omega &= \varepsilon^T \left(\left[\mathcal{A}(h(\bar{z})) - LC \right]^T P + \right. \\ & P \left[\mathcal{A}(h(\bar{z})) - LC \right] + I_n \Big) \varepsilon \\ & + 2\varepsilon^T P \left(E_\omega - LD_\omega \right) \omega - \mu \omega^T \omega \end{aligned} \quad (27)$$

and

$$\begin{aligned} W_\omega &= x^T \left(\left[\mathcal{A}(h(z)) - A_u K \right]^T S^{-1} + \right. \\ & S^{-1} \left[\mathcal{A}(h(z)) - A_u K \right] + I_n \Big) x \\ & + 2x^T S^{-1} \left[A_u \ E_\omega \right] \bar{\omega} - \nu \bar{\omega}^T \bar{\omega} \end{aligned} \quad (28)$$

where

$$V_\omega \triangleq \dot{V} + \varepsilon^T \varepsilon - \mu \omega^T \omega$$

$$W_\omega \triangleq \dot{W} + x^T x - \nu \bar{\omega}^T \bar{\omega}$$

and

$$\mu = \lambda_1^2, \quad \nu = \lambda_2^2$$

Using the notation $Y = L^T P$, we deduce that

$$V_\omega = \begin{bmatrix} \varepsilon \\ \omega \end{bmatrix}^T \begin{bmatrix} \mathbb{M}(P, Y, h(\bar{z})) & PE_\omega - Y^T D_\omega \\ (\star) & -\mu I_s \end{bmatrix} \begin{bmatrix} \varepsilon \\ \omega \end{bmatrix} \quad (29)$$

From (21), the convexity principle leads to

$$\begin{bmatrix} \mathbb{M}(P, Y, h(\bar{z})) & PE_\omega - Y^T D_\omega \\ (\star) & -\mu I_s \end{bmatrix} < 0$$

which means that $V_\omega < 0$. From (28), we have $W_\omega < 0$ if

$$\begin{bmatrix} \mathbb{A}^T(h)S^{-1} + S^{-1}\mathbb{A}(h) + I_n & S^{-1} \begin{bmatrix} A_u & E_\omega \end{bmatrix} \\ (\star) & -\nu I_s \end{bmatrix} < 0$$

where

$$\mathbb{A}(h) = \mathcal{A}(h(z)) - A_u K.$$

By pre- and post multiplying the last inequality by $\begin{bmatrix} S & 0 \\ 0 & I_s \end{bmatrix}$, we deduce that $W_\omega < 0$ if

$$\begin{bmatrix} S\mathbb{A}^T(h) + \mathbb{A}(h)S + S^2 & \begin{bmatrix} A_u & E_\omega \end{bmatrix} \\ (\star) & -\nu I_s \end{bmatrix} < 0 \quad (30)$$

Using the Schur lemma and the notation $X = SK^T$, (30) is equivalent to

$$\begin{bmatrix} \mathbb{N}(S, X, h(z)) & \begin{bmatrix} A_u & E_\omega \end{bmatrix} & S \\ (\star) & -\nu I_s & 0 \\ (\star) & (\star) & -I_{n+s} \end{bmatrix} < 0 \quad (31)$$

Hence, from (22) the convexity principle leads to (31), which means that $W_\omega < 0$. Finally, to obtain best values of μ and ν , the convex optimization problems 1) – 2) are required. This ends the proof of Theorem 3.2. The proof for the discrete-time case may be obtained using similar manipulations. ■

IV. A NUMERICAL EXAMPLE

Hereafter, we provide a numerical example in order to show performances of the results. Consider the discrete-time model of the chaotic system of *Lorenz*. Using *Euler* discretization and adding an input in the dynamics as in [19], the system can be written under the form (1) with the following parameters :

$$A_x = \begin{bmatrix} 1 - 10T & 10T & 0 \\ 28T & 1 - T & 0 \\ 0 & 0 & 1 - \frac{8}{3}T \end{bmatrix}, \quad B = T \begin{bmatrix} 0 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$C = [1 \ 0 \ 0], \quad A_u = T \begin{bmatrix} 30 \\ 28 \\ 10 \end{bmatrix}$$

and

$$f(x) = [x_1 x_3 \ x_1 x_2]^T$$

where $T = 0.001s$ is the sampling period. Note that the Lorenz's system exhibits a chaotic behavior with to bounded trajectories, which means, according to the Remark 2.1, that the condition (3) holds. Thus, our approach is applicable.

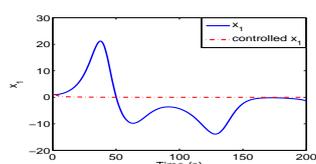
Assume that the output signal is affected, in a finite time interval, by a Gaussian distributed random signal with mean zero and standard deviation $\sigma = 0.05$. Using Theorem 3.2, we obtain, after solving the convex optimization problem 23-24

by Matlab LMI toolbox, the following solutions :

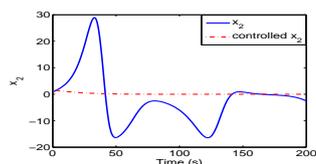
$$K^T = \begin{bmatrix} 2.7842 \\ 27.0817 \\ 2.0446 \end{bmatrix}, \quad L = \begin{bmatrix} 0.9959 \\ 0.6224 \\ 0.0056 \end{bmatrix} \quad \text{and} \quad \lambda^* = 2.1062$$

The simulation results are shown in figures 1 and 2. The initial values used for numerical simulations are

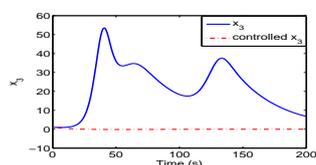
$$x_0 = [1 \quad 1 \quad 1]^T, \quad \hat{x}_0 = [-1 \quad -1 \quad -1]^T.$$



(a) Behavior of x_1



(b) Behavior of x_2



(c) Behavior of x_3

Fig. 1. Behaviors of the state variables without disturbances

V. CONCLUSION

In this paper, we presented a simple and systematic method to solve the problem of \mathcal{H}_∞ observer-based control for a class of nonlinear systems. The main advantage of this approach lies in the fact that the synthesis of the controller and the observer gains is reduced to solve two independent nonrestrictive LMIs without any equality constraint. Indeed, this latter, which is the obstacle of many observer-based control methods established in literature, is difficult to be solved for certain nonlinear systems with single input as shown in [7]. The asymptotic convergence of the state and the estimation error towards zero is also established using some mathematical tools such as the Barbalat's lemma and the convergence of series. Finally, the proposed approach is validated successfully by a numerical example.

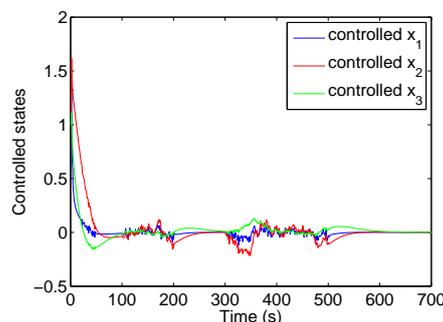


Fig. 2. The controlled state variables with $\sigma = 0.05$

REFERENCES

- [1] C. Kravaris, V. Sotiropoulos, C. Georgiou, N. Kazantzis, M. Q. Xiao, and A. J. Krener. Nonlinear observer design for state and disturbance estimation. *Systems and Control Letters*, 54:877–886, 2005.
- [2] N. Kazantzis and R. A. Wright. Nonlinear observer design in the presence of delayed output measurements. *Systems and Control Letters*, 54:877–886, 2005.
- [3] V. Sundarapandian. Reduced order observer design for discrete-time nonlinear systems. *Applied Mathematics Letters*, 19, 2006.
- [4] A. Zemouche, M. Boutayeb, and G. I. Bara. Observers for a class of Lipschitz systems with extension to \mathcal{H}_∞ performance analysis. *Systems & Control Letters*, 57(1):18–27, 2008.
- [5] S. Sundaram and C. N. Hadjicostis. Partial state observers for linear systems with unknown inputs. *Automatica*, doi:10.1016/j.automatica.2008.05.011, 2008.
- [6] C. P. Tan, F. Crusca, and M. Aldeen. Extended results on robust state estimation and fault detection. *Automatica*, 44, 2008.
- [7] S. Ibrir, Wen Fang Xie, and Chun-Yi Su. Observer-based control of discrete-time Lipschitzian non-linear systems : application to one-link flexible joint robot. *International Journal of Control*, 78(6):385–395, 2005.
- [8] A. Pisano and E. Usai. Output-feedback control of an underwater vehicle prototype by higher-order sliding modes. *Automatica*, 40, 2004.
- [9] R. Marino, P. Tomei, and C. M. Verrelli. A global tracking control for speed-sensorless induction motors. *Automatica*, 40, 2004.
- [10] R. Hedjar, R. Toumi, P. Boucher, and D. Dumur. Feedback nonlinear predictive control of rigid link robot manipulators. In *2002 American Control Conference*, Anchorage, AK, 2002.
- [11] Chang-Hua Lien. Robust observer-based control of systems with state perturbations via LMI approach. *IEEE Trans. on Automatic Control*, 49(8):1365–1370, 2004.
- [12] P. R. Pagilla, E. O. King, L. H. Dreinhoefer, and S. S. Garimella. Robust observer-based control of an aluminum strip processing line. *IEEE Trans. on Industry Applications*, 36(3):865–870, 2000.
- [13] A. N. Atassi and H. K. Khalil. Separation principle for the stabilization of a class of nonlinear systems. *IEEE Trans. on Automatic Control*, 44(9):1672–1687, 1999.
- [14] A. N. Atassi and H. K. Khalil. Separation results for the stabilization of nonlinear systems using different high-gain observer designs. *Systems & Control Letters*, 39(3):183–191, 2000.
- [15] A. N. Atassi and H. K. Khalil. A separation principle for the control of a class of nonlinear systems. *IEEE Trans. on Automatic Control*, 46(5):742–746, 2001.
- [16] W. Lohmiller and J.-J. E. Slotine. Stability and robustness analysis of nonlinear systems via contraction metrics and SOS programming. *Automatica*, 34(6):683–696, 1998.
- [17] E. M. Aylward, P. A. Parrilo, and J.-J. E. Slotine. Stability and robustness analysis of nonlinear systems via contraction metrics and SOS programming. *Automatica*, 44(8):2163–2170, 2008.
- [18] M. Farza, M. M'Saad, and L. Rossignol. Observer design for a class of MIMO nonlinear systems. *Automatica*, 40, 2004.
- [19] M. Boutayeb. Synchronization and input recovery in digital non-linear systems. *IEEE Trans. Circuits Syst. II:Express Briefs*, 51(8):393–399, 2004.

Mathematical study for a tuberculosis model with two differential infectivity and n classes of latent

Samuel Bowong, Yves Emvudu, Dany Pascal Moualeu and Jean Jules Tewa

Abstract—We consider a very general tuberculosis model with two differential infectivity, n classes of latent individuals and mass action incidence. This general system exhibits the traditional threshold behavior. There is always a globally asymptotically stable equilibrium state. Depending on the value of the basic reproduction ratio \mathcal{R}_0 , this state can be either endemic ($\mathcal{R}_0 > 1$), or infection-free ($\mathcal{R}_0 \leq 1$). The global stability of this model is derived through the use of Lyapunov stability theory and LaSalle's invariant set theorem. Computer simulations are given to illustrate analytical results.

Index Terms—Nonlinear dynamical systems, Epidemiological models, Tuberculosis models, Global stability.

I. INTRODUCTION

Tuberculosis (TB) is primarily a disease of the respiratory system with variable degrees of infectiousness. It can follow infection with the airborne bacteria germ *Mycobacterium tuberculosis*. Bacilli only live in the air for approximately 2 hours so individuals who have intense contact with TB bacilli in poorly ventilated areas are the most likely to become infected. Thus, TB morbidity and mortality rates are strongly affected by living conditions. Infectiousness of the source case, duration and frequency of exposure and characteristics of shared environments, all contribute to the overall risk of transmission [1-10]. It is also known that factors such as endogenous reactivation, emergence of multi-drug resistant TB, and increase in HIV incidence in the recent years call for improved control strategies for TB. Another issue that is essential to the epidemiology of TB is the exogenous re-infection, where a latently-infected individual acquires a new infection from another infectious individual (see [11-13] and references therein).

Many mathematical models for tuberculosis differentiate between individuals according to their history of infection. In particular, the population under investigation is subdivided into four epidemiological classes: susceptible individuals, latently infected individuals (those who are infected but not

infectious), infectious and the recovered or cured individuals [3,17]. Some researchers, however, model tuberculosis taking into consideration three of the four epidemiological classes (excluding the recovered class) [1,9,18], with the assumption that recovered individuals revert back to the latent class. These models, irrespective of the number of epidemiological classes, have given rise to interesting results both qualitatively and quantitatively. However, the division of a population into various compartments give rise to compartmental models (whose complexity increases with increasing compartments, interactions compartments, interactions and specific population characteristics such as age, periodicity, susceptibility, infectivity, etc..).

On the other hand, it is well known that the duration of latency varies greatly from case to case [14]. It is possible for a tuberculosis infection to become active within a few months of infection. It is also possible that activation may occur several years or decades after exposure has taken place. Until such time, the individual suffers no ill-effects of the disease, and cannot transmit the disease to others [14,15]. Also, the risk of activation seems to decrease over time [11,12]. One way in which this can be modeled is by including a sequence of several latent classes through which latently infected individuals can pass before entering the infectious class. Each latent class can be assigned its own activation rate. Thus, TB infected individuals, generally classified as "infective", play a major role in the transmission of TB. In this work, we divided the infective class into two subgroups with different properties: infectious and loss of slight. Indeed, in Sub-Saharan Africa, some infectious that begun their effective therapy in the hospital never return to the hospital for the sputum examinations for many reasons such as negligence, lack of information about TB, long duration of treatment regimen, poverty, mentality, etc... In this case, the health personal don't know their epidemiological status, i.e., if they are died, recovered or still infectious. We call these infective individuals loss of slight. It should be pointed out that according to the Direct Observation Therapy Strategy (DOTS) applied in most developed countries, a patient of pulmonary tuberculosis must make three sputum examinations during his treatment and will be considered cured when the last result of the examination of sputum is negative. According to the National Comity of Fight against Tuberculosis of Cameroon [38], about 10% of infectious that begun their therapy treatment never return to the hospital for the sputum examinations and then become loss of slight. Thus, this lack of epidemiological status of some patients can has some effects on the spread of TB. Indeed, what is happening with the

Samuel Bowong is with the Department of Mathematics and Computer Science, Faculty of Science, University of Douala, PO Box 24157 Douala, Cameroon, e-mail: sbowong@gmail.com.

Yves Emvudu is with the Department of Mathematics, Faculty of Science, University of Yaounde I, PO Box 812 Yaounde, Cameroon, e-mail: yemvudu@yahoo.fr

Dany Pascal Moualeu is with the Department of Mathematics, Faculty of Science, University of Yaounde I, PO Box 812 Yaounde, Cameroon, e-mail: danyemoualeu@gmail.com

Jean Jules Tewa is with the Department of Mathematics and Physics, National Advanced School of Engineering, University of Yaounde I, PO Box 8390 Yaounde, Cameroon, e-mail: tewajules@gmail.com.

Samuel Bowong and Jean Jules Tewa are with UMI 209 UMMISCO (IRD/UPMC), Yaoundé-Cameroun

loss of sight? How are these affect the dynamics of TB? Therefore, this epidemiological fact cannot be neglected in the mathematical modeling of TB. Unfortunately, a frequent focus in qualitative mathematical epidemiology is to determine the long-term dynamics exhibited by a given model. Global results of stability for the disease free equilibrium as well as the endemic equilibrium for epidemic models are not so common [3]. In recent years, Lyapunov methods have been used (see [19-28] and references therein). A Volterra-like function has been used in [22] to prove the global stability of the endemic equilibrium for SEIR models. For higher-dimensional systems, the most promising method may be that of Lyapunov methods. Nevertheless, in mathematical epidemiology, applications of Lyapunov's method tend to be for low-dimensional [22] or for systems with comparatively simple interactions between the different dimensions. As a consequence, the global stability of epidemiological models with different infectivity, many classes of latent and mass action incidence is an important issue.

In this paper, to explore the role of the lack of epidemiological status of some patients in the hospital on the dynamics of tuberculosis, we formulate a mathematical model taking into cognisance of variability of the duration of latency and that certain loss of sight can still infectious. We introduce a direct transfer from the class of susceptible toward the compartment of infectious. The reason is the co-infection with HIV infection. We address the global dynamics of the generalized tuberculosis model with two differential infectivity and n latent classes. To the best of authors knowledge, the global analysis of tuberculosis models with differential infectivity is not well discussed in the literature. The global dynamics of the model is resolved through the use of Lyapunov functions. We use the same Lyapunov functions as those used recently in Refs. [19-28] to demonstrate the global stability of the endemic equilibrium of SEIR, SEIS, and SIR models. The results in this paper are notable in that very few restrictions are placed on the compartmental structure of the model, except that the susceptible and recovered populations can each be represented by a scalar variable. However, this class of TB epidemiological models can be extend to many class of infective individuals. Numerical studies are presented to validate analytical results.

II. THE MODEL

We consider a finite population of N people. We assume that latently infected individuals (inactive TB) becomes infectious (active TB) after a variable (typically long) latency period. At any given time, an individual is in one of the following states: susceptible, latently infected with n stage (i.e., exposed to TB but are not infectious), infectious (i.e., have active TB), recovered (i.e., not suffer of disease) and loss of sight (i.e., the health personal don't know their epidemiological status). We will denote these states by S , E , I , L and R , respectively. All recruitment is into the susceptible class, and occurs at a constant rate Λ . Transmission of M. tuberculosis occurs following adequate contact between a susceptible and infectious individual or loss sight who continue to suffer of disease. We use the standard mass balance incidence expressions $\beta_1 SI$ and

$\beta_2 SL$ to indicate successful transmission of M. tuberculosis due to nonlinear contact dynamics in the population. On adequate contact with infectious or loss of sight who continue to suffer of disease, a susceptible individual becomes infected but yet infectious. This individual remains in the latently infected class for a certain latent period through n stages (E_1, \dots, E_n) before infectious. Once latently infected with M. Tuberculosis, an individual will remain so for life unless reactivation occurs. To account for treatment, we define $r_i E_i$ as the fraction of latently infected individuals receiving effective chemoprophylaxis. We assume that chemoprophylaxis of latently infected individuals E_i reduces their reactivation at a constant rate r_i . Thus, a fraction $1 - r_i$ of latently infected individuals who does not received effective chemoprophylaxis progress to the next stage of latently infected class with a rate constant k_i . We assume that latently infected individuals leave the subclass E_i to the infectious class I at rate α_i . After receiving an effective therapy, infectious can spontaneously recover from the disease with a constant rate r , entering the recovered class R . Recovered individuals still have the bacterium in their body and can undergo a reactivation of the disease with rate γ . We also assume that among the fraction $1 - r$ of infectious who does not recovered, some of them who begun their treatment will not return to the hospital for the examination of sputum at a rate ϕ and enters the class of loss of sight L . After some time, some of them will continue to suffer of the disease and will return to the hospital at a constant rate δ . Because of natural recover and traditional medicine (practiced in Sub-Saharan Africa), a fraction $1 - \delta$ of loss of sight who does not continue to have disease can recover at a constant rate ε and enters the recovered class R . The rate constant for non-disease related death is μ , thus $1/\mu$ is the average lifetime. Latently infected, infectious and loss sight have addition death rates due to infection and disease with rates constant d_i , d_I and d_L , respectively.

Thus, the corresponding transfer diagram is

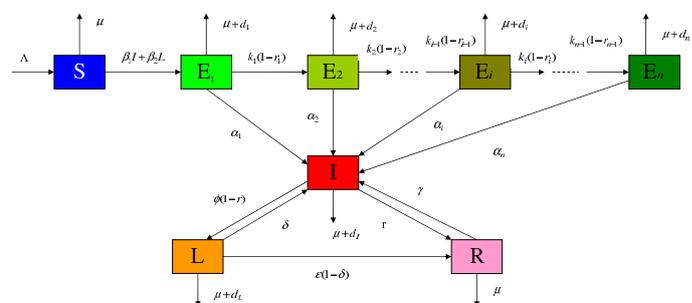


Fig. 1. A transfer diagram for a general TB model with two differential infectivity and n latent classes.

The corresponding equations are

$$\left\{ \begin{array}{l} \dot{S} = \Lambda - S(\beta_1 I + \beta_2 L) - \mu S, \\ \dot{E}_1 = S(\beta_1 I + \beta_2 L) - [\mu + d_1 + \alpha_1 + k_1(1 - r_1)]E_1, \\ \vdots \\ \dot{E}_i = k_{i-1}(1 - r_{i-1})E_{i-1} - [\mu + d_i + \alpha_i + k_i(1 - r_i)]E_i, \\ \quad i = 2, \dots, n-1, \\ \vdots \\ \dot{E}_n = k_{n-1}(1 - r_{n-1})E_{n-1} - (\mu + d_n + \alpha_n)E_n, \\ \dot{I} = \sum_{i=1}^n \alpha_i E_i + \gamma R + \delta L - [\mu + d_I + r + \phi(1 - r)]I, \\ \dot{L} = \phi(1 - r)I - [\mu + d_L + \delta + \varepsilon(1 - \delta)]L, \\ \dot{R} = rI + \varepsilon(1 - \delta)L - (\mu + \gamma)R. \end{array} \right. \quad (1)$$

III. MATHEMATICAL ANALYSIS OF THE MODEL

The system (1) can be written in the following compact form:

$$\left\{ \begin{array}{l} \dot{x} = \varphi(x) - x\langle\beta | y\rangle, \\ \dot{y} = x\langle\beta | y\rangle B + A y, \end{array} \right. \quad (2)$$

where $x = S \in \mathbb{R}_{\geq 0}$ is a state representing the compartment of non transmitting individuals (susceptible), $y = (E_1, \dots, E_n, I, L, R)^T = (y_1, \dots, y_n, y_{n+1}, y_{n+2}, y_{n+3})^T \in \mathbb{R}_{\geq 0}^{n+3}$ is the vector representing the state compartment of different infected individuals (latently infected, infectious, loss of slight and recovered individuals), $\varphi(x) = \Lambda - \mu x$ is a function that depends on x , $\beta = (0, \dots, 0, \beta_1, \beta_2, 0) \in \mathbb{R}^{n+3}$ and $B = (1, 0, \dots, 0)^T \in \mathbb{R}^{n+3}$, $\langle \cdot | \cdot \rangle$ is the usual scalar product in \mathbb{R}_+^{n+3} and A is a $(n+3) \times (n+3)$ constant matrix defined as

$$A = \begin{bmatrix} -a_1 & \dots & 0 & 0 & 0 & 0 \\ k_1(1 - r_1) & \dots & 0 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \dots & \dots & \vdots & \vdots & \vdots \\ 0 & \dots & -a_n & 0 & 0 & 0 \\ \alpha_1 & \dots & \alpha_n & -a_I & \delta & \gamma \\ 0 & \dots & 0 & \phi(1 - r) & -a_L & 0 \\ 0 & \dots & 0 & r & \varepsilon(1 - \delta) & -a_R \end{bmatrix},$$

with

$$a_i = \mu + d_i + \alpha_i + k_i(1 - r_i), \quad i = 1, \dots, n-1,$$

$$a_n = \mu + d_n + \alpha_n, \quad a_I = \mu + d_I + r + \phi(1 - r),$$

$$a_L = \mu + d_L + \delta + \varepsilon(1 - \delta) \quad \text{and} \quad a_R = \mu + \gamma.$$

It should be pointed out that A is a Metzler matrix, that is, a matrix with off-diagonal entries nonnegative [29,30].

With respect to system (2), we made the following hypothesis:

Hypothesis: For the system

$$\dot{x} = \varphi(x),$$

there exists a unique point $x_0 > 0$ such that

$$\left\{ \begin{array}{l} \varphi(x_0) = 0, \quad \varphi(x) > 0 \quad \text{for} \quad 0 \leq x < x_0, \\ \text{and} \quad \varphi(x) < 0 \quad \text{for} \quad x > x_0. \end{array} \right. \quad (3)$$

A. Positive invariance of the nonnegative orthant

We have the following result:

Proposition 1: The nonnegative orthant $\mathbb{R}_{\geq 0}^{n+4}$ is positively invariant for the system (2).

Proof: The system (2) may be rewritten as

$$\left\{ \begin{array}{l} \dot{x} = \varphi(x) - x\langle\beta | y\rangle, \\ \dot{y} = (x B \beta^T + A) y, \end{array} \right. \quad (4)$$

where $(x B \beta^T + A)$ is a Metzler matrix since $x \geq 0$. With the hypothesis (3), $\varphi(0) > 0$ and the half line $\mathbb{R}_{\geq 0}$ is positively invariant by $\dot{x} = \varphi(x) - x\langle\beta | y\rangle$ since a linear Metzler system let invariant the nonnegative orthant [29,30]. This proves the positive invariance of the nonnegative orthant $\mathbb{R}_{\geq 0}^{n+4}$ for the system (2) and this concludes the proof. \square

B. Boundedness and dissipativity of the trajectories

The trajectories of the model (1) are bounded and dissipative. Indeed, adding all equations of the model (1), one has $\dot{N}(t) = \Lambda - \mu \left(S + \sum_{i=1}^n E_i + I + L + R \right) (t) - \sum_{i=1}^n d_i E_i(t) - d_1 I(t) - d_2 L(t)$. Thus, one can deduce that $\dot{N}(t) \leq \Lambda - \mu N(t)$. It then follows that $\lim_{t \rightarrow +\infty} N(t) = \frac{\Lambda}{\mu}$. It is straightforward to prove that for $\rho \geq 0$ the simplex:

$$\Omega_\rho = \left\{ (S, E_i, I, R, L) \in \mathbb{R}_{\geq 0}^{n+4}, \quad N(t) \leq \frac{\Lambda}{\mu} + \rho \right\}, \quad (5)$$

is a compact forward invariant set for the system (1) and that for $\rho > 0$ this set is absorbing. So, we limit our study to this simplex for $\rho > 0$.

C. Basic reproduction number

Global behavior for this model crucially depends on the basic reproduction number, that is, an average number of secondary cases produced by a single infective individual which is introduced into an entirely susceptible population. The system (2) has an evident equilibrium $P_0 = (x_0, 0)$ with $x_0 = \Lambda/\mu$ when there is no disease. This equilibrium point is the disease free equilibrium. We calculate the basic reproduction ratio, \mathcal{R}_0 , using the next generation approach, developed in van den Driessche and Watmough [32]. Using the techniques reported in Refs. [31-34], the next generation matrix is given by $M = x_0 \langle \beta | (-A^{-1}) B \rangle$ and the basic reproduction ratio \mathcal{R}_0 is the spectral radius of the matrix M .

Since M is a rank one matrix, the only nonzero eigenvalue is given by

$$\mathcal{R}_0 = x_0 \langle \beta | (-A^{-1}) B \rangle. \quad (6)$$

We use the expression $(-A^{-1})$ to put the emphasis on the fact that $(-A^{-1}) \geq 0$ because the matrix A is Metzler stable. Using the expressions of β and B defined as in Eq. (2) and after the computation of $(-A^{-1})$, the basic reproduction ratio (6) may be rewritten as

$$\mathcal{R}_0 = \frac{a_1 a_R \mathcal{R}_0^1 [\beta_1 a_L + \beta_2 \phi (1-r)] x_0}{\phi (1-r) \mathcal{R}_0^2 + a_L [(\mu + \gamma)(\mu + d_I) + \mu r]}, \quad (7)$$

where

$$\mathcal{R}_0^1 = \sum_{i=1}^n \left[\alpha_i \frac{\prod_{l=1}^{i-1} k_l (1-r_l)}{\prod_{j=2}^i a_j} \right] \quad \text{and}$$

$$\mathcal{R}_0^2 = \mu[\mu + d_L + \varepsilon(1 - \delta)] + \gamma(\mu + d_L).$$

Now, let us determine, using the threshold quantity \mathcal{R}_0 , whether or not loss of sight can influence the propagation of tuberculosis in the considered population. Suppose that there is no loss of sight in the considered population, that is, $\phi = \delta = \varepsilon = 0$. In this case the basic reproduction ratio (7) becomes

$$\mathcal{R}_0^* = \frac{\beta_1 a_1 a_R \mathcal{R}_0^1 x_0}{(\mu + \gamma)(\mu + d_I) + \mu r}. \quad (8)$$

From Eqs. (7) and (8), one has

$$\mathcal{R}_0^* < \mathcal{R}_0, \quad (9)$$

which implies that the parameters ϕ , δ and ε influence the propagation of the disease in the host population. This means that the lost of sight play an important role in the propagation of TB. Also, it is evident from (7) that

$$\lim_{\phi \rightarrow \infty} \mathcal{R}_0 = \frac{\beta_2 a_1 a_R \mathcal{R}_0^1 x_0}{\mathcal{R}_0^2} > 0. \quad (10)$$

Thus, a sufficiently effective TB treatment program can lead to effective disease control if it results in making the right-hand side of (10) less than unity.

Further sensitivity analysis on the rate at which infectious become loss of sight is carried out by computing the partial derivatives of \mathcal{R}_0 with respect to the parameter ϕ giving

$$\frac{\partial \mathcal{R}_0}{\partial \phi} = \frac{a_1 a_R a_L \mathcal{R}_0^1 x_0 [\beta_2 [(\mu + \gamma)(\mu + d_I) + \mu r] - \beta_1 \mathcal{R}_0^2]}{[\phi (1-r) \mathcal{R}_0^2 + a_L [(\mu + \gamma)(\mu + d_I) + \mu r]]^2}. \quad (11)$$

It then comes from the above equation that $\frac{\partial \mathcal{R}_0}{\partial \phi} < 0$ if

$$\beta_2 < \Delta = \frac{\beta_1 \mathcal{R}_0^2}{(\mu + \gamma)(\mu + d_I) + \mu r}. \quad (12)$$

Thus, the rate at which infectious become loss of sight will have positive impact in reducing the propagation of TB only if $\beta_2 < \Delta$. Such a rate at which infectious become loss of sight will fail to reduce TB propagation if $\beta_2 = \Delta$, and will have detrimental impact in the considered population (increase \mathcal{R}_0) if $\beta_2 > \Delta$. This result is summarized below:

Lemma 1: The rate at which infectious become loss of sight ϕ will have positive impact if $\beta_2 < \Delta$, no impact if $\beta_2 = \Delta$ and will have detrimental impact if $\beta_2 > \Delta$ on the propagation of TB in the considered population.

It is worth emphasizing that the quantity Δ decreases when the treatment parameter of infectious r increases. So, this quantity can be made as small as possible by increasing the treatment parameter of infectious. Note that if the condition (12) does not hold (in this case the treatment parameter of infectious is small), then the use of the corresponding treatment strategy would increase TB propagation in the considered population (since $\mathcal{R}_0 > 1$). That is, the use of drug will increase the disease propagation if it fails to reduce the infectiousness of those treated below a certain threshold ($\beta_2 < \Delta$). So, in order to better control the disease, we have to take care of all the infectious individuals in the center of health to avoid loss of sight and to have a high value of the treatment parameter. This is, not the case in developing countries where the organization of health centers is practically non-existent.

D. Equilibria

The system (1) has a trivial equilibrium when there is no disease. This equilibrium is called the disease free equilibrium and is given by $P_0 = (x_0, 0) = (S_0, 0, \dots, 0)$, where $S_0 = x_0 = \frac{\Lambda}{\mu}$. The disease free equilibrium exists and is locally asymptotically stable when $\mathcal{R}_0 \leq 1$. But this equilibrium is unstable when $\mathcal{R}_0 > 1$. A new equilibrium different from the P_0 exists when the disease persists. This equilibrium is called the endemic equilibrium.

Let $P^* = (x^*, y^*)$ be the positive endemic equilibrium of model (2). Then, the positive endemic equilibrium (steady state with $y^* > 0$) can be obtained by setting the right hand side of all equations in the model (2) equal to zero, that is,

$$\begin{cases} \varphi(x^*) - x^* \langle \beta | y^* \rangle = 0, \\ x^* \langle \beta | y^* \rangle B + A y^* = 0. \end{cases} \quad (13)$$

From the second equation of (13), one has $y^* = x^* (-A^{-1}) B \langle \beta | y^* \rangle$, and replacing in $\langle \beta | y^* \rangle$ yields

$$\langle \beta | y^* \rangle = x^* \langle \beta | (-A^{-1}) B \rangle \langle \beta | y^* \rangle.$$

The case $\langle \beta | y^* \rangle = 0$ implies that $\varphi(x^*) = 0$ and $-A y^* = 0$. Since A is nonsingular, this gives the disease free equilibrium P_0 . For the other case, simplifying by $\langle \beta | y^* \rangle$ gives

$$x^* = \frac{1}{\langle \beta | (-A^{-1}) B \rangle} = \frac{x_0}{\mathcal{R}_0} > 0.$$

With $\mathcal{R}_0 > 1$, one has $x^* < x_0$, $\varphi(x^*) > 0$ and

$$y^* = (-A^{-1}) B \varphi(x^*).$$

Hence, the model (2) has a unique endemic equilibrium $P^* = (x^*, y^*)$ where x^* and y^* are given by

$$x^* = \frac{1}{\langle \beta | (-A^{-1}) B \rangle} \quad \text{and} \quad y^* = (-A^{-1}) B \varphi(x^*). \quad (14)$$

Now, using the expressions of β and B defined as in Eq. (2) and the expression of $(-A^{-1})$ obtained after a calculation, the endemic equilibrium of the model (1) is given by

$$\left\{ \begin{array}{l} S^* = \frac{x_0}{\mathcal{R}_0} = \frac{\Lambda}{\mu \mathcal{R}_0}, \\ E_1^* = \frac{\Lambda(\mathcal{R}_0 - 1)}{a_1 \mathcal{R}_0}, \\ E_i^* = \frac{\prod_{l=1}^{i-1} k_l(1-r_l)}{\prod_{j=2}^i a_j} \frac{\Lambda(\mathcal{R}_0 - 1)}{a_1 \mathcal{R}_0}, \text{ for } i = 2, 3, \dots, n, \\ I^* = \frac{a_L \mu (\mathcal{R}_0 - 1)}{\beta_1 a_L + \beta_2 \phi(1-r)}, \\ L^* = \frac{\phi(1-r) \mu (\mathcal{R}_0 - 1)}{\beta_1 a_L + \beta_2 \phi(1-r)}, \\ R^* = \frac{\mu[r a_L + \varepsilon(1-\delta)\phi(1-r)](\mathcal{R}_0 - 1)}{a_R [\beta_1 a_L + \beta_2 \phi(1-r)]}. \end{array} \right. \quad (15)$$

Thus, we have established the following result.

Lemma 2: When $\mathcal{R}_0 > 1$, then the system (1) has a unique endemic equilibrium defined as in Eq. (15).

E. Global stability of the disease free equilibrium

We have the following result about the global stability of the disease free equilibrium P_0 .

Theorem 1: If $\mathcal{R}_0 \leq 1$, then the system (1) has no positive equilibrium states and the disease free equilibrium P_0 is globally asymptotically stable in Ω_ρ . This implies the global asymptotic stability of the disease free equilibrium P_0 on the nonnegative orthant $\mathbb{R}_{\geq 0}^{n+4}$. This means that the disease naturally dies out.

Proof: Let us consider the following LaSalle-Lyapunov candidate function:

$$V(x, y) = \frac{1}{x_0}(x - x_0 \ln x) + \beta^T (-A^{-1}) y - \frac{1}{x_0}(x_0 - x_0 \ln x_0) \quad (16)$$

It is easy to see that at the disease free equilibrium P_0 , the fonction $V(x, y)$ reaches its global minimum in Ω_ρ , and hence $V(x, y)$ is a Lyapunov fonction since we know that $\beta^T (-A^{-1}) > 0$. Its time derivative along the trajectories of

(2) satisfies

$$\begin{aligned} \dot{V}(x, y) &= \frac{1}{x_0} \left[\varphi(x) - x \langle \beta | y \rangle - \frac{x_0}{x} \varphi(x) + x_0 \langle \beta | y \rangle \right] \\ &+ \beta^T (-A^{-1}) x \langle \beta | y \rangle B - \beta^T y, \\ &= \frac{1}{x_0} \left[\frac{(x - x_0)}{x} \varphi(x) - x \beta^T y + x_0 \beta^T y \right] \\ &+ x \beta^T y \beta^T (-A^{-1}) B - \beta^T y, \\ &= \frac{(x - x_0)}{x_0 x} \varphi(x) + \frac{x \beta^T y}{x_0} [x_0 \beta^T (-A^{-1}) B - 1], \\ &= \frac{(x - x_0)}{x_0 x} \varphi(x) + \frac{x \beta^T y}{x_0} (\mathcal{R}_0 - 1). \end{aligned} \quad (17)$$

Recalling that at the disease free equilibrium $\Lambda = \mu x_0$ so that $\varphi(x) = \mu(x_0 - x)$. In this case, Eq. (17) becomes

$$\dot{V}(x, y) = \frac{-\mu(x - x_0)^2}{x_0 x} + \frac{x}{x_0} (\beta_1 I + \beta_2 L) (\mathcal{R}_0 - 1). \quad (18)$$

Thus, $\mathcal{R}_0 \leq 1$ ensures that $\dot{V}(x, y) \leq 0$ for all $x, y \geq 0$, and that $\dot{V}(x, y) = 0$ holds when $\mathcal{R}_0 = 1$ for $x = x_0$. It is easy to verify that the disease free equilibrium state P_0 is the only fixed point of the system in the space $x = x_0$, and hence the system has no equilibria in Ω_ρ apart from P_0 . Then, by the Lyapunov-LaSalle's asymptotic stability theorem [35-37], the equilibrium state P_0 is globally asymptotically stable in Ω_ρ . This proves the global asymptotic stability on Ω_ρ and then in the nonnegative orthant $\mathbb{R}_{\geq 0}^{n+4}$ (see [37], Theorem 3.7.11, page 346). This achieves the proof. \square

F. Global stability of the endemic equilibrium

The global stability of the endemic equilibrium is given by Theorem 2, stated below.

Theorem 2: If $\mathcal{R}_0 > 1$, then the positive endemic equilibrium state P^* of the model (1) is globally asymptotically stable on the set Ω_ρ when

$$\frac{L^*}{L} \leq \frac{R^*}{R} \leq \frac{I^*}{I} \leq 1 \text{ and } \frac{E_{i+1}^*}{E_{i+1}} \leq \frac{I^*}{I} \leq \frac{E_i^*}{E_i} \leq 1, \quad (19)$$

$$i = 1, \dots, n - 1.$$

Proof: Consider the following Lyapunov function:

$$\begin{aligned} U(S, E_i, L, R) &= (S - S^* \ln S) + \sum_{i=1}^n A_i (E_i - E_i^* \ln E_i) \\ &+ B(I - I^* \ln I) + C(L - L^* \ln L) + D(R - R^* \ln R), \end{aligned} \quad (20)$$

where A_i, B, C and D are positive constants to be determined

later. Differentiating this function with respect to time yields

$$\begin{aligned}
\dot{U} &= \left(1 - \frac{S^*}{S}\right) \dot{S} + \sum_{k=1}^n A_k \left(1 - \frac{E_k^*}{E_k}\right) \dot{E}_k \\
&+ B \left(1 - \frac{I^*}{I}\right) \dot{I} + C \left(1 - \frac{L^*}{L}\right) \dot{L} \\
&+ D \left(1 - \frac{R^*}{R}\right) \dot{R} \\
&= \left(1 - \frac{S^*}{S}\right) (\Lambda - \beta_1 SI - \beta_2 SL - \mu S) \\
&+ A_1 \left(1 - \frac{E_1^*}{E_1}\right) (\beta_1 SI + \beta_2 SL - a_1 E_1) \\
&+ \sum_{i=2}^n A_i \left(1 - \frac{E_i^*}{E_i}\right) [k_{i-1}(1 - r_{i-1})E_{i-1} - a_i E_i] \\
&+ B \left(1 - \frac{I^*}{I}\right) \left[\sum_{i=1}^n \alpha_i E_i + \gamma R + \delta L - a_I I \right] \\
&+ C \left(1 - \frac{L^*}{L}\right) [\phi(1 - \delta)I - a_L L] \\
&+ D \left(1 - \frac{R^*}{R}\right) [rI + \varepsilon(1 - \delta)L - a_R R] \\
&= \left(1 - \frac{S^*}{S}\right) (\Lambda - \beta_1 SI - \beta_2 SL - \mu S) \\
&+ A_1 \beta_1 SI + A_1 \beta_2 SL - A_1 a_1 E_1 \\
&- A_1 \beta_1 SI \frac{E_1^*}{E_1} - A_1 \beta_2 SL \frac{E_1^*}{E_1} + A_1 a_1 E_1^* \\
&+ \sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1} - \sum_{i=2}^n A_i a_i E_i \\
&- \sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1} \frac{E_i^*}{E_i} + \sum_{i=2}^n A_i a_i E_i^* \\
&+ B \sum_{i=1}^n \alpha_i E_i + B \gamma R + B \delta L - B a_I I \\
&- B \sum_{i=1}^n \alpha_i E_i \frac{I^*}{I} - B \gamma R \frac{I^*}{I} - B \delta L \frac{I^*}{I} \\
&+ B a_I I^* + C \phi(1 - r)I - C a_L L \\
&- C \phi(1 - r)I \frac{L^*}{L} + C a_L L^* + D r I \\
&+ D \varepsilon(1 - \delta)L - D a_R R - D r I \frac{R^*}{R} \\
&- D \varepsilon(1 - \delta)L \frac{R^*}{R} + D a_R R^*,
\end{aligned} \tag{21}$$

where a_i , a_I , a_L and a_R are defined as in Eq. (2). By considering Eq. (1) at the positive endemic equilibrium $P^* = (S^*, E_i^*, I^*, L^*, R^*)$, one has

$$\begin{cases}
\Lambda &= \beta_1 S^* I^* + \beta_2 S^* L^* + \mu S^*, \\
a_1 E_1^* &= \beta_1 S^* I^* + \beta_2 S^* L^*, \\
&\vdots \\
a_i E_i^* &= k_{i-1} (1 - r_{i-1}) E_{i-1}^*, \quad i = 2, \dots, n, \\
a_I I^* &= \sum_{i=1}^n \alpha_i E_i^* + \gamma R^* + \delta L^*, \\
a_L L^* &= \phi(1 - r) I^*, \\
a_R R^* &= r I^* + \varepsilon(1 - \delta) L^*.
\end{cases} \tag{22}$$

After plugging Eq. (22) into Eq. (21), one obtains

$$\begin{aligned}
\dot{U} &= -\mu \frac{(S - S^*)^2}{S} + (A_1 - 1)(\beta_1 SI + \beta_2 SL) \\
&+ \left(1 - \frac{S^*}{S}\right) (\beta_1 S^* I^* + \beta_2 S^* L^*) \\
&+ A_1 \beta_1 S^* I^* \left(1 - \frac{S}{S^*} \frac{I}{I^*} \frac{E_1}{E_1^*}\right) \\
&+ A_1 \beta_2 S^* L^* \left(1 - \frac{S}{S^*} \frac{L}{L^*} \frac{E_1}{E_1^*}\right) \\
&+ \sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1}^* \left(1 - \frac{E_{i-1}}{E_{i-1}^*} \frac{E_i}{E_i^*}\right) \\
&+ B \sum_{i=1}^n \alpha_i E_i^* \left(1 - \frac{E_i}{E_i^*} \frac{I}{I^*}\right) + B \gamma R^* \left(1 - \frac{R}{R^*} \frac{I}{I^*}\right) \\
&+ B \delta L^* \left(1 - \frac{L}{L^*} \frac{I}{I^*}\right) + C \phi(1 - r) I^* \left(1 - \frac{I}{I^*} \frac{L}{L^*}\right) \\
&+ D r I^* \left(1 - \frac{I}{I^*} \frac{R}{R^*}\right) + D \varepsilon(1 - \delta) L^* \left(1 - \frac{L}{L^*} \frac{R}{R^*}\right) \\
&+ (-D a_R + B \gamma) R + [\beta_1 S^* - B a_I + C \phi(1 - r) + D r] I \\
&+ [\beta_2 S^* - C a_L + D \varepsilon(1 - \delta) + B \delta] L \\
&+ \sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1} - a_n A_n E_n \\
&+ B \sum_{i=1}^n \alpha_i E_i - \sum_{i=1}^n a_i A_i E_i.
\end{aligned} \tag{23}$$

Now, let $(u_1, u_2, u_3, u_4, v_i) = \left(\frac{S^*}{S}, \frac{I^*}{I}, \frac{L^*}{L}, \frac{R^*}{R}, \frac{E_i^*}{E_i}\right)$. Re-

marking that

$$\sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1} = \sum_{i=1}^{n-1} A_{i+1} k_i (1 - r_i) E_i,$$

then, Eq. (23) becomes

$$\begin{aligned} \dot{U} = & -\mu \frac{(S - S^*)^2}{S} + (A_1 - 1)(\beta_1 S I + \beta_2 S L) \\ & + (1 - u_1)(\beta_1 S^* I^* + \beta_2 S^* L^*) \\ & + A_1 \beta_1 S^* I^* \left(1 - \frac{v_1}{u_1 u_2}\right) \\ & + A_1 \beta_2 S^* L^* \left(1 - \frac{v_1}{u_1 u_3}\right) \\ & + A_2 k_1 (1 - r_1) E_1^* \left(1 - \frac{v_2}{v_1}\right) \\ & + \sum_{i=2}^{n-1} A_{i+1} k_i (1 - r_i) E_i^* \left(1 - \frac{v_{i+1}}{v_i}\right) \\ & + B \sum_{i=1}^n \alpha_i E_i^* \left(1 - \frac{u_2}{v_i}\right) + B \gamma R^* \left(1 - \frac{u_2}{u_4}\right) \\ & + B \delta L^* \left(1 - \frac{u_2}{u_3}\right) + C \phi (1 - r) I^* \left(1 - \frac{u_3}{u_2}\right) \\ & + D r I^* \left(1 - \frac{u_4}{u_2}\right) + D \varepsilon (1 - \delta) L^* \left(1 - \frac{u_4}{u_3}\right) \\ & + [\beta_1 S^* - B a_I + C \phi (1 - r) + D r] I \\ & + [\beta_2 S^* - C a_L + D \varepsilon (1 - \delta) + B \delta] L \\ & + (-D a_R + B \gamma) R \\ & + \sum_{i=1}^{n-1} [-a_i A_i + B \alpha_i + A_{i+1} k_i (1 - r_i)] E_i \\ & + (-a_n A_n + B \alpha_n) E_n. \end{aligned} \tag{24}$$

The coefficients A_i , B , C and D are chosen such that the coefficients of I , L , R and E_i are equal to zero, that is,

$$\begin{cases} A_1 - 1 = 0, \\ \beta_1 S^* - B a_I + C \phi (1 - r) + D r = 0, \\ \beta_2 S^* - C a_L + D \varepsilon (1 - \delta) + B \delta = 0, \\ -D a_R + B \gamma = 0, \\ -a_i A_i + B \alpha_i + A_{i+1} k_i (1 - r_i) = 0, \\ i = 1, 2, \dots, n - 1, \\ -A_n a_n + B \alpha_n = 0. \end{cases} \tag{25}$$

At this point, it is important to mention that when the second equation of (25) is satisfied, then all equations of (25) are also satisfied. Indeed, multiplying the second equation of (25) by I^* and using the expression of $a_I I^*$ defined as in Eq. (22) yields

$$\begin{aligned} \beta_1 S^* I^* - B a_I I^* + C \phi (1 - r) I^* + D r I^* = & -B \sum_{i=1}^n \alpha_i E_i^* \\ & - B \gamma R^* - B \delta L^* + \beta_1 S^* I^* + C \phi (1 - r) I^* + D r I^*. \end{aligned} \tag{26}$$

On the other hand, from Eq. (22), one has

$$\begin{aligned} a_1 A_1 E_1^* - A_1 \beta_1 S^* I - A_1 \beta_2 S^* L^* + \sum_{i=2}^n a_i A_i E_i^* \\ - \sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1}^* + C a_L L^* - C \phi (1 - r) I^* \\ + D a_R R^* - D r I^* - D \varepsilon (1 - \delta) L^* = 0. \end{aligned} \tag{27}$$

Adding Eq. (27) in the right hand side of Eq. (26) and noting that

$$\begin{aligned} a_1 A_1 E_1^* + \sum_{i=2}^n a_i A_i E_i^* - \sum_{i=2}^n A_i k_{i-1} (1 - r_{i-1}) E_{i-1}^* = \\ \sum_{i=1}^{n-1} [a_i A_i - A_{i+1} k_i (1 - r_i)] E_i^* + a_n A_n E_n^*, \end{aligned}$$

gives

$$\begin{aligned} \beta_1 S^* I^* - B a_I I^* + C \phi (1 - r) I^* + D r I^* = \\ (1 - A_1)(\beta_1 I^* + \beta_2 L^*) S^* + (-B \gamma + D a_R) R^* \\ + [-\beta_2 S^* - B \delta - D \varepsilon (1 - \delta) + C a_L] L^* \\ + \sum_{i=1}^{n-1} [-B \alpha_i + a_i A_i - A_{i+1} k_i (1 - r_i)] E_i^* \\ + (-B \alpha_n + a_n A_n) E_n^* = 0. \end{aligned} \tag{28}$$

Therefore, we only consider the following system of equations:

$$\begin{cases} A_1 - 1 = 0, \\ \beta_2 S^* - C a_L + D \varepsilon (1 - \delta) + B \delta = 0, \\ -D a_R + B \gamma = 0, \\ -a_i A_i + B \alpha_i + A_{i+1} k_i (1 - r_i) = 0, \\ i = 1, \dots, n - 1, \\ -A_n a_n + B \alpha_n = 0. \end{cases} \tag{29}$$

Using the fourth equation of (29), one can easily prove that

$$A_i = \prod_{j=1}^{i-1} \frac{a_j}{k_j (1 - r_j)} - \left(\sum_{l=1}^{i-1} \frac{\alpha_l}{k_l (1 - r_l)} \prod_{k=l+1}^{i-1} \frac{a_k}{k_k (1 - r_k)} \right) B, \\ i = 2, \dots, n.$$

Now using the expression of A_n obtained through the above equation and plugging it in the last equation of (29), one obtains

$$B = \frac{a_n \prod_{j=1}^{n-1} \frac{a_j}{k_j(1-r_j)}}{\alpha_n + \sum_{l=1}^{n-1} \frac{\alpha_l}{k_l(1-r_l)} \prod_{k=l+1}^{n-1} \frac{a_k}{k_k(1-r_k)}}.$$

Thus, the solutions of the system of equations (29) are given by

$$A_1 = 1, \quad B = \frac{a_n \prod_{j=1}^{n-1} \frac{a_j}{k_j(1-r_j)}}{\alpha_n + \sum_{l=1}^{n-1} \frac{\alpha_l}{k_l(1-r_l)} \prod_{k=l+1}^{n-1} \frac{a_k}{k_k(1-r_k)}},$$

$$A_i = \prod_{j=1}^{i-1} \frac{a_j}{k_j(1-r_j)}$$

$$- \left(\sum_{l=1}^{i-1} \frac{\alpha_l}{k_l(1-r_l)} \prod_{k=l+1}^{i-1} \frac{a_k}{k_k(1-r_k)} \right) B, \quad i = 2, \dots, n,$$

$$D = \frac{\gamma}{a_R} B \quad \text{and}$$

$$C = \frac{1}{a_L} \left[\beta_2 S^* + \left(\delta + \frac{\varepsilon \gamma (1-\delta)}{a_R} \right) B \right]. \quad (30)$$

Replacing the expressions of A_i , B , C and D given in Eq. (30) into Eq. (24), one obtains

$$\begin{aligned} \dot{U} &= -\mu \frac{(S-S^*)^2}{S} + \beta_1 S^* I^* \left(2 - u_1 - \frac{v_1}{u_1 u_2} \right) \\ &+ \beta_2 S^* L^* \left(2 - u_1 - \frac{v_1}{u_1 u_3} \right) \\ &+ \sum_{i=1}^{n-1} A_{i+1} k_i (1-r_i) E_i^* \left(1 - \frac{v_{i+1}}{v_i} \right) \\ &+ B \sum_{i=1}^n \alpha_i E_i^* \left(1 - \frac{u_2}{v_i} \right) \\ &+ B \gamma R^* \left(1 - \frac{u_2}{u_4} \right) + B \delta L^* \left(1 - \frac{u_1}{u_3} \right) \\ &+ C \phi (1-r) I^* \left(1 - \frac{u_3}{u_2} \right) \\ &+ D r I^* \left(1 - \frac{u_4}{u_2} \right) \\ &+ D \varepsilon (1-\delta) L^* \left(1 - \frac{u_4}{u_3} \right). \end{aligned} \quad (31)$$

Now, multiplying the second, third, fourth and fifth equations of (29) by L^* , R^* , E_i^* , $i = 1, \dots, n-1$ and E_n^* , respectively, one has

$$\begin{cases} \beta_2 S^* L^* - C a_L L^* + D \varepsilon (1-\delta) L^* + B \delta L^* = 0, \\ -D a_R R^* + B \gamma R^* = 0, \\ -a_i A_i E_i^* + B \alpha_i E_i^* + A_{i+1} k_i (1-r_i) E_i^* = 0, \\ i = 1, \dots, n-1, \\ -A_n a_n E_n^* + B \alpha_n E_n^* = 0 \end{cases} \quad (32)$$

Using the expressions of $a_L L^*$, $a_R R^*$, $A_i E_i^*$, $A_n E_n^*$ defined as in Eq. (22), the above equation becomes

$$\begin{cases} \beta_2 S^* L^* - C \phi (1-r) I^* + D \varepsilon (1-\delta) L^* + B \delta L^* = 0, \\ -D r I^* - D \varepsilon (1-\delta) L^* + B \gamma R^* = 0, \\ -\beta_1 S^* I^* - \beta_2 S^* L^* + B \alpha_1 E_1^* + A_2 k_1 (1-r_1) E_1^* = 0, \\ -A_i k_{i-1} (1-r_{i-1}) E_{i-1}^* + B \alpha_i E_i^* + A_{i+1} k_i (1-r_i) E_i^* = 0, \\ i = 2, \dots, n-1, \\ -A_n k_{n-1} (1-r_{n-1}) E_{n-1}^* + B \alpha_n E_n^* = 0. \end{cases} \quad (33)$$

Let $F_1(u)$, $F_2(u)$, $F_3(u)$ and $G_i(u)$, $i = 1, \dots, n$ where $u = (u_1, u_2, u_3, v_i)^T$ be $n+3$ functions to be determined later. Then, multiplying the first, second, third, fourth and fifth equations of (33) by $F_1(u)$, $F_2(u)$, $F_3(u)$, $G_1(u)$, $G_i(u)$ ($i = 2, \dots, n-1$) and $G_n(u)$, respectively, one has

$$\begin{cases} \beta_2 S^* L^* F_1(u) - C \phi (1-r) I^* F_1(u) + D \varepsilon (1-\delta) L^* F_1(u) \\ + B \delta L^* F_1(u) = 0, \\ -D r I^* F_2(u) - D \varepsilon (1-\delta) L^* F_2(u) + B \gamma R^* F_2(u) = 0, \\ -\beta_1 S^* I^* G_1(u) - \beta_2 S^* L^* G_1(u) + B \alpha_1 E_1^* G_1(u) \\ + A_2 k_1 (1-r_1) E_1^* G_1(u) = 0, \\ -A_i k_{i-1} (1-r_{i-1}) E_{i-1}^* G_i(u) + B \alpha_i E_i^* G_i(u) \\ + A_{i+1} k_i (1-r_i) E_i^* G_i(u) = 0, \quad i = 2, \dots, n-1, \\ -A_n k_{n-1} (1-r_{n-1}) E_{n-1}^* G_n(u) + B \alpha_n E_n^* G_n(u) = 0. \end{cases} \quad (34)$$

Now, adding the three last equations of (34), one has

$$\begin{aligned}
& -\beta_1 S^* I^* G_1(u) - \beta_2 S^* L^* G_1(u) + B\alpha_1 E_1^* G_1(u) \\
& + A_2 k_1 (1 - r_1) E_1^* G_1(u) - \sum_{i=2}^{n-1} A_i k_{i-1} (1 - r_{i-1}) E_{i-1}^* G_i(u) \\
& + B \sum_{i=2}^{n-1} \alpha_i E_i^* G_i(u) + \sum_{i=2}^{n-1} A_{i+1} k_i (1 - r_i) E_i^* G_i(u) \\
& - A_n k_{n-1} (1 - r_{n-1}) E_{n-1}^* G_n(u) + B\alpha_n E_n^* G_n(u) = \\
& -\beta_1 S^* I^* G_1(u) - \beta_2 S^* L^* G_1(u) \\
& + \sum_{i=1}^{n-1} A_{i+1} k_i (1 - r_i) E_i^* [G_i(u) - G_{i+1}(u)] \\
& + \sum_{i=1}^n \alpha_i E_i^* G_i(u) = 0.
\end{aligned}$$

In this case, Eq. (34) may be rewritten as follows:

$$\left\{ \begin{array}{l}
\beta_2 S^* L^* F_1(u) - C\phi(1-r)I^* F_1(u) \\
+ D\varepsilon(1-\delta)L^* F_1(u) + B\delta L^* F_1(u) = 0, \\
-DrI^* F_2(u) - D\varepsilon(1-\delta)L^* F_2(u) + B\gamma R^* F_2(u) = 0, \\
-\beta_1 S^* I^* G_1(u) - \beta_2 S^* L^* G_1(u) \\
+ \sum_{i=1}^{n-1} A_{i+1} k_i (1 - r_i) E_i^* [G_i(u) - G_{i+1}(u)] \\
+ \sum_{i=1}^n \alpha_i E_i^* G_i(u) = 0.
\end{array} \right. \quad (35)$$

Adding Eq. (35) to the right hand side of Eq. (31) yields

$$\begin{aligned}
\dot{U} &= -\mu \frac{(S - S^*)^2}{S} + \beta_1 S^* I^* \left(2 - u_1 - \frac{v_1}{u_1 u_2} - G_1 \right) \\
&+ \beta_2 S^* L^* \left(2 - u_1 - \frac{v_1}{u_1 u_3} - G_1 + F_1 \right) \\
&+ \sum_{i=1}^{n-1} A_{i+1} k_i (1 - r_i) E_i^* \left(1 - \frac{v_{i+1}}{v_i} + G_i - G_{i+1} \right) \\
&+ B \sum_{i=1}^n \alpha_i E_i^* \left(1 - \frac{u_2}{v_i} + G_i \right) \\
&+ B\gamma R^* \left(1 - \frac{u_2}{u_4} + F_2 \right) \\
&+ B\delta L^* \left(1 - \frac{u_2}{u_3} + F_1 \right) \\
&+ C\phi(1-r)I^* \left(1 - \frac{u_3}{u_2} - F_1 \right) \\
&+ DrI^* \left(1 - \frac{u_4}{u_2} - F_2 \right) \\
&+ D\varepsilon(1-\delta)L^* \left(1 - \frac{u_4}{u_3} + F_1 - F_2 \right). \quad (36)
\end{aligned}$$

Now, we shall choose the functions $F_1(u)$, $F_2(u)$, $F_3(u)$ and $G_i(u)$, which make \dot{U} non positive. To do so, the functions $F_1(u)$, $F_2(u)$, $F_3(u)$ and $G_i(u)$ are chosen such that the coefficients of $C\phi(1-r)I^*$, DrI^* and $\alpha_i E_i^*$ are equal to zero, that is,

$$\begin{aligned}
F_1 &= 1 - \frac{u_3}{u_2}, & F_2 &= 1 - \frac{u_4}{u_2} & \text{and} \\
G_i &= -1 + \frac{u_2}{v_i} \quad (i = 1, \dots, n). \quad (37)
\end{aligned}$$

Replacing the expressions of $F_1(u)$, $F_2(u)$, $F_3(u)$ and $G_i(u)$ given as in Eq. (37) into Eq. (36), one obtains

$$\begin{aligned}
\dot{U} &= -\mu \frac{(S - S^*)^2}{S} + \beta_1 S^* I^* \left(3 - u_1 - \frac{v_1}{u_1 u_2} - \frac{u_2}{v_1} \right) \\
&+ \beta_2 S^* L^* \left(4 - u_1 - \frac{v_1}{u_1 u_3} - \frac{u_2}{v_1} - \frac{u_3}{u_2} \right) \\
&+ \sum_{i=1}^{n-1} A_{i+1} k_i (1 - r_i) E_i^* \left(1 - \frac{v_{i+1}}{v_i} + \frac{u_2}{v_i} - \frac{u_2}{v_{i+1}} \right) \\
&+ B\gamma R^* \left(2 - \frac{u_2}{u_4} - \frac{u_4}{u_2} \right) + B\delta L^* \left(2 - \frac{u_2}{u_3} - \frac{u_3}{u_2} \right) \\
&+ D\varepsilon(1-\delta)L^* \left(1 - \frac{u_4}{u_3} - \frac{u_3}{u_2} + \frac{u_4}{u_2} \right). \quad (38)
\end{aligned}$$

Now let

$$F = 1 + \frac{u_4}{u_2} - \frac{u_4}{u_3} - \frac{u_3}{u_2} \quad \text{and} \quad (39)$$

$$H_i = 1 + \frac{u_2}{v_i} - \frac{v_{i+1}}{v_i} - \frac{u_2}{v_{i+1}}, \quad i = 1, \dots, n-1.$$

The next step is to show that the functions F and H_i are non-positive for all $u_1, u_2, u_3, v_i \in \mathbb{R}_{\geq 0}$.

By using lemma 4 in Appendix A with $w = 2$, $y_1 = u_3$, $y_2 = u_4$ and $Y = u_2$, when $u_3 \leq u_4 \leq u_2 \leq 1$ one has $1 + \frac{u_4}{u_2} - \frac{u_4}{u_3} - \frac{u_3}{u_2} \leq 0$, i.e., $F \leq 0$. Also, using the same lemma with $w = 2$, $y_1 = v_{i+1}$, $y_2 = u_2$ and $Y = v_i$, then if $v_{i+1} \leq u_2 \leq v_i \leq 1$ one has $1 + \frac{u_2}{v_i} - \frac{v_{i+1}}{v_i} - \frac{u_2}{v_{i+1}} \leq 0$, i.e., $H_i \leq 0$. Thus $\dot{U} \leq 0$ and Eq. (38) implies that \dot{U} is less than or equal to zero with equality only if $S = S^*$. Therefore, $\dot{U} \leq 0$ for all $S, E_i, I, L, R \geq 0$, provided that $S^*, E_i^*, I^*, L^*, R^*$ are positive, where the equality $\dot{U} = 0$ holds only on the straight line $S = S^*, E_i^*/E_i = I^*/I = L^*/L = R^*/R$. It is easy to see that for the system (1), P^* is the only equilibrium state on this line. Therefore, by Lyapunov-LaSalle asymptotic stability theorem [35-37], the positive equilibrium state P^* is globally asymptotically stable in the positive region $\Omega_\rho \subset \mathbb{R}_{\geq 0}^{n+4}$, except on the S -axis which is the stable manifold for the fixed point P_0 . This achieves the proof. \square

Remark 1: It is possible for inequality (19) to fail, in which case the global stability of P^* has not been established. The local stability result and numerical simulations, however, seem to support the idea that P^* is still global asymptotically stable even in this case.

IV. NUMERICAL STUDIES

To illustrate the various theoretical results contained in this paper, the model (1) is simulated with two latent classes ($n = 2$) and using the parameter value/range of Cameroon in the following table.

Table: Description and estimation of parameters

Parameters	Description	Estimated value/range
Λ	Recruitment rate of susceptible individuals into the community	2 (year)^{-1}
β_1	Transmission coefficient of infectious variable	
β_2	Transmission coefficient of loss of slight who continue to have disease	variable
μ	Naturally death rate	0.0101^2
k_1	Rate of progression from the first latent class to the second latent class	0.5^1
r_1	Rate of effective chemoprophylaxis of individuals of the first latent class	0^3
α_1	Rate of progression from the first latent class to infectious	0.003
α_2	Rate of progression from the second latent class to infectious	0.005
r	Rate of effective therapy of infectious	0.8182^3
ϕ	Rate of progression from infectious to loss of slight	0.2^3
δ	Rate at which loss of slight return to the hospital	0.1
γ	Relapse rate of recovered individuals	0.002
ε	Recovered rate of loss of slight	0.001
d_I	Death rate of infectious	0.022722^3
d_L	Death rate of loss of slight	0.020^3
d_1	Additional death rate of latently infected individuals of the first latent class	0.001
d_2	Additional death rate of latently infected individuals of the second latent class	0.002

Note. Sources of estimates:

¹World Health Organization

²Ministry of Administration and Decentralization, Cameroon

³National Comity of Fight against Tuberculosis, Cameroon [38].

Numerical results for the model (1) are depicted in Figs. 2-4.

Figure 2 presents the basic reproduction ratio \mathcal{R}_0 as a function of the parameter at which infectious become loss of slight ϕ . In this figure the line blue stands for $\beta_2 < \Delta$ and the line red for $\beta_2 > \Delta$. As predicted by Lemma 1, when the parameter ϕ increases, one can see that for $\beta_2 < \Delta$ the basic reproduction ratio decreases while for $\beta_2 > \Delta$, the basic reproduction ratio increases.

Figure 3 presents the trajectory plot when $\mathcal{R}_0 \leq 1$. From this figure, one can see that the trajectory of model (1) converges to the disease free equilibrium. This means that the disease disappears in the host population. Figure 4 gives the trajectory plot when $\mathcal{R}_0 > 1$. One can observe that the trajectory converge to the unique endemic equilibrium point. Thus, when $\mathcal{R}_0 > 1$, the disease persists in the host population as shown in Theorem 2.

V. CONCLUSION

In this paper, we have give a complete analysis of a tuberculosis model with two differential infectivity, n classes of latently infected individuals and mass balance incidence. By analyzing this model, we found that it is globally asymptotically stable and possesses the only globally stable equilibrium state. Depending on the basic reproduction ratio, this steady state is either the endemic or the disease-free. The global stability of the infection-free equilibrium state implies that for any initial level of infection the disease will eventually fade out from the population when the condition for this stability, namely $\mathcal{R}_0 \leq 1$, holds. The condition $\mathcal{R}_0 > 1$ implies that the disease will persist in a population. Numerical results are presented to illustrate the analytical results.

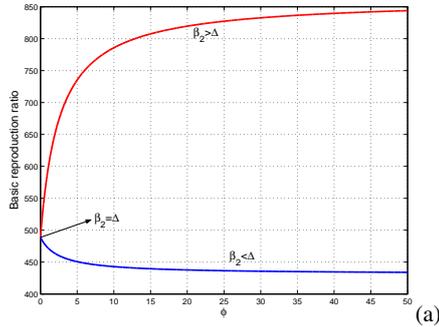


Fig. 2. Basic reproduction ratio as a function of ϕ .

APPENDIX A USEFUL INEQUALITIES

In this appendix, we give inequalities which are necessary to demonstrate that the time derivative of the Lyapunov function $U(S, E_i, I, L, R)$ is non-positive. A key tool is the Arithmetic-Geometric Means Inequality, which we state here.

Lemma 3: (Arithmetic-Geometric Means Inequality): Let z_1, \dots, z_w be positive real numbers. Then,

$$\sqrt[w]{z_1 \dots z_w} \leq \frac{z_1 + \dots + z_w}{w}. \quad (40)$$

Furthermore, exact equality only occur if $z_1 = \dots = z_w$.

An immediate consequence of the Arithmetic-Geometric Means Inequality follows.

Lemma 4: [16]: Let $y_1 \leq \dots \leq y_w \leq Y$ be positive real numbers. Then

$$\frac{y_w}{Y} + (w-1) - \left(\frac{y_1}{Y} + \frac{y_2}{y_1} + \dots + \frac{y_w}{y_{w-1}} \right) \leq 0. \quad (41)$$

ACKNOWLEDGMENT

Part of this work has been done during the visit of Samuel Bowong at the Postdam Institute for Climate Impact Research, Germany. He gratefully acknowledges, with thanks, the support in part of the Humboldt Foundation and the Postdam Institute for Climate Impact Research, Germany. The authors would like to thank Professor Gauthier Sallet (IRD, Project MASAIE, University of Metz, France) for very useful comments pertaining to the model design.

REFERENCES

- [1] B. M. Murphy, B. H. Singer, D. Kirschner, Comparing epidemic tuberculosis in demographically distinct populations, *Maths. Biosci.* 180 (2002) 161-185.
- [2] S. M. Blower, A. R. McLean, T. C. Porco, P. M. Small, P. C. Hopwell, M. A. Sanchez, A. R. Ross, The intrinsic transmission dynamics of tuberculosis epidemics, *Nature Medicine* 1 (1995) 815-821.
- [3] C. Castillo-Chavez, B. Song, Dynamical models of tuberculosis and their applications, *Math. Biosci. and Eng.* 1 (2004) 361-404.
- [4] J. Snider, M. Rabiglion, A. Kochi, Global burden of tuberculosis. In B. R. Bloom Ed.: *Tuberculosis, Pathogenesis, Protection and Control*, ASM Press, Washington, DC, pp. 47-59, 1994.
- [5] B. R. Bloom, *Tuberculosis: Pathogenesis, Protection and Control*, ASME Press, Washington, DC, 1994.
- [6] B. R. Bloom, C. Murray, Tuberculosis: commentary on a reemergent killer, *Science* 257 (1992) 1055-1064.
- [7] C. Dye, S. Scheele, P. Dolin, Global burden of tuberculosis, *JAMA* 181 (1999) 677-685.
- [8] E. Ziv, C. L. Daley, S. M. Blower, Early therapy for latent tuberculosis infection, *Amer. J. Epid.* 153 (2001) 381-385.
- [9] S. M. Blower, J. L. Gerberding, Understanding, predicting and controlling the emergence of drug-resistant tuberculosis: a theoretical framework, *J. Mod. Med.* 76 (1998) 624-636.

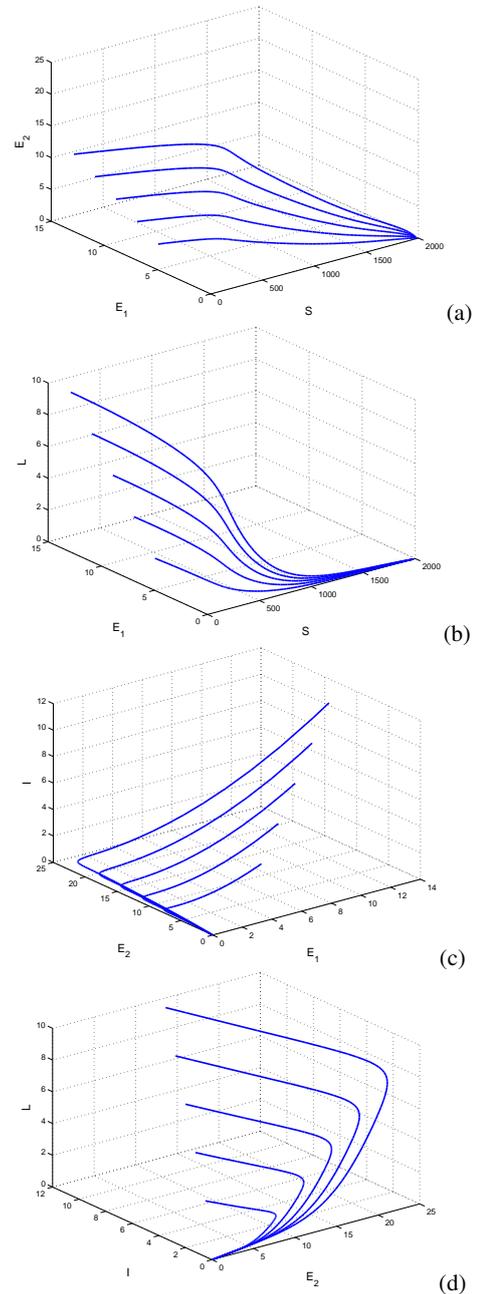


Fig. 3. The trajectory of model (1) for different initial conditions when $\mathcal{R}_0 \leq 1$.

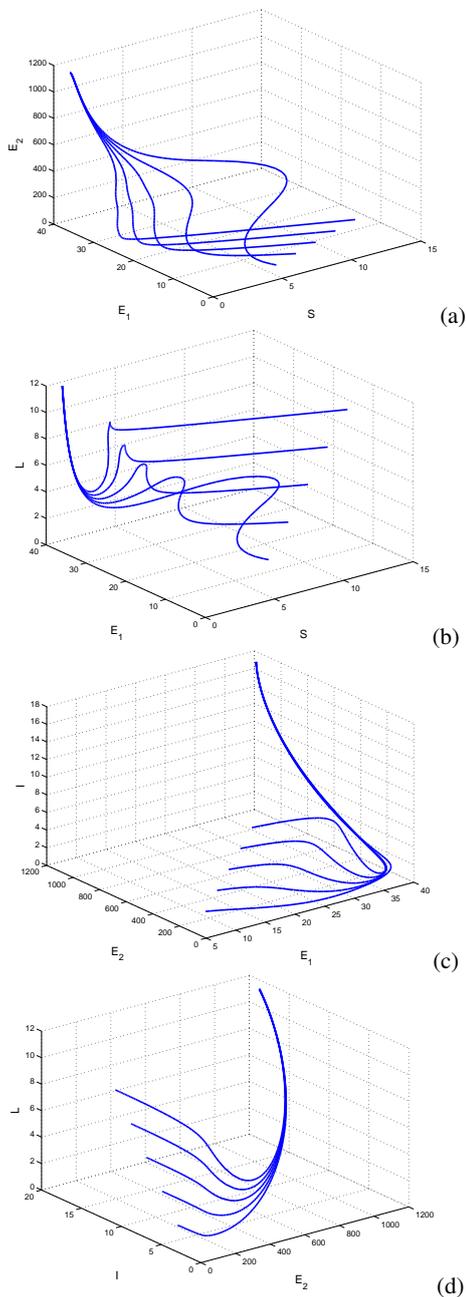


Fig. 4. The trajectory of model (1) for different initial conditions when $\mathcal{R}_0 > 1$.

[10] S. M. Blower, P. Small, P. Hopewell, Control strategies for tuberculosis epidemics: new method for old problem, *Science* 273 (1996) 497-500.

[11] WHO/IUATLD, Global Working Group on Antituberculosis Drug Resistance Surveillance, Guidelines for surveillance of drug resistance in tuberculosis, WHO Geneva/IUATLD, Paris. *Int. J. Tuberc. Lung Dis.* 2 (1998) 72-89.

[12] WHO/IUATLD, Global Project on Anti-tuberculosis Drug Resistance Surveillance. Anti-Tuberculosis Drug Resistance in the World: third global report, WHO/HTM/TB/2004.343, 2004.

[13] World Health Organization, Global tuberculosis control: surveillance, planning, financing, WHO/HTM/TB/2005.349, Geneva.

[14] A. S. Benenson (Ed.), *Control of Communicable Diseases Manual*, 16th ed., American Public Health Association, 1995

[15] N. E. Dunlap, D. E. Briles, Immunology of tuberculosis, *Med. Clin. North Amer.* 77 (1993) 1235-1251.

[16] B. Miller, Preventive therapy for tuberculosis, *Med. Clin. North Amer.* 77 (6) (1993) 1263-1275.

[17] C. P. Bhunu, W. Garira, Z. Mukandawire, M. Zimba, Tuberculosis transmission model with chemoprophylaxis and treatment, *Bulletin of Mathematical Biology*, doi:10.1007/S11538-008-9295-4.

[18] N. Bacaer, R. Ouifki, C. Pretorius, R. Wood, B. William, Modelling the joint epidemics of TB and HIV in a South African township, *J. Math. Biol.* 57 (2008) 557-593.

[19] A. Korobeinikov, Lyapunov functions and global properties for SIER and SEIS epidemic models, *Math. Med. Biol.* 21 (2001) 75-83.

[20] A. Korobeinikov, P. K. Maini, A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence, *Math. Biosci. Eng.* 1 (2004) 57-60.

[21] C. C. McCluskey, Global stability for a class of mass action systems allowing for latency in tuberculosis, *J. Math. Ana. App.* 338 (2008) 518-555.

[22] C. C. McCluskey, Lyapunov functions for tuberculosis models with fast and slow progression, *Math. Bios. Eng.* 3 (2006) 603-614.

[23] N. Bame, S. Bowong, J. Mbang, G. Sallet, J. J. Tewa, Global stability for SEIS models with n latent classes, *Math. Biosci. Eng.* 5 (2008) 20-33.

[24] A. Igdir, J. C. Kamgang, G. Sallet, J. J. Tewa, Global analysis of new malaria intrahost models with a competitive exclusion principle, *SIAM J. App. Math.* 1 (2007) 260-278.

[25] P. Adda, J. L. Dimi, A. Igdir, J. C. Kamgang, G. Sallet, J. J. Tewa, General models of host-parasite systems, *Global analysis, Discrete Contin. Dyn. Syst. Ser. B* 8 (2007) 1-17.

[26] A. Igdir, J. Mbang, G. Sallet, J. J. Tewa, Multi-compartment models, *DCDS* (2007) 506-519.

[27] A. Igdir, J. Mbang, G. Sallet, Stability analysis of within-host parasite models with delays, *Math. Biosci.* 209 (2007) 51-75.

[28] J. J. Tewa, J. L. Dimi, S. Bowong, Lyapunov functions for a dengue disease transmission model, *Chaos, Solitons & Fractals* 39 (2009) 936-941.

[29] A. Berman, R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, SIAM, 1994.

[30] J. A. Jacquez, C. P. Simon, *Qualitative theory of compartmental systems*, *SIAM Rev.* 35 (1993) 43-79.

[31] D. G. Luenberger, *Introduction to dynamic systems: theory, models, and applications*, John Wiley & Sons Ltd., 1979.

[32] P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Bios.* 180 (2002) 29-28.

[33] O. Diekmann, J. A. P. Heesterbeek, J. A. J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, *J. Math. Biol.* 28 (1990) 365-382.

[34] O. Diekmann, J. A. P. Heesterbeek, J. A. J. Metz, *Mathematical epidemiology of infectious diseases*, Wiley Series in Mathematical and Computational Biology, John Wiley & Sons Ltd., Chichester, 2000, Model building, analysis and interpretation.

[35] J. P. LaSalle, *The stability of dynamical systems*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1976. With an appendix: "Limiting equations and stability of nonautonomous ordinary differential equations" by Z. Artstein, *Regional Conference Series in Applied Mathematics*.

[36] J.P. LaSalle, Stability theory for ordinary differential equations, *J. Differ. Equ.* 41 (1968) 57-65.

[37] N. P. Bhatia, G. P. Szegö, *Stability Theory of Dynamical Systems*, Springer-Verlag, 1970.

[38] National Comity of Fight Against Tuberculosis, "Guide de personnel de la santé," 2001.

Complex dynamics of cellular automata emerging in chaotic rules

Genaro J. Martínez¹, Andrew Adamatzky¹ and Ramon Alonso-Sanz²

Abstract—We show novel techniques of analysing complex dynamics of cellular automata (CA) with chaotic behaviour. CA are well known computational substrates for studying emergent collective behaviour, complexity, randomness and interaction between order and disorder. A number of attempts have been made to classify CA functions on their spatio-temporal dynamics and to predict behaviour of any given function. Examples include mechanical computation, λ and Z -parameters, mean field theory, differential equations and number conserving features. We propose to classify CA based on their behaviour when they act in a historical mode, i.e. as CA with memory. We demonstrate that cell-state transition rules enriched with memory quickly transform a chaotic system converging to a complex global behaviour from almost any initial condition. Thus just in few steps we can select chaotic rules without exhaustive computational experiments or recurring to additional parameters. We provide analysis of well-known chaotic functions in one-dimensional CA, and decompose dynamics of the automata using majority memory.

Index Terms—Cellular automata, memory, complex dynamics, chaos, self-organization and filter

I. INTRODUCTION

IN this paper we will introduce a simple tool to extract complex systems from a family of chaotic discrete dynamical system. We will employ a technique — memory based rule analysis [4] of using past history of a system to construct its present state and to predict its future.

We focus on one-dimensional cellular automata (CA). CA are well known computational substrates for studying emergent collective behaviour, complexity, randomness and interaction between order and disorder. A number of efforts have been made to classify CA functions on their spatio-temporal dynamics and to predict behaviour of any given function. Examples include mechanical computation, λ and Z -parameters, mean field theory, differential equations and number conserving features. We propose to classify CA based on their behaviour when they act in a historical mode, i.e. as CA with memory.

Particularly we study elementary CA (ECA) where each function evaluates a central cell with their two neighborhoods (left and right) and every cell takes a value of its binary alphabet. ECA were introduced by Wolfram and extensively studied [28]. In ECA there is a set of functions determining global chaotic behavior where global configurations are disordered, many configurations have many ancestors, and attractors are dense [27].

¹Bristol Institute Technology, University of the West of England, Bristol, United Kingdom. Email: {genaro.martinez, andrew.adamatzky}@uwe.ac.uk

²ETSI Agrónomos, Polytechnic University of Madrid, Madrid, Spain. Email: ramon.alonso@upm.es

Manuscript received April 21, 2009

ECA is a one-dimensional array of finite automata, each automaton takes two states and updates its state in discrete time depending on its own state and states of its two closest neighbours, all cells update their state synchronously. A general classification of ECA was introduced in [29] as follow:

- class I.* CA evolving to a homogeneous state.
- class II.* CA evolving periodically.
- class III.* CA evolving chaotically.
- class IV.* Include all previous cases, as well known as class of *complex rules*.

In this classification class IV is of particular interest because the rules of the class exhibit non-trivial behaviour with rich diversity of patterns emerging and non-trivial interactions between travelling localisations, or gliders, e.g. ECA Rule 54 [16].

In present paper we aim to transform a chaotic evolution rule to a complex system by using memory

$$\text{chaotic ECA} \xrightarrow{\text{memory}} \text{complex ECA}$$

and derive a new class of CA functions with historic evolution.

We believe that by employing historic evolution we are able to explore hidden properties of chaotic systems, and select chaotic rules with ordered dynamics.

II. BASIC NOTATION

A. One-dimensional cellular automata

One-dimensional CA is represented by an array of *cells* x_i where $i \in \mathbb{Z}$ (integer set) and each x takes a value from a finite alphabet Σ . Thus, a sequence of cells $\{x_i\}$ of finite length n represents a string or *global configuration* c on Σ . This way, the set of finite configurations will be represented as Σ^n . An evolution is represented by a sequence of configurations $\{c_i\}$ given by the mapping $\Phi : \Sigma^n \rightarrow \Sigma^n$; thus their global relation is following

$$\Phi(c^t) \rightarrow c^{t+1} \quad (1)$$

where t is time steps and every global state of c is defined by a sequence of cell states. Also the cell states in configuration c^t are updated at the next configuration c^{t+1} simultaneously by a local function φ as follow

$$\varphi(x_{i-r}^t, \dots, x_i^t, \dots, x_{i+r}^t) \rightarrow x_i^{t+1}. \quad (2)$$

Wolfram represents one-dimensional CA with two parameters (k, r) . Where $k = |\Sigma|$ is the number of states, and r is radius of neighbourhood. ECA are defined by parameters $(2, 1)$.

There are Σ^n different neighborhoods (where $n = 2r + 1$) and k^{k^n} different evolution rules.

In computer experiments we are using automata with periodic boundary conditions.

B. Cellular automata with memory

Conventional cellular automata are ahistoric (memoryless): i.e., the new state of a cell depends on the neighborhood configuration solely at the preceding time step of φ (e.g. 2).

CA with *memory* can be considered as an extension of the standard framework of CA where every cell x_i is allowed to remember some period of its previous evolution.

Thus to implement a memory we design a memory function ϕ , as follow:

$$\phi(x_i^{t-\tau}, \dots, x_i^{t-1}, x_i^t) \rightarrow s_i \quad (3)$$

such that $\tau < t$ determines the degree of memory backwards and each cell $s_i \in \Sigma$ being a state function of the series of states of the cell x_i with memory up to time-step. Finally to execute the evolution we apply the original rule as follows:

$$\varphi(\dots, s_{i-1}^t, s_i^t, s_{i+1}^t, \dots) \rightarrow x_i^{t+1}.$$

Thus in CA with memory, while the mapping φ remains unaltered, historic memory of all past iterations is retained by featuring each cell as a summary of its past states from ϕ . Therefore cells *canalize* memory to the map φ .

As an example, we can consider memory function ϕ as a *majority memory*:

$$\phi_{maj} \rightarrow s_i \quad (4)$$

where in case of a tie given by $\Sigma_1 = \Sigma_0$ from ϕ then we will take the last value x_i . So ϕ_{maj} function represents the classic majority function [21] on the cells $(x_i^{t-\tau}, \dots, x_i^{t-1}, x_i^t)$ and define a temporal ring before to get finally the next global configuration c .

Note that memory is a simple function as CA self and as well its global behavior Φ produced is totally unpredictable from its local function ϕ and φ .

III. CLASSES OF ECA BY POLYNOMIALS

A. Mean filed approximation

Mean field theory is a proven technique for discovering statistical properties of CA without analyzing evolution spaces of individual rules [19].

The method assumes that elements of the set of states Σ are independent, uncorrelated between each other in the rule's evolution space φ . Therefore we can study probabilities of states in neighborhood in terms of probability of a single state (the state in which the neighborhood evolves), thus probability of a neighborhood is the product of the probabilities of each cell in the neighborhood.

In this way, it was proposed to explain Wolfram's classes by a mixture of probability theory and de Bruijn diagrams in [20], resulting in a classification based on mean field theory curve:

- class I: monotonic, entirely on one side of diagonal;
- class II: horizontal tangency, never reaches diagonal;
- class III: no tangencies, curve crosses diagonal.
- class IV: horizontal plus diagonal tangency, no crossing;

Thus for one dimension all their neighborhoods must be considered as follow:

$$p_{t+1} = \sum_{j=0}^{k^{2r+1}-1} \varphi_j(X) p_t^v (1-p_t)^{n-v} \quad (5)$$

such that j is a number of relations from their neighborhoods and X the combination of cells $x_{i-r}, \dots, x_i, \dots, x_{i+r}$. n represents the number of cells in neighborhood, v indicates how often state one occurs in Moore's neighborhood, $n-v$ shows how often state zero occurs in the neighborhood, p_t is a probability of cell being in state one, q_t is a probability of cell being in state zero (therefore $q = 1-p$).

IV. COMPLEX DYNAMICS EMERGING FROM CHAOTIC ECA

A. Chaotic ECA

Let us consider two cases of classic ECA with chaotic behavior to demonstrate our results: the evolution rules 86 and 101.

We need to provide their mean field approximation to verify that both function have a chaotic global behavior before selecting the memory.

The local rule φ corresponding to rule 86 is following:

$$\varphi_{R86} = \begin{cases} 1 & \text{if } 110, 100, 010, 001 \\ 0 & \text{if } 111, 101, 011, 000 \end{cases}.$$

Initially φ_{R86} has produces states zero and one equiprobably. There is an equilibrium of states in Φ . On the other hand, φ_{R86} determines a surjective correspondence and therefore all the configuration has at least one ancestor and no Garden of Eden configurations [1]. Also this rule is the reflexion of well-known ECA rule 30 [27].

The local function for rule 101 is following:

$$\varphi_{R101} = \begin{cases} 1 & \text{if } 110, 101, 010, 000 \\ 0 & \text{if } 111, 100, 011, 001 \end{cases}.$$

In this case, φ_{R101} has the same probability as φ_{R86} to produce states one and zero. However φ_{R101} is not a surjective rule and therefore has the Garden of Eden configurations, i.e., not all configurations have ancestors.

To classify global behavior properly of φ_{R86} and φ_{R101} we should calculate their mean field polynomials. Mean field polynomial to φ_{R86} is:

$$p_{t+1} = 3p_t q_t^2 + p_t^2 q_t \quad (6)$$

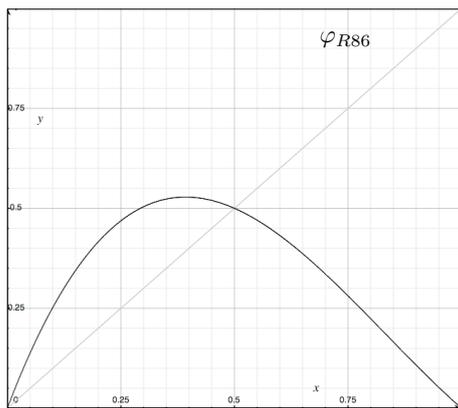
and for φ_{R101} we have:

$$p_{t+1} = 2p_t^2 q_t + p_t q_t^2 + q_t^3. \quad (7)$$

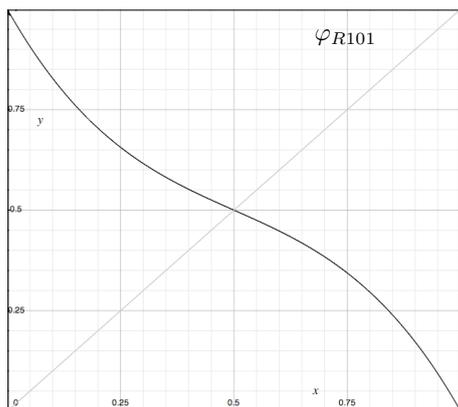
Quickly the polynomial for φ_{R86} satisfies the mean field classification (section III.A). Where rules in CA class III do not have tangencies and therefore the curve crosses the

identity. Consequently, φ_{R86} evolves with a chaotic global behavior (see fig. 1(a)).

This mean field polynomial has an stable fixed point when e.q. 6 is $f = 0.5$. This value relate the existence of densities where the population of cells in state one is preserved with few changes. Also such fixed point confirm its initial probability since φ_{R86} . Of course, if there are extreme densities of zeros and ones then then next time Φ will be filled of states zeros only.



(a)



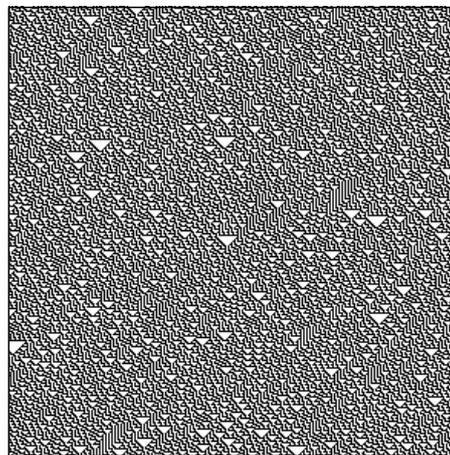
(b)

Fig. 1. Mean field curves for (a) φ_{R86} and (b) φ_{R101} respectively.

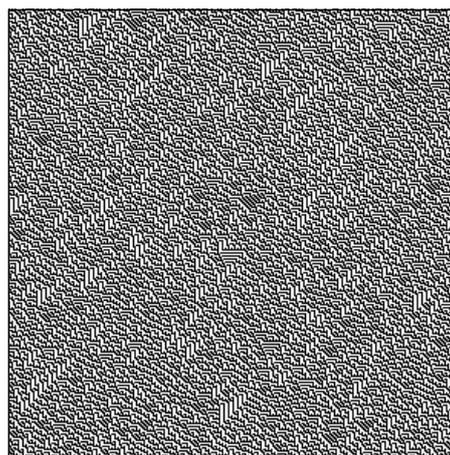
Mean field curve for φ_{R101} (see fig. 1(b)) presents another characteristic. Again the curve does not cross the identity and its global behavior Φ should then be chaotic. Its stable fixed point $f = 0.5$ relates to the initial probability estimated since φ_{R101} . The curve displays what would happen if some initial configuration c_0 is dominated by state one, at the next step Φ will be dominated by states zero and therefore this behavior should repeat periodically. Such phenomenon also is balanced with its 50% of density to each step.

Finally fig. 2 displays two evolutions with typical chaotic behavior in ECA. First evolution (a) displays the chaotic global evolution of φ_{R86} since a random initial condition with a 50% of density. That confirm an evolution without some order or pattern defined. Second evolution (b) displays the

chaotic global behavior for φ_{R101} with the same parameters. Inclusive the same initial condition was used to calculate both evolutions.



(a)



(b)

Fig. 2. Chaotic global behavior in ECA evolution rules (a) φ_{R86} and (b) φ_{R101} evolving over an array of 294 cells in 296 generations. Both evolutions start since a random initial density of 50%. Black cells represent the state one and white cells the state zero.

Now we will select a kind of memory and discover hidden information in those chaotic ECA's.

B. Filtering evolutions

Filters selected in CA are an useful tool for understand properties that at the first view they are hide. This tool was amply utilized by Wuensche in [31] in the context of automatic classification of CA. Filters were deduced form mechanical computation paradigm [12] and frequencies by states [31].

Others derivations deducing filters relate to tiling, as used for ECA rule 110 and rule 54 [23], [17]. In general, such filters are not widely exploited in studying CA. We consider the tile representation to identify filters as block of cells on one or two

dimensions. We explain each tile filtering φ_{R86} and φ_{R101} in the next section.

C. Complex dynamics emerging in φ_{R86} and φ_{R101} with majority memory

Firstly we should consider a kind of memory, in this case the majority memory ϕ_{maj} (see eq. 4) and then a value for τ . This value represent the number of cells backward to consider in the memory. Therefore a way to represent functions with memory and one ECA associated is proposed as follow:

$$\phi_{mRca:\tau} \quad (8)$$

such that ca represents the decimal notation of an specific ECA and m a kind of memory given. This way the majority memory working in ECA rule 86 checking tree cells on its history is denoted simply as $\phi_{majR86:3}$.

Implementing the majority memory ϕ_{maj} we can select some ECA and experimentally look what is the effect. Figures 3 and 4 shows the result of selecting memory τ in φ_{R86} and φ_{R101} respectively. The result is a new family of ECA with majority memory, the rules: $\phi_{majR86:3}, \dots, \phi_{majR86:\infty}$ and $\phi_{majR101:3}, \dots, \phi_{majR101:\infty}$.

As a characteristic while the memory is working on ϕ_{majR86} and $\phi_{majR101}$ a periodic background was more evident and it can be represented as a tile. But also these filters work as well at the original rules φ as have been illustrated the first evolutions at the figs. 3 and 4.

The memory effect produces an emergency of patterns and some of them interacting quickly. In fact, ϕ_{majR86} and some values of τ change dynamics dramatically. By previous results in [15] we consider only even values that offer better global dynamics. The new rule $\phi_{majR86:8}$ displays patterns as particles traveling in different velocities and a periodic background that avoid an space filled of patterns.

The second case $\phi_{majR101}$ displays more attractive result. These three new evolution rules are able to support stationary and mobile particles, traveling and colliding, some of them as soliton reactions.

Also on all evolutions a filter was used to clarify the evolutions and patterns.¹ Filters really are useful to recognize periodic dominant patterns of objects moving into such local universes.

The first two-dimensional tile working in ϕ_{majR86} is represented as $t_{R86} = \begin{bmatrix} 101 \\ 101 \end{bmatrix}$. Also this tile works at the original evolution rule as show the fig. 3. The tile reported for $\phi_{majR101}$ is determined for the two-dimensional tile $t_{R101} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}$. So this filter works on the original evolution rule as well as show the fig. 4.

The effect of memory producing new evolution rules is preserved in some way. Initially the existence of a filter that can evolve on all different function, that is not rare because the memory only read the history and process the new generation with the original rule.

¹All evolutions simulated to ECA and ECA with memory they are calculated with OSXLCAU21 system, available from <http://uncomp.uwe.ac.uk/genaro/OSXCASystems.html>

D. Coding particles

1) *Self-organization by structure formation:* Patterns as particles and non-trivial behavior emerging in these new ECA with memory ϕ_{majR86} and $\phi_{majR101}$, naturally conduce to known problems as self-organization.

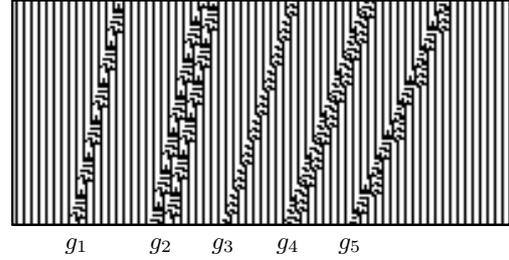


Fig. 5. Set of particles \mathcal{G} emerging and living in $\phi_{majR101:4}$.

Considering the evolution rule $\phi_{majR101:4}$, we have done a classification of particles in this local universe (see fig. 5). The universe is not bigger compared with other complex rules, moreover is interesting how all particles in $\phi_{majR101:4}$ can be produced from them self with binary collisions, i.e., a self-organization by structure formation [14], given the next relation of reactions:

$$g_i \rightarrow g_j = g_k$$

such that $i \neq j \neq k$ and $i, j, k \in \mathcal{G}_{\phi_{majR101:4}}$.

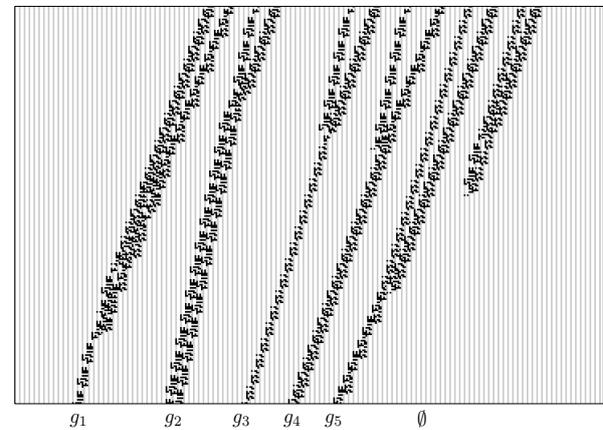


Fig. 6. Self-organization by particle collisions to form the set $\mathcal{G}_{\phi_{majR101:4}}$ evolution is filtered.

By coding such particles in $\phi_{majR101:4}$ then we can select each particle from other different particles. Figure 6 presents the set of reactions to get every particle including an annihilation reaction. Thus a set of collisions to reproduce this list is following:

- 1) $g_4 - b^3 - g_5 = g_1$
- 2) $g_1 - b^2 - g_4 = g_2$
- 3) $g_1 - b^4 - g_4 = g_3$
- 4) $g_1 - b^6 - g_5 = g_4$

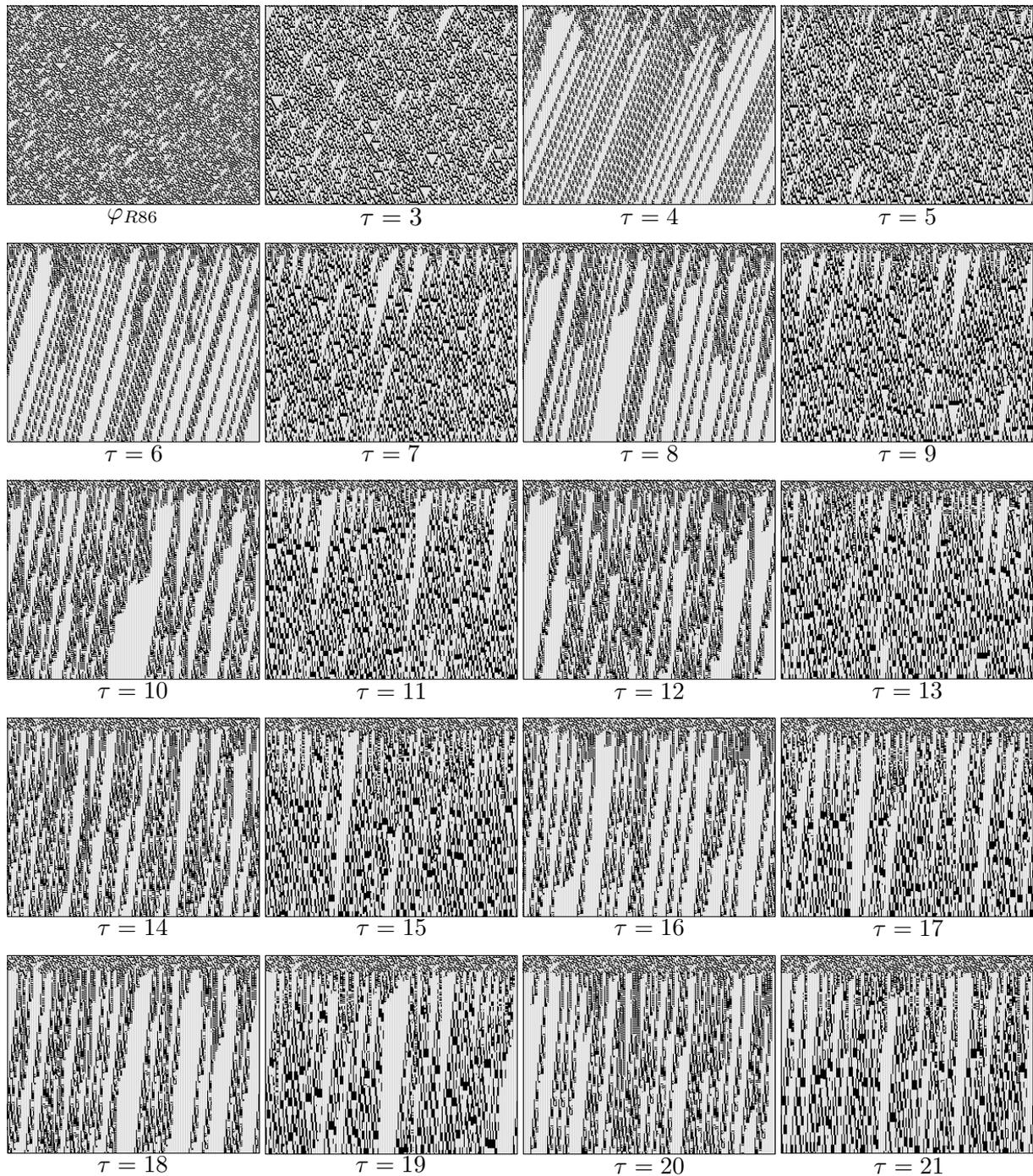


Fig. 3. Majority memory ϕ_{maj} working in φ_{R86} with τ values of 3 to 21, respectively. The first one is the original ECA Rule 86 evolution. All snapshots evolve with the same random initial condition to 50% over an array of 287 cells to 228 generations, and also all evolutions are filtered.

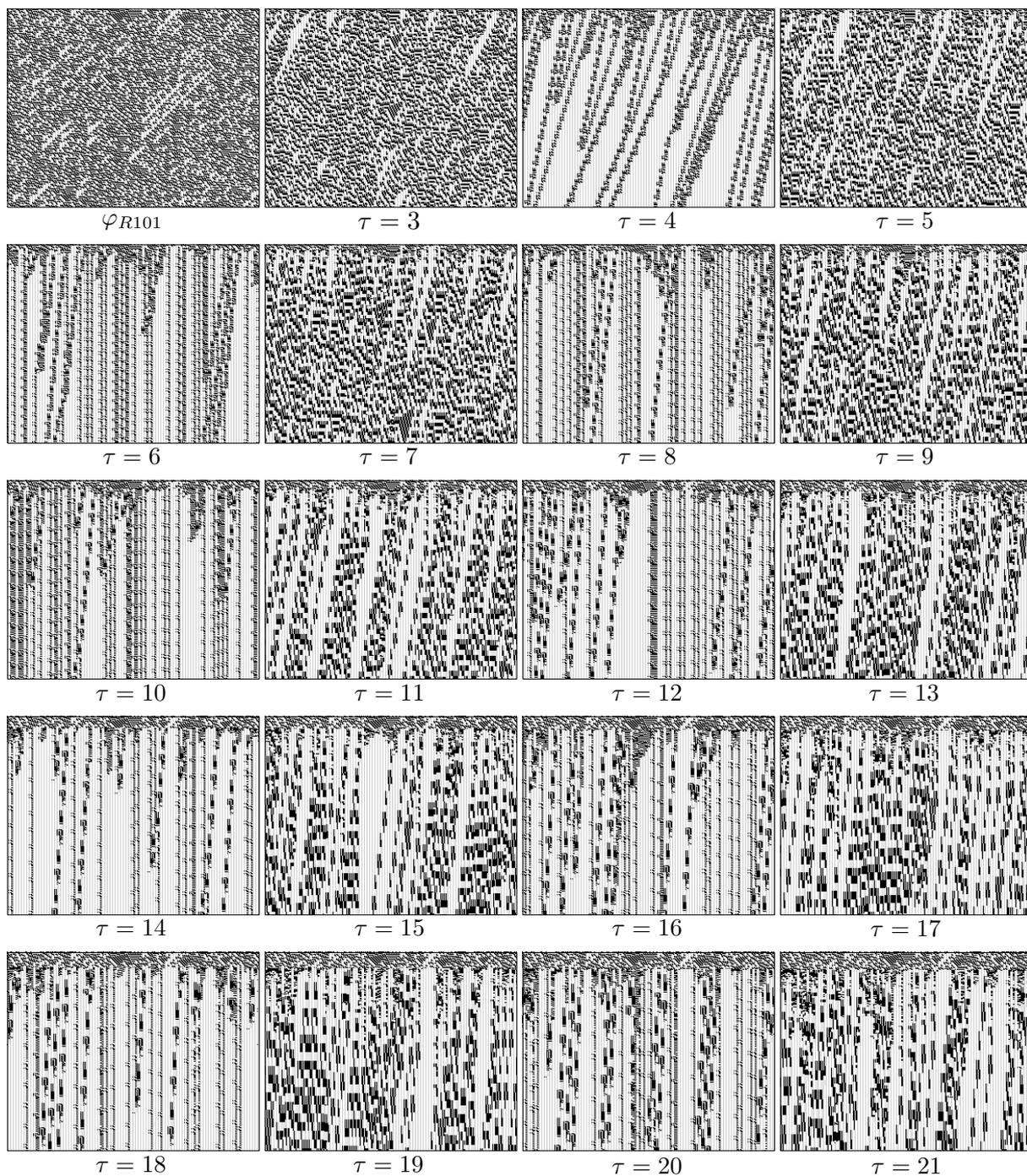


Fig. 4. Majority memory ϕ_{maj} working in φ_{R101} with τ values of 3 to 21, respectively. The first one is the original ECA Rule 101 evolution. All snapshots evolve with the same random initial condition (the same initial condition for all cases in this and the previous figure) to 50% over an array of 287 cells to 228 generations, and also all evolutions are filtered.

- 5) $g_3 - b^3 - g_4 = g_5$
- 6) $g_3 - b^2 - g_4 = \emptyset$

Of course, they are not all possibilities to get every particle and a organization of several particles could be produce even more complex behavior with capacities for simulate physics, biology, chemical or computational phenomena; as wave propagation, reaction-diffusion, morphogenesis, particle collision, fluid-dynamics, (tissue) grown, pattern formation, self-reproduction, self-assembly, artificial life, synthetic constructions (engineering), tessellation, differential equations, formal languages, or unconventional computing [3], [7], [18], [22], [25].

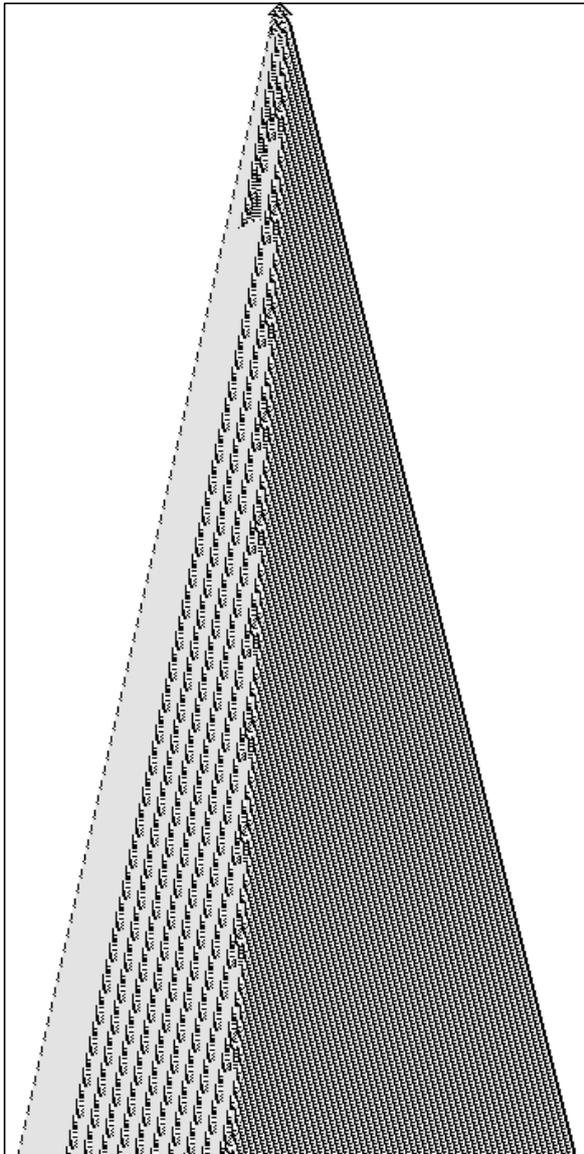


Fig. 7. Stream of particles and fuse patterns emerging from a single cell in $\phi_{majR86:8}$. These patterns exhibit unlimited growth.

2) *Generator pattern:* Figure 7 displays the evolution of a single cell in state 1 with the evolution rule $\phi_{majR86:8}$. On this evolution a fuse pattern is organized by stream of gliders (left) emitted periodically every 62 steps and a fixed periodic pattern (right) growing with a velocity of $-1/4$.

Finally both previous examples are just two simple cases showing the memory effect on chaotic ECA. Another case was developed for the ECA rule 30 in [15].

3) *Implementing basic functions:* Also we can use the particles codification to represent solutions of some basic functions. Of course, thinking how a complex systems could be organized and controlled to get a construction, as computation [3].

Let consider the new ECA rule $\phi_{majR86:8}$. Because we want to implement a simple function as **addToHead** working on two strings $w_1 = A_1, \dots, A_n$ and $w_2 = B_1, \dots, B_m$, such that, $n, m \geq 1$. For example, if $w_1 = AAA$, $w_2 = BBB$ and $w_3 = w_1w_2$ then the **addToHead**(w_2) will yield: $w_3 = w_2w_1$. As the next diagram shows.

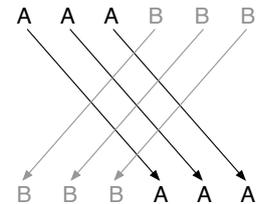


Fig. 8. Schematic diagram adding the string w_2 to head of the list w_3 .

To implement such function on $\phi_{majR86:8}$ we must consider represent every data as a particle. Thus g_1 and g_2 gliders to $\phi_{majR86:8}$ are coded to reproduce a soliton reaction.² The second problem is to synchronize several gliders and obtain the same result with multiple collisions.

The codification was not complicated to get. However a systematic analysis by reactions is necessary. We known than a periodic gap and one fixed phase between particles is sufficient to reproduce the **addToHead** function for any string $A^n B^m$.

Figure 9 displays fragments of evolutions of $\phi_{majR86:8}$ from an initial condition coded by gliders representing the string AAAAAAAAAAABBBBBBBBBB. So operating the **addToHead** function to get the final string BBBBBBBBBB-BAAAAAAAAAAAA late of 6,888 generations. The first snapshot displays its initial configuration and the first 400 steps, the middle snapshot mainly presents how the string w_1 across the string w_2 preserving the information (soliton reaction), and the third snapshot shows the final global configuration so given the string w_2w_1 processed in parallel with $\phi_{majR86:8}$.

V. DISCUSSION

We have demonstrated that memory in ECA offers a new approach to discover complex dynamics based on particles and

²These gliders are a reflexion of ECA rule 30 with memory, because rule 86 is the reflexion of rule 30, and consequently their gliders can be coded in a similar way [15].

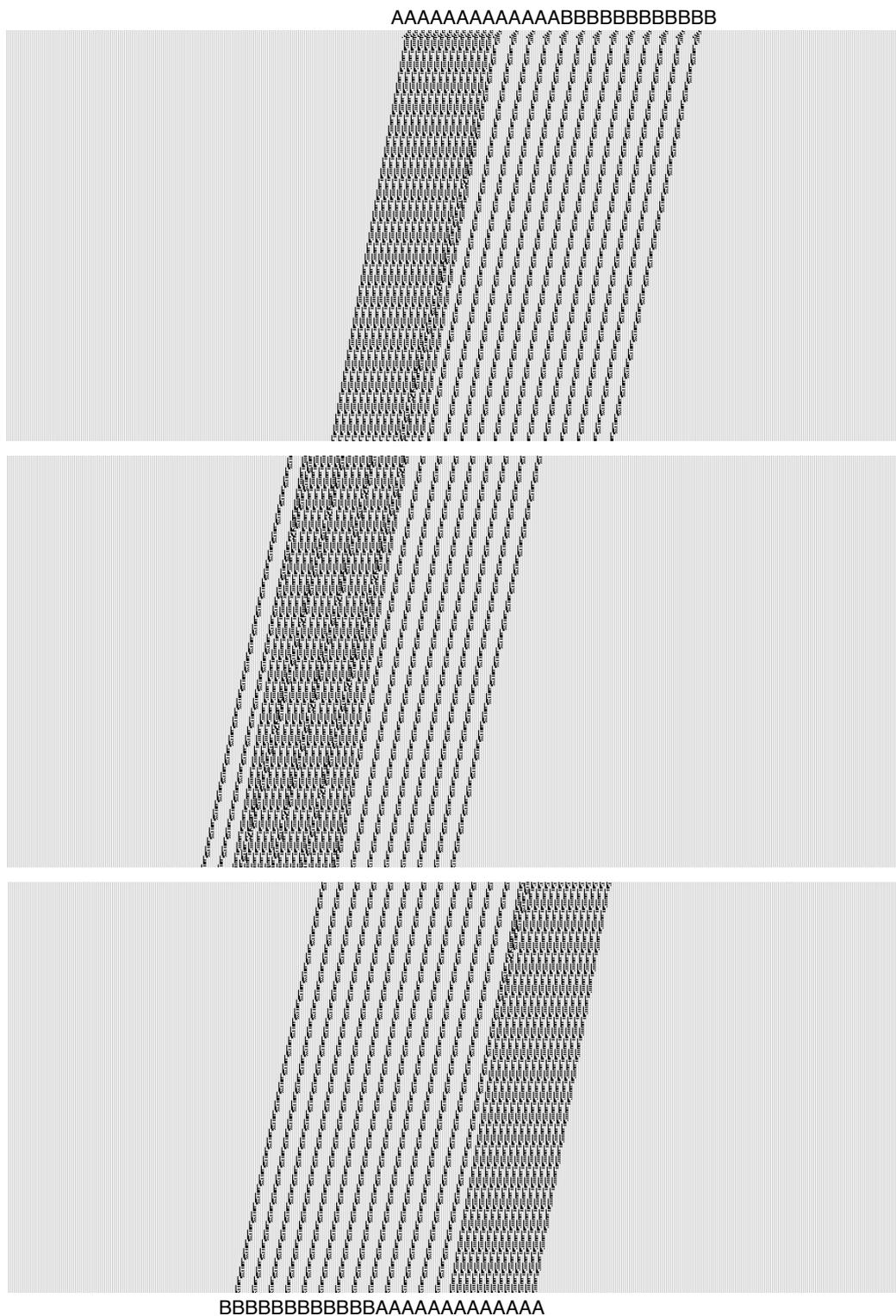


Fig. 9. A simple substitution system processing the word $A^{12}B^{12}$ to $B^{12}A^{12}$ with ECA $\phi_{majR86:8}$. The final production is reached by 6,888 generations with synchronization of soliton reactions.

non-trivial reactions across them. This can be substantiated by a number of different techniques, e.g. number-conservation [8], [13], exhaustive search [11], tiling [23], [18], de Bruijn diagrams [17], Z -parameter [31], genetic algorithms [10], mean field theory [20] or from a differential equations point view [9]. Thus the memory function ϕ offers a more easy way to get similar and, in some cases, more strong results reporting new complex rules in ECA with memory.

We have enriched some chaotic ECA rules with majority memory and demonstrated that by applying certain filtering procedures we can extract rich dynamics of travelling localizations, or particles.

Therefore, we can deduce a relation on chaotic systems decomposed in complex dynamics as a self-contained set. Generally a relation of sets of complex dynamics can be self-contained describing Φ as attractors, as a cascade (fig. 10 [6]).

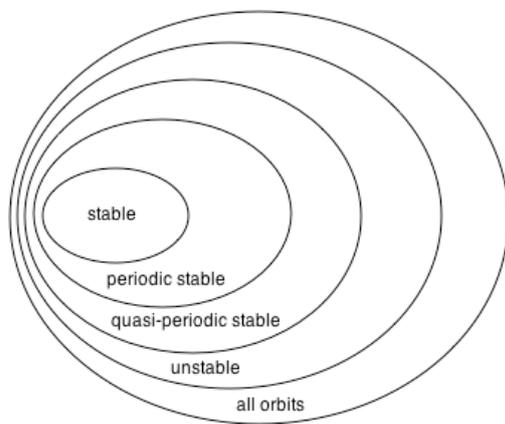


Fig. 10. Classes of global behavior.

This way, the most bigger set in fig. 10 ‘all orbits’ corresponds to complex dynamics and the ‘unstable’ set represents the chaotic systems. In fact there are a number of properties between orbits and characteristics that cannot be inferred directly. However the memory plays a role of a powerful tool to discover such properties. Finally, the memory ϕ can be applied to any CA or dynamical system.

ACKNOWLEDGMENT

G.J. Martínez is supported by EPSRC grant EP/F054343/1 and R. Alonso-Sanz by EPSRC grant EP/E049281/1.

REFERENCES

[1] S. Amoroso and G. Cooper, The Garden-of-Eden theorem for finite configurations, *Proceedings of the American Mathematical Society*, 1970.
 [2] A. Adamatzky, *Identification of Cellular Automata*, Taylor and Francis, London, 1994.
 [3] A. Adamatzky, (ed.) *Collision-Based Computing*, Springer, 2003.
 [4] R. Alonso-Sanz, Cellular automata with memory, In Meyers, R. (ed.) *Encyclopedia of Complexity and Systems Science*, Springer NY, (in press).
 [5] R. Alonso-Sanz, Elementary rules with elementary memory rules: the case of linear rules, *Journal of Cellular Automata* 1 71–87, 2006.

[6] A. Adamatzky, G.J. Martínez, and J.C. Seck-Tuoh-Mora, Phenomenology of reaction-diffusion binary-state cellular automata, *Int. J. Bifurcation and Chaos* 16 (10) 1–21, 2006.
 [7] Y. Bar-Yam, *Dynamics of Complex Systems*, Perseus Books, 1997.
 [8] N. Boccara and H. Fukš, Number-Conserving Cellular Automaton Rules, *Fundamenta Informaticae* 52 1–13, 2002.
 [9] L.O. Chua, *A Nonlinear Dynamics Perspective of Wolfram’s New Kind of Science*, World Scientific Publishing Company, 2007.
 [10] R. Das, M. Mitchell, and J.P. Crutchfield, A genetic algorithm discovers particle-based computation in cellular automata, *Lecture Notes in Computer Science* 866 344–353, 1994.
 [11] D. Eppstein, Searching for spaceships, *MSRI Publications* 42 433–452, 2002.
 [12] J.E. Hanson and J.P. Crutchfield, Computational Mechanics of Cellular Automata: An Example, *Physics D* 103 169–189, 1997.
 [13] K. Imai, A. Ikazaki, C. Iwamoto, and K. Morita, A logically universal number-conserving cellular automaton with a unary table-lookup function, *IEICE Trans. on Information and Systems E87-D* 694–699, 2004.
 [14] S.A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York, 1993.
 [15] G.J. Martínez, A. Adamatzky, A. Alonso-Sanz, and J.C. Seck-Tuoh-Mora, Complex dynamic emerging in Rule 30 with majority memory, *Complex Systems*, by publish.
 [16] G.J. Martínez, A. Adamatzky, and H.V. McIntosh, Phenomenology of glider collisions in cellular automaton Rule 54 and associated logical gates, *Chaos, Solitons and Fractals* 28 100–111, 2006.
 [17] G.J. Martínez, A. Adamatzky, and H.V. McIntosh, On the representation of gliders in Rule 54 by de Bruijn and cycle diagrams, *Lecture Notes in Computer Science* 5191 83–91, 2008.
 [18] M. Margenstern, *Cellular Automata in Hyperbolic Spaces*, Old City Publishing, Inc., 2007.
 [19] H.V. McIntosh, *One Dimensional Cellular Automata*, Luniver Press, 2009.
 [20] H.V. McIntosh, Wolfram’s Class IV and a Good Life, *Physica D* 45 105–121, 1990.
 [21] M. Minsky, *Computation: Finite and Infinite Machines*, Prentice Hall, 1967.
 [22] M. Mitchell, *Complexity: A Guided Tour*, Oxford University Press, 2009.
 [23] G.J. Martínez, H.V. McIntosh, and J.C. Seck-Tuoh-Mora, Gliders in Rule 110, *Int. J. of Unconventional Computing* 2 (1) 1–49, 2006.
 [24] M. Mitchell and M. Newman, Complex Systems Theory and Evolution, In *Encyclopedia of Evolution* (Pagel, M., ed.), Oxford University Press, 2002.
 [25] K. Morita, Cellular Automata and Artificial Life, *6th. Summer School on Complex Systems*, 1998.
 [26] C. Robinson, *Dynamical Systems: stability, symbolic dynamics and chaos*, CRC Press, 1994.
 [27] A. Wuensche and M. Lesser, *The Global Dynamics of Cellular Automata*, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company, 1992.
 [28] S. Wolfram, Universality and complexity in cellular automata, *Physica D* 10 1–35, 1984.
 [29] S. Wolfram, *Cellular Automata and Complexity*, Addison-Wesley Publishing Company, 1994.
 [30] A. Wuensche, Complexity in one-d cellular automata, *Santa Fe Institute working paper* 94-04-025, 1994.
 [31] A. Wuensche, Classifying Cellular Automata Automatically, *Complexity* 4 (3) 47–66, 1999.

Bifurcations of a neuron oscillator

Nathalie Corson

Abstract—This work adresses the study of the three-dimensional autonomous ordinary differential equations Hindmarsh-Rose neuronal model. General bifurcation diagrams are first given after a brief presentation of the model. Then, the existence of a Hopf bifurcation according to a small parameter which corresponds to the ratio of time scales between the fast and the slow dynamics is proved. Using the Hassard method we show that, under some conditions, a Hopf bifurcation occurs for a critical value of this parameter. The direction, stability and period of this bifurcation are also discussed. Numerical simulations are done to observe this bifurcation and to illustrate theoretical results.

Index Terms—Neuronal model, asymptotic dynamics, Hopf bifurcation.

I. INTRODUCTION

In 1952, a mathematical model that describes neuron activity has been given by two neurophysiologists, A.L. Hodgkin and A.F.Huxley, see [10]. Different neuron models have been then developed and studied, see for example [12], [13], [15] and references therein cited. In this paper, we focus on one of them, the Hindmarsh-Rose model (HR), which results from a simplification and a generalization of the Hodgkin-Huxley model, see [8], [9]. As observed in various biology systems, neuron activity presents different time scales. This can be explicitly observed in HR, which is a *slow-fast* autonomous three ordinary differential equations. The two first equations control the fast dynamics while the third one controls the slow one. Besides, periodic phenomena or oscillations are observed as in many natural systems such as neuron models. Those phenomena can be closely related to Hopf bifurcation.

The HR model reads as follows,

$$(HR) \begin{cases} \dot{x} &= y + ax^2 - x^3 - z + I \\ \dot{y} &= 1 - dx^2 - y \\ \dot{z} &= \epsilon(b(x - c_x) - z) \end{cases} \quad (1)$$

Parameters a, b and d are experimentally determined, c_x is the equilibrium x -coordinate of the two-dimensional system given by the two first equations of (1) when $I = 0$ and $z = 0$ and parameter I corresponds to the applied current. It is easy to experimentally change its value and it is therefore often used as the bifurcation parameter. Indeed, in the next part, bifurcation diagrams according to I are presented. Finally, parameter ϵ represents the ratio of time scales between fast and slow fluxes across the membrane of a neuron and, therefore, plays a very special role in neuron activity. It is chosen, in

nathalie.corson@univ-lehavre.fr
Laboratoire de Mathématiques Appliquées du Havre, 25 rue Philippe Lebon,
BP 540, 76058 Le Havre Cedex, France

this paper, as the bifurcation parameter, as in [2] or in [5], in which numerical simulations are done, among other, to study this system according to parameter ϵ . In the last section of this paper, parameters a, b, d and c_x are fixed as follows,

$$a = 3, \quad b = 4, \quad d = 5, \quad c_x = -\frac{1}{2}(1 + \sqrt{5}). \quad (2)$$

Equilibria are given by $\dot{x} = \dot{y} = \dot{z} = 0$, that is to say by,

$$x^3 + (d - a)x^2 + bx - bc_x - I - 1 = 0 \quad (3)$$

Let us denote,

$$\begin{aligned} x &= \xi + \frac{a - d}{3}, \\ p &= b - \frac{(a - d)^2}{3} \\ q &= -\frac{2(a - d)^3}{27} + \frac{b(a - d)}{3} - bc_x - I - 1 \end{aligned} \quad (4)$$

Then, (3) reads as, $\xi^3 + p\xi - q = 0$. Solving this equation gives the equilibria of system (1).

Proposition 1. *With notations (4), if $4p^3 + 27q^2 > 0$, then system (1) has a unique equilibrium $S_e = (x_e, y_e, z_e)$ given by,*

$$\begin{cases} x_e &= \left(-\frac{q}{2} + \left(\frac{q^2}{4} + \frac{p^3}{27} \right)^{\frac{1}{2}} \right)^{\frac{1}{3}} \\ &+ \left(-\frac{q}{2} - \left(\frac{q^2}{4} + \frac{p^3}{27} \right)^{\frac{1}{2}} \right)^{\frac{1}{3}} \\ &+ \frac{a - d}{3} \\ y_e &= 1 - dx_e \\ z_e &= b(x_e - c_x) \end{cases} \quad (5)$$

In the next section, a presentation of some bifurcation diagrams of the Hindmarsh-Rose model according to parameter I and parameter ϵ is done. Then the existence of a Hopf bifurcation according to parameter ϵ is studied. Indeed, even if this HR model dates from 1984 and has been widely numerically studied, see for example [1], [2], [4], [5], [11], no theoretical proof has ever been published as far as we know.

II. BIFURCATION DIAGRAMS

A bifurcation diagram shows the evolution of the asymptotic behaviour of solutions according to one parameter.

Parameter I corresponds to the current which is injected in the neuron. Thus, it can be controlled during experiments and can then play the role of bifurcation parameter.

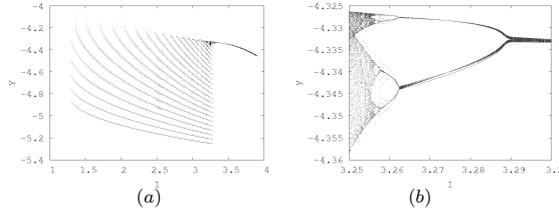


Figure 1. (a) Bifurcation diagram of the HR model for parameters (2) and $\epsilon = 0.001$. As the magnitude of injected current I increases, the number of branches on the diagram also increases. Biologically, the fast dynamics of the neuron is evolving. (b) Enlargements of (a) for $I \in [3.25; 3.3]$.

Figure 2(a) gives the bifurcation diagram with respect to the control parameter ϵ in the range $[0, 0.05]$. In order to have a more accurate analysis of the dynamics of system (1), we present in figure 2(b),(c),(d) enlargements of figure 2(a). Figure 2(b) shows, among other things, that there is an $\epsilon_1 \in [0.00041, 0.00049]$ for which the neuron behaviour changes abruptly. Indeed, $\forall \epsilon < \epsilon_1$, the neuron exhibits a tonic spiking motion and, $\forall \epsilon > \epsilon_1$, the neuron exhibits a bursting motion. Moreover, this figure shows that system (1) with parameters given in (2) and $I = 3.25$ does not exhibit chaotic behaviour for $\epsilon \in [\epsilon_1, 0.002]$.

The enlargement of figure 2(a) for $\epsilon \in [0.005, 0.015]$ shown in figure 2(c) exhibits not only inverse period doubling cascades starting with period 3, period 4 or period 5 but also some dark parts, which is a numerical sign of chaotic motion. Of course, this argument is not sufficient to claim that this system is chaotic for some given ranges of parameters. A more accurate study is done, for example, in [2].

The enlargement of figure 2(c) for $\epsilon \in [0.0138, 0.0148]$ shown in figure 2(d) also exhibits a chaotic behaviour of system (1). The right part of figure 2(a) exhibits a reverse period doubling cascade. As ϵ becomes larger, the number of spikes within a burst decreases until the bursting motion of the neuron disappears to let the spiking motion arises.

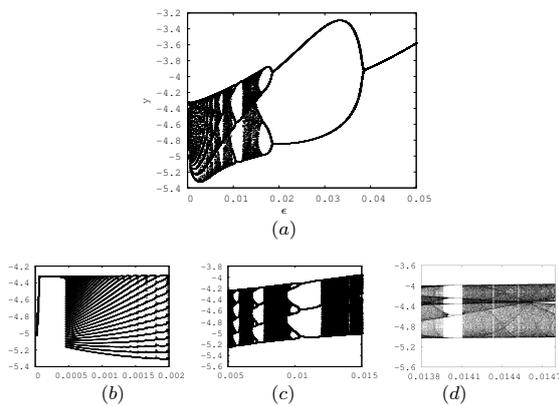


Figure 2. Bifurcation diagrams in (ϵ, y) plane for system (1) with parameters given in (2) and with $I = 3.25$. (a) An inverse period doubling cascade is observed for $\epsilon \in [0, 0.05]$. (b) Enlargement of figure (a) for $\epsilon \in [0; 0.002]$. (c) Enlargement of figure (a) for $\epsilon \in [0.005; 0.015]$. (d) Enlargement of figure (c) for $\epsilon \in [0.0138; 0.0147]$.

III. EXISTENCE, DIRECTION, STABILITY AND PERIOD OF A HOPF BIFURCATION ACCORDING TO ϵ

In this section, the existence, direction, stability and period of a Hopf bifurcation according to ϵ is studied, see [3].

Under the coordinate transformation, $x_1 = x - x_e$, $y_1 = y - y_e$ and $z_1 = z - z_e$, system (1) becomes,

$$\begin{cases} \dot{x}_1 = (2ax_e - 3x_e^2)x_1 + y_1 - z_1 \\ \quad + \hat{F}_1(x_1, y_1, z_1) \\ \dot{y}_1 = -2dx_ex_1 - y_1 + \hat{F}_2(x_1, y_1, z_1) \\ \dot{z}_1 = \epsilon bx_1 - \epsilon z_1 + \hat{F}_3(x_1, y_1, z_1) \end{cases} \quad (6)$$

where $\hat{F}_j(x_1, y_1, z_1)$, $j = 1, 2, 3$ are the nonlinear terms.

The Poincaré-Andronov-Hopf theorem applied to system (1) leads to the following proposition,

Proposition 2. With notations (12) and (13), if the two following conditions hold,

$$4r^3 + 27s^2 > 0 \quad (7)$$

$$\frac{2}{3}(a - d) < x_e < 0 \quad (8)$$

then, when parameter ϵ passes the value ϵ_c , system (1) undergoes a Hopf bifurcation at the equilibrium S_e , where,

$$\epsilon_c = \frac{-((1 - m_{11})^2 - m_{11}b) + \Delta^{\frac{1}{2}}}{2(1 - m_{11} + b)} \quad (9)$$

and $m_{11} = 2ax_e - 3x_e^2$, $m_{21} = -2dx_e$ and $\Delta = [(1 - m_{11})^2 - m_{11}b]^2 + 4(1 - m_{11} + b)(m_{11} + m_{21})(1 - m_{11})$.

Proof: The existence of a Hopf bifurcation point in system (1) is studied using the linearized system (6) at S_e . First of all, its jacobian matrix $M(\epsilon)$ is,

$$M(\epsilon) = (m_{ij})_{1 \leq i, j \leq 3} \quad (10)$$

The corresponding characteristic equation is,

$$f(\lambda(\epsilon)) = \lambda^3(\epsilon) + P(\epsilon)\lambda^2(\epsilon) + Q(\epsilon)\lambda(\epsilon) + R(\epsilon) \quad (11)$$

where,

$$\begin{aligned} P(\epsilon) &= 1 - m_{11} + \epsilon \\ Q(\epsilon) &= (1 - m_{11} + b)\epsilon - m_{11} - m_{21} \\ R(\epsilon) &= \epsilon(b - m_{11} - m_{21}) \end{aligned} \quad (12)$$

Setting,

$$\begin{aligned} \lambda(\epsilon) &= \nu(\epsilon) - \frac{P(\epsilon)}{P^2(\epsilon)} \\ r(\epsilon) &= Q(\epsilon) - \frac{P^2(\epsilon)}{3} \\ s(\epsilon) &= \frac{2P^3(\epsilon)}{27} - \frac{P(\epsilon)Q(\epsilon)}{3} + R(\epsilon) \end{aligned} \quad (13)$$

equation (11) reads as,

$$\nu^3(\epsilon) + r(\epsilon)\nu(\epsilon) + s(\epsilon) = 0,$$

which is the equation giving $M(\epsilon)$ eigenvalues.

The sign of $4r^3(\epsilon) + 27s^2(\epsilon)$ provides the number of real and complex eigenvalues of this matrix. Indeed, if

$4r^3(\epsilon) + 27s^2(\epsilon) > 0$, that is if condition (7) holds, then $M(\epsilon)$ has two complex eigenvalues $\lambda_{1,2}(\epsilon) = \alpha(\epsilon) + i\omega(\epsilon)$ and one real, $\lambda_3(\epsilon)$.

Now, let us study the existence of a critical value ϵ_c of parameter ϵ .

From (10), (11) and (12), polynomial rules lead to the existence of,

$$\epsilon_c = \frac{-((1 - m_{11})^2 - m_{11}b) \pm \Delta^{\frac{1}{2}}}{2(1 - m_{11} + b)}.$$

Algebraic computations show that under condition (8), $\epsilon_c > 0$.

Moreover, since $x_e < 0$, it is obvious that $m_{11} < 0$ and thus, $P(\epsilon_c) > 0$. Therefore, $\lambda_3(\epsilon) < 0$.

The derivative according to ϵ of the characteristic equation given in (11) is,

$$\begin{aligned} \frac{\partial f(\epsilon)}{\partial \epsilon} &= 3\lambda^2(\epsilon) \frac{\partial \lambda(\epsilon)}{\partial \epsilon} + \frac{\partial P(\epsilon)}{\partial \epsilon} \lambda^2(\epsilon) \\ &+ 2P(\epsilon) \lambda(\epsilon) \frac{\partial \lambda(\epsilon)}{\partial \epsilon} + \frac{\partial Q(\epsilon)}{\partial \epsilon} \lambda(\epsilon) \\ &+ Q(\epsilon) \frac{\partial \lambda(\epsilon)}{\partial \epsilon} + \frac{\partial R(\epsilon)}{\partial \epsilon}. \end{aligned} \quad (14)$$

Therefore, solving $\frac{\partial f}{\partial \epsilon}(\epsilon_c) = 0$ and separating imaginary and real parts, we obtain,

$$\frac{\partial \alpha}{\partial \epsilon}(\epsilon_c) = \frac{\frac{\partial R}{\partial \epsilon}(\epsilon_c) - \frac{\partial P}{\partial \epsilon}(\epsilon_c)Q(\epsilon_c) - P(\epsilon_c)\frac{\partial Q}{\partial \epsilon}(\epsilon_c)}{2Q(\epsilon_c) + 2P(\epsilon_c)^2}.$$

Since $2Q(\epsilon_c) + 2P(\epsilon_c)^2 > 0$, $\frac{\partial \alpha}{\partial \epsilon}(\epsilon_c) < 0$.

Finally, if $4r^3(\epsilon) + 27s^2(\epsilon) > 0$ and $\frac{2}{3}(a - d) < x_e < 0$, then all the conditions of the Poincaré-Andronov-Hopf theorem hold and (S_e, ϵ_c) is a Hopf bifurcation point of system (1). ■

Let us now study direction, stability and period of this Hopf bifurcation occurring at ϵ_c using Hassard method, see [7] and see also [6], [14], [16].

Let us denote by ω_0 the value $\omega(\epsilon_c) > 0$ and let v_j , $j = 1, 2, 3$, be the eigenvectors of the matrix $M(\epsilon_c)$, given in (10), corresponding to the eigenvalues λ_j . We have, $\lambda_j = \pm i\omega_0 = \pm iQ^{1/2}(\epsilon_c)$, $j = 1, 2$, and $\lambda_3 = -P(\epsilon_c)$.

The eigenvector v_1 associated with $\lambda_1 = i\omega_0$ is,

$$v_1 = \left(1, \frac{m_{21}(1 - i\omega_0)}{1 + \omega_0^2}, \frac{\epsilon_c b(\epsilon_c - i\omega_0)}{\epsilon_c^2 + \omega_0^2}\right)^T,$$

and the eigenvector v_3 associated with λ_3 is,

$$v_3 = \left(1, \frac{m_{21}}{m_{11} - \epsilon_c}, \frac{\epsilon_c b}{m_{11} - 1}\right)^T.$$

Let us define P such that $(x_1, y_1, z_1)^T = [P(x_2, y_2, z_2)]^T$,

$$P = (Re(v_1), -Im(v_1), v_3) = (p_{ij})_{1 \leq i, j \leq 3} \quad (15)$$

The inverse matrix is given by $P^{-1} = (p_{ij}^{-1})_{1 \leq i, j \leq 3}$.

Thus,

$$\begin{cases} \dot{x}_2 &= \omega_0 y_2 + F_1(x_2, y_2, z_2), \\ \dot{y}_2 &= -\omega_0 x_2 + F_2(x_2, y_2, z_2), \\ \dot{z}_2 &= \lambda_3 z_2 + F_3(x_2, y_2, z_2), \end{cases}$$

where F_1 , F_2 and F_3 are the nonlinear terms, satisfying $F_i(x_2, y_2, z_2) = P^{-1}(\bar{F}_i(x_1, y_1, z_1))$,

Procedures proposed by Hassard *et al.* [7] are used to calculate the following quantities, evaluated at $\epsilon = \epsilon_c$.

$$\begin{aligned} g_{11} &= \frac{1}{2} [p_{11}^{-1}(a - 3x_e) - p_{12}^{-1}d] \\ &+ \frac{i}{2} [p_{21}^{-1}(a - 3x_e) - p_{22}^{-1}d] \\ g_{02} &= g_{11} \\ g_{20} &= g_{11} \\ G_{21} &= -\frac{3}{4} (p_{11}^{-1} + ip_{21}^{-1}). \end{aligned}$$

Moreover, let us calculate the quantities,

$$\begin{aligned} h_{11} &= \frac{1}{2} [p_{31}^{-1}(a - 3x_e) - p_{32}^{-1}d] \\ h_{20} &= h_{11}. \end{aligned}$$

Then, solving the two equations,

$$\lambda_3 w_1 = -h_{11}$$

$$(\lambda_3 - 2i\omega_0)w_{20} = -h_{20}$$

gives,

$$\begin{aligned} w_{11} &= \frac{[p_{31}^{-1}(a - 3x_e) - p_{32}^{-1}d]}{2(1 - m_{11} + \epsilon_c)}, \\ w_{20} &= \frac{1}{2} \frac{[p_{31}^{-1}(a - 3x_e) - p_{32}^{-1}d]}{(1 - m_{11} + \epsilon_c)^2 + 4\omega_0^2} \\ &\quad \cdot (1 - m_{11} + \epsilon_c + 2i\omega_0) \end{aligned}$$

Furthermore, calculating the quantities,

$$\begin{aligned} G_{110} &= \frac{1}{4} [p_{11}^{-1}(a - 3x_e) - p_{12}^{-1}d] \\ &+ \frac{i}{4} [p_{21}^{-1}(a - 3x_e) - p_{22}^{-1}d] \\ G_{101} &= G_{110} \\ g_{21} &= G_{21} + (2G_{110}w_{11} + G_{101}w_{20}) \end{aligned}$$

and by setting,

$$c_1 = \frac{i}{2\omega_0} \left[g_{20}g_{11} - 2|g_{11}|^2 - \frac{1}{3}|g_{02}|^2 \right] + \frac{1}{2}g_{21}$$

we can give the main result below in which,

$$\begin{aligned} \mu_2 &= -\frac{Re(c_1)}{\frac{\partial \alpha}{\partial \epsilon}(\epsilon_c)}, \\ \tau_2 &= -\frac{Im(c_1) + \mu_2 \frac{\partial \omega}{\partial \epsilon}(\epsilon_c)}{\omega_0}, \\ \beta_2 &= 2Re(c_1). \end{aligned}$$

Theorem 1. *Under the hypothesis of proposition 2, system (1) undergoes a Hopf bifurcation at the equilibrium point*

(x_e, y_e, z_e) as ϵ passes through ϵ_c with the following properties.

- 1) If $\mu_2 < 0$ (reps. $\mu_2 > 0$), then the direction of bifurcation is $\epsilon < \epsilon_c$ (resp. $\epsilon > \epsilon_c$) and the bifurcation is supercritical (resp. subcritical),
- 2) If $\beta_2 < 0$ (resp. $\beta_2 > 0$), the bifurcating periodic solutions are orbitally stable (resp. unstable),
- 3) If $\tau_2 > 0$ (resp. $\tau_2 < 0$), the period of bifurcating periodic solutions increases (resp. decreases).

The period and characteristic exponents are given by,

$$\begin{aligned} T &= \frac{2\pi}{\omega_0}(1 + \tau_2 E^2 + O(E^4)) \\ \beta &= \beta_2 E^2 + O(E^4) \end{aligned}$$

Where, $E^2 = \frac{\epsilon - \epsilon_c}{\mu_2} + O(\epsilon - \epsilon_c)^2$ (provided $\mu_2 \neq 0$).

The periodic solutions themselves are,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_e \\ y_e \\ z_e \end{pmatrix} + P \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad (16)$$

where,

$$\begin{aligned} u_1 &= \text{Re}(\zeta), & u_2 &= \text{Im}(\zeta) \\ u_3 &= w_{11}|\zeta|^2 + \text{Re}(w_{20}\zeta^2) + O(|\zeta|^3) \end{aligned}$$

and

$$\begin{aligned} \zeta &= Ee^{2i\pi t/T} + \frac{iE^2}{6\omega_0}(g_{02}e^{-4i\pi t/T} \\ &\quad - 3g_{20}e^{4i\pi t/T} + 6g_{11}) + O(E^3) \end{aligned}$$

Now, numerical computations are done to illustrate these theoretical results. Thereafter, we consider system (1), with parameters a, b, c_x and d fixed as in (2) and $I = 3.25$. Equilibria of this system are studied as presented in the first section of this paper, and using notation of (4), we obtain,

$$4p^3 + 27q^2 \approx 76.443755 > 0$$

Therefore, proposition 1 leads to the existence an uniqueness of system (1) equilibrium (x_e, y_e, z_e) , given by (5),

$$x_e \approx -0.722126, \quad y_e \approx -1.607329, \quad z_e \approx 3.583632$$

Let us verify if proposition 2 can be applied to system (1) with the fixed values of parameters (2).

The bifurcation value of ϵ , given by equation (9) of proposition 2, is,

$$\epsilon_c \approx 0.125912.$$

This value of ϵ_c is really close to the one we observe on the bifurcation diagram given in figure 3.

For this value of ϵ_c , we have, $4r^3(\epsilon_c) + 27s^2(\epsilon_c) \approx 443.299666 > 0$ and condition (7) holds. Therefore, the jacobian matrix $M(\epsilon_c)$ has one real eigenvalue $\lambda_3(\epsilon_c)$ and two complex ones $\lambda_{1,2} = \alpha(\epsilon_c) \pm i\omega(\epsilon_c)$.

Since $2(a-d)/3 = -4/3$, the condition $2(a-d)/3 < x_e < 0$ is verified.

Furthermore, $\frac{\partial \alpha}{\partial \epsilon}(\epsilon_c) \approx -0.748444 \neq 0$, and $\lambda_3(\epsilon_c) \approx -7.025406 < 0$.

Thus, thanks to proposition 2, $\epsilon_c \approx 0.125912$ is the Hopf bifurcation value of parameter ϵ for system (1) at $(-0.722126, -1.607329, 3.583632)$.

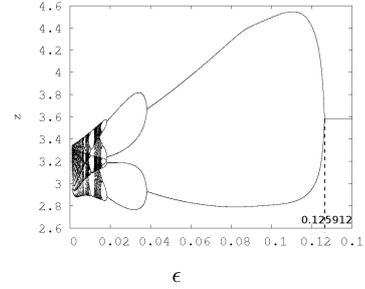


Figure 3. Bifurcation diagram of system (1) with parameters given in (2) according to parameter ϵ .

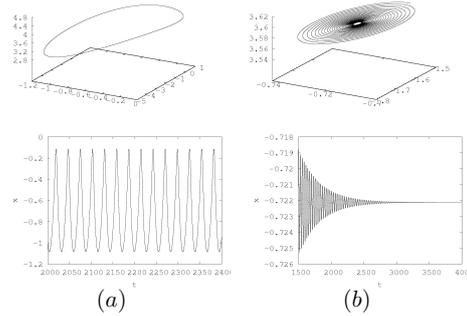


Figure 4. (a) (x, y, z) view of the phase portrait and time series of system (1) with parameter fixed as in (2) and $\epsilon = 0.12 < \epsilon_c$, the asymptotic solution is a stable limit cycle. (b) (x, y, z) view of the phase portrait and time series of system (1) with parameter fixed as in (2) and $\epsilon = 0.13 > \epsilon_c$, the asymptotic solution is stable focus.

The computation of matrix P and its inverse matrix P^{-1} gives,

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 6.890683 & 1.509269 & -1.198934 \\ 0.993530 & 1.728299 & -0.073022 \end{pmatrix}$$

and ,

$$P^{-1} = \begin{pmatrix} 0.158582 & 0.139699 & -0.121995 \\ -0.055612 & -0.086210 & 0.653888 \\ 0.841418 & -0.139699 & 0.121995 \end{pmatrix}$$

The various useful coefficient presented in the previous section are then,

$$\begin{aligned}
g_{11} &= 0.060399 + 0.071870i \\
G_{21} &= -0.118936 + 0.041709i \\
h_{11} &= 2.522790 \\
w_{11} &= 0.359215 \\
w_{20} &= 0.357823 + 0.022319i \\
G_{110} &= 0.030199 + 0.035935i \\
g_{21} &= -0.087236 + 0.081058i \\
c_1 &= -0.063437 - 0.009879i
\end{aligned}$$

Finally, computations give,

$$\begin{aligned}
\mu_2 &= -0.084758 < 0, \\
\beta_2 &= -0.126873 < 0, \\
\tau_2 &= 0.384677 > 0.
\end{aligned}$$

According to theorem 1, the Hopf bifurcation occurring at ϵ_c is supercritical and the direction of bifurcation is $\epsilon < \epsilon_c$. Moreover, the bifurcating periodic solutions are asymptotically orbitally stable and the period of bifurcating periodic solutions increases.

The period of the solution is given by,

$$T = 28.686363 - 130.193932(\epsilon - 0.125912) + O((\epsilon - 0.125912)^2)$$

and this period increases as ϵ decreases. The characteristic exponents is given by,

$$\beta = 1.496888(\epsilon - 0.125912) + O((\epsilon - 0.125912)^2)$$

The periodic solutions are,

$$\begin{aligned}
\begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -0.722126 \\ -1.607329 \\ 3.583632 \end{pmatrix} \\
&+ \begin{pmatrix} 1 & 0 & 1 \\ 6.890683 & 1.509269 & -1.198934 \\ 0.993530 & 1.728299 & -0.073022 \end{pmatrix} \\
&\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}
\end{aligned}$$

where,

$$\begin{aligned}
u_1 &= Re(\zeta), \\
u_2 &= Im(\zeta), \\
u_3 &= 0.359215|\zeta|^2 + Re((0.357823 + 0.022319i)\zeta^2) + O(|z\eta|^3)
\end{aligned}$$

and,

$$\begin{aligned}
\zeta &= Ee^{2i\pi t/T} \\
&+ \frac{iE^2}{6\omega_0} \left((0.060399 + 0.071870i)e^{-4i\pi t/T} \right. \\
&- 3(0.060399 + 0.071870i)e^{4i\pi t/T} \\
&\left. + 6(0.060399 + 0.071870i) \right) + O(E^3)
\end{aligned}$$

IV. CONCLUSION

In this paper, after a presentation of the three-dimensional autonomous ordinary differential equations Hindmarsh-Rose neuronal model, bifurcation diagrams according to parameters I and ϵ are presented. Then, the existence of a Hopf bifurcation according to parameter ϵ in this model is discussed. Indeed, for

a critical value ϵ_c of this parameter, a Hopf bifurcation occurs under some conditions. Using Hassard algorithm, the direction, stability and period of this bifurcation are then studied. Finally, numerical simulations are done to observe this bifurcation and to illustrate theoretical results.

REFERENCES

- [1] P. Arena, L. Fortuna, M. Frasca, M. La Rosa, Locally active Hindmarsh-Rose neurons, *Chaos Sol. and Fract.* 27 (2006) 405-412.
- [2] N. Corson, M.A. Aziz-Alaoui, Asymptotic dynamics of the slow-fast Hindmarsh-Rose neuronal system, *DCDIS-B* (2009) to appear.
- [3] N. Corson, N. Verdière, M.A. Aziz-Alaoui, Hopf bifurcation in the slow-fast Hindmarsh-Rose neuronal system, (2009) submitted.
- [4] L-X. Duan, Q-S. Lu, Codimension-two bifurcation analysis in Hindmarsh-Rose model with two parameters, *Chin. Phy. Lett.* vol. 22 (2005) 1325-1328.
- [5] J.M. Gonzalez-Miranda, Complex bifurcation structures in the Hindmarsh-Rose neuron model, *Int. J. Bif. Chaos* vol. 17, 9 (2007) 3071-3083.
- [6] B.D. Hassard, Y.H. Wan, Bifurcation formulae derived from center manifold theory, *J. Math. Anal. Appl.* 63 (1978) 297-312.
- [7] B.D. Hassard, N.D. Kazarinoff, Y. H. Wan, *Theory and Applications of Hopf bifurcation*, London Mathematical Society Lecture Note Series 41, London, 1981.
- [8] J.L. Hindmarsh, R.M. Rose, A model of the nerve impulse using two first-order differential equations, *Nature* vol. 296, (1982) 162-164.
- [9] J.L. Hindmarsh, R.M. Rose, A model of neuronal bursting using three coupled first order differential equations, *Proc. R. Sc. Lond. B221* (1984) 87-102.
- [10] A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol* 117 (1952) 500-544.
- [11] G. Innocenti, A. Morelli, R. Genesio, A. Torcini, Dynamical phases of the Hindmarsh-Rose neuronal model: Studies of the transition from bursting to spiking chaos, *Chaos* 17 (2007) 043128 1 - 043128 11.
- [12] E.M. Izhikevich, Which model to use for cortical spiking neurons, *IEEE Trans. on Neuron Net.* 15 (5) (2004) 1063-1070.
- [13] E.M. Izhikevich, *Dynamical systems in Neuroscience*, MIT Press, Cambridge, 2007.
- [14] J. Lü, T. Zhou, G. Chen, S. Zhang, Local bifurcation of the Chen system, *Int. J. Bif. Chaos* vol. 12, 10 (2002) 2257-2270.
- [15] C. Morris, H. Lecar, Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* 35 (1981) 193-213.
- [16] X. Zhou, Y. Wu, Y. Li, Z. Wei, Hopf bifurcation of the Liu system, *Chaos Sol. and Fract.* 36 (2008) 1385-1391.

The chikungunya epidemic: modelling and dynamics

Djamila Moulay

Abstract—Models for the transmission of the chikungunya virus to human population are done. The chikungunya virus is an alpha arbovirus, first identified in 1953, transmitted by *Aedes*, mosquitoes, responsible for a little documented uncommon acute tropical disease. Models describing the mosquito population dynamics and the virus transmission to human population are discussed. Global analysis of endemic equilibria is given, which use on the one hand Lyapunov function techniques and on the other hand results borrowed to theory of competitive systems and satbility of periodic orbits.

Index Terms—chikungunya virus, Epidemiologic model, Endemic equilibrium, Global stability, Competitive system.

I. INTRODUCTION

An unprecedented chikungunya epidemic has appeared on the Reunion Island (775,000 inhabitants) with over 244,000 reported and 205 deaths (directly or indirectly linked) as of April 20 2006. *Aedes albopictus*, see [1], long present on the island, is the main vector of this disease. After the Grande Comore Island epidemic, the first cases were reported in the Reunion Island in March 2005. This is the first time that a chikungunya epidemic was described in this part of the world. The Asian tiger mosquito or forest day mosquito (*Aedes albopictus*), from the mosquito family Culicidae, is characterized by its black and white striped legs, small black and white body. It is native to the tropical and subtropical areas of Southeast Asia. In the past couple of decades this species has invaded many countries throughout the world, through the transport of goods and increasing international travel.

The chikungunya is an arthropod-borne viral disease (arbovirus). The name is derived from the Makonde word meaning “that which bends up” in reference to the stooped posture developed as a result of the arthritic symptoms of the disease. It was first described by Marion Robinson and W.H.R. Lumsden [2], following an outbreak in 1952 on the Makonde Plateau. Some Arboviruses are able to cause emergent diseases. Arthropods are able to transmit the virus upon biting, allowing it to enter the bloodstream which can cause viraemia. The dynamics of arboviral disease like Dengue, [3],[4] or chikungunya [5] are influenced by many factors such as humans, the mosquito vector and the virus, as well as the environment which affects all the present mechanisms directly or indirectly. This paper deals with a succession of two model involving differential equations for the mosquito population and virus transmission to human population and is organized

D. Moulay is with Laboratoire de Mathématiques Appliquées du Havre (LMAH), University of Le Havre, France, e-mail: djamila.moulay@univ-lehavre.fr.

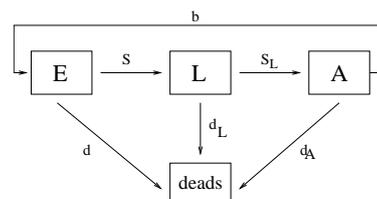


Fig. 1. A stage structured model for *Aedes Albopictus* population dynamics. E states for eggs, L for larvae and pupae, A for female adult. s, s_L, b, d, d_L, d_A are system parameters.

In the diagram:

- b = eggs laying rate;
- s, s_L = transfer rates;
- d, d_L, d_A = mortality rates of eggs, larvae and adult population.

as follows. The second section deals with the formulation of the dynamical models, first of all for the population growth that is the *Aedes Albopictus* mosquito, and secondly for the virus transmission to human population. The first model use an stage structured model and the second SI and SIR type models. The third section is devoted to the mathematical analysis of both models, focusing on the boundedness, the positivity of the solutions, local or global satbility for the endemic equilibria. For the first system we use the Lyapunov theory to establish the global stability of the endemic equilibrium, while for the second we use the general theory of competitive systems and compound matrix. As usual in mathematical epidemiology studies, we have also found thresholds parameters which determine the global dynamical behaviour.

II. VECTOR POPULATION AND TRANSMISSION MODELLING

A. Formulation of a dynamical model for vector population growth

To describe the *Aedes Albopictus* population dynamic we use a stage structured model, which consider three main stages (see Fig. 1), embryonic (E), larvae (L, which consists here of the larvae and pupae populations) and adult (A), see [6] for more details.

We assume that the number of laid eggs is proportional to the number of females. Moreover, it has been observed that mosquitoes are able to detect the best breeder site for the eggs developpement. Indeed if there are too much eggs in the oviposition habitat or too few nutrients and water ressources, then females laid less eggs or choose an over site. It seems reasonable to express this biological phenomenon with

a mathematical model which explicitly incorporates the idea of limited carrying capacity resources. This model should take into account the availability of nutrients and the occupation by eggs or larvae of the available breeder sites.

That is why we assume that,

- *per capita* oviposition rate is then given by,

$$b \left(1 - \frac{E(t)}{K_E} \right) A(t)$$

where K_E is the availability of nutrients and space, b represents the rate at which the population would grow if it were unencumbered by environmental degradation,

- the number of possible larvae (only a part of the eggs class became larvae), see [6], is given by,

$$s \left(1 - \frac{L(t)}{K_L} \right) E(t)$$

where K_L is the availability of nutrients and space, and s the transfert rate from eggs class to larvae class.

The above hypothesis lead to the following equations,

$$\begin{cases} \frac{dE}{dt}(t) = b \left(1 - \frac{E(t)}{K_E} \right) A(t) - (s + d)E(t) \\ \frac{dL}{dt}(t) = s \left(1 - \frac{L(t)}{K_L} \right) E(t) - (s_L + d_L)L(t) \\ \frac{dA}{dt}(t) = s_L L(t) - d_A A(t) \end{cases} \quad (1)$$

This system is defined on the bounded subset of \mathbb{R}^3 , see [6],

$$\Delta = \left\{ (E, L, A) \mid \begin{array}{l} 0 \leq E \leq K_E \\ 0 \leq L \leq K_L \\ 0 \leq A \leq \frac{s_L}{d_A} K_L \end{array} \right\}$$

B. A compartmental model for the virus transmission to human population

Denote by N_H the human population size for which we assume an exponential growth. Then, its dynamics is described by,

$$\frac{dN_H}{dt}(t) = (b_H - d_H)N_H(t)$$

where b_H and d_H are the birth and natural death rates, respectively.

Let \bar{S}_H , \bar{I}_H and \bar{R}_H denote the total number of susceptibles, infectives, and immunes in the population, respectively, and \bar{S}_m , \bar{I}_m be the total number of susceptible and infective mosquitoes. The immune class in the vector population does not exist, since mosquitoes carry the infection along their life. The model is schematically represented in Fig.2.

The effective contact rate β_H is the average number of contacts per day that would result in infection if the vector is infectious. The effective contact rate β_m is the average number of contacts per day that effectively transmits the infection to vectors. β_H and β_m depends on the average number of bites per mosquito per day and the proportion of bites that result in vector infection.

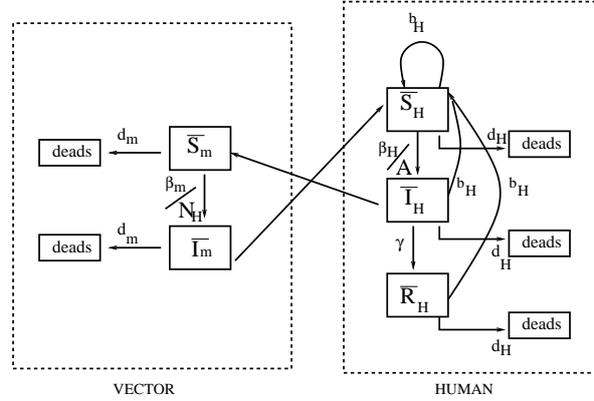


Fig. 2. A compartmental model for the chikungunya virus transmission

- β_m = effective contact rate between susceptible vectors and humans
- β_H = effective contact rate between susceptible humans and vectors
- γ = recovery rate of infected humans
- d_H =mortality rate of human population
- d_m =mortality rate of human vector population

The above hypotheses lead to the following equations,

$$\begin{cases} \frac{d\bar{S}_H}{dt}(t) = b_H(\bar{S}_H(t) + \bar{I}_H(t) + \bar{R}_H(t)) - \beta_H \frac{\bar{I}_m(t)}{A(t)} \bar{S}_H(t) - d_H \bar{S}_H(t) \\ \frac{d\bar{I}_H}{dt}(t) = \beta_H \frac{\bar{I}_m(t)}{A(t)} \bar{S}_H(t) - \gamma \bar{I}_H(t) - d_H \bar{I}_H(t) \\ \frac{d\bar{R}_H}{dt}(t) = \gamma \bar{I}_H(t) - d_H \bar{R}_H(t) \\ \frac{d\bar{S}_m}{dt}(t) = s_L L(t) - d_m \bar{S}_m(t) - \beta_m \frac{\bar{I}_H(t)}{N_H(t)} \bar{S}_m(t) \\ \frac{d\bar{I}_m}{dt}(t) = \beta_m \frac{\bar{I}_H(t)}{N_H(t)} \bar{S}_m(t) - d_m \bar{I}_m(t) \end{cases} \quad (2)$$

All parameters in this model are non-negative. We assume that human and vector populations remain constant. Therefore, without loss of generality, we can introduce the proportions $S_H = \bar{S}_H/N_H$, $I_H = \bar{I}_H/N_H$, $R_H = \bar{R}_H/N_H$, $S_m = \bar{S}_m/A$, $I_m = \bar{I}_m/A$ in system (2), and using the relation $S_H + I_H + R_H = 1$, we obtain the following system that describes the dynamics of the proportion of individuals in each class,

$$\left\{ \begin{array}{l} \frac{dE}{dt}(t) = bA(t) \left(1 - \frac{E(t)}{K_E}\right) - (s+d)E(t) \\ \frac{dL}{dt}(t) = sE(t) \left(1 - \frac{L(t)}{K_L}\right) - (s_L + d_L)L(t) \\ \frac{dA}{dt}(t) = s_L L(t) - d_m A(t) \\ \frac{dS_H}{dt}(t) = -(bs_H + \beta_H I_m(t)) S_H(t) + b_H \\ \frac{dI_H}{dt}(t) = \beta_H I_m(t) S_H(t) - (\gamma + b_H) I_H(t) \\ \frac{dI_m}{dt}(t) = - \left(s_L \frac{L(t)}{A(t)} + \beta_m I_H(t) \right) I_m(t) + \beta_m I_H(t) \end{array} \right. \quad (3a)$$

This system is defined on the bounded subset of \mathbb{R}_+^6 , which is the region of biological interest, $\Delta \times \Omega$ where

$$\Omega = \{(S_H, I_H, I_m) \in \mathbb{R}_+^3 : 0 \leq S_H + I_H \leq 1, 0 \leq I_m \leq 1\}. \quad (4)$$

Obviously, this model may be enhanced by taking into account the delay between the transfer to mosquitoes and the transmission to humans (from five to six days). In this case the model would be of type SEI for the vector (see [5]). However if we consider a huge mosquito population, the number of mosquitoes in state E (exposed) can be neglected in comparison to the whole population.

The next sections give the mathematical results on those models. For the proofs see [6].

III. ANALYSIS OF THE POPULATION DYNAMIC MODEL

We will investigate the asymptotic behaviour of orbits starting in the non-negative cone :

$$\mathbb{R}_+^3 = \{(x, y, z) \in \mathbb{R}^3 : x \geq 0, y \geq 0, z \geq 0\}$$

$$\mathbb{R}_+^3 = \{(x, y, z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0\}$$

Lemma III.1. Let $(t_0, X_0 = (E_0, L_0, A_0)) \in \mathbb{R}_+ \times \mathbb{R}_+^3$ and $([t_0, T], X = (E, L, A))$ ($T \in]t_0, +\infty[$) be the maximal solution of the Cauchy problem associated to (1) with the initial condition (t_0, X_0) .

Then,

$$\forall t \geq t_0, X(t) \in \mathbb{R}_+^3.$$

Lemma III.2. The set,

$$\Delta = \left\{ (E, L, A) \mid \begin{array}{l} 0 \leq E \leq K_E \\ 0 \leq L \leq K_L \\ 0 \leq A \leq \frac{s_L}{d_A} K_L \end{array} \right\}$$

is an invariant region under the flow induced by (1).

Proposition III.3. The set Δ is an attraction basin of system (1).

We shall use the following threshold parameter for computing the steady state,

$$r = \left(\frac{b}{s+d} \right) \left(\frac{s}{s_L + d_L} \right) \left(\frac{s_L}{d_A} \right) \quad (5)$$

Proposition III.4. System (1) always has the disease free equilibrium $N_0^* = (0, 0, 0)$.

- If $r \leq 1$, then system (1) has no other equilibria.
- If $r > 1$, there is a unique endemic equilibrium

$$N^* = \left(1 - \frac{1}{r} \right) \begin{pmatrix} \frac{K_E}{\gamma_E} \\ \frac{K_L}{\gamma_L} \\ \frac{s_L K_L}{d_A \gamma_L} \end{pmatrix} = \begin{pmatrix} E^* \\ L^* \\ A^* \end{pmatrix}$$

where,

$$\gamma_E = 1 + \frac{(s+d)d_A K_E}{bs_L K_L} \quad \text{and} \quad \gamma_L = 1 + \frac{(s_L + d_L)K_L}{sK_E}$$

Proposition III.5. The equilibrium $N_0^* = (0, 0, 0)$ is locally asymptotically stable if $r \leq 1$ and unstable if $r > 1$.

Proposition III.6. If $r > 1$, N^* is locally asymptotically stable.

Proposition III.7. If $r > 1$ the equilibrium N^* is globally asymptotically stable (GAS) in $\text{int}(\Delta)$.

IV. DYNAMICS ANALYSIS OF THE TRANSMISSION MODEL

We will focus on the case $r > 1$, r given by (5) which is the condition of survival of all populations as we studied in the previous section.

Our work addresses the existence and stability of equilibrium point of (3). For this aim, we shall use the following threshold parameter,

$$R_0 = \frac{\beta_m \beta_H}{d_A (\gamma + b_H)} = \frac{\beta_m \beta_H}{s_L \frac{L^*}{A^*} (\gamma + b_H)} \quad (6)$$

A. Existence and GAS of the equilibria

Proposition IV.1. We assume that $r > 1$. System (3) always has the disease free equilibrium (E^*, L^*, A^*, X_1^*) where $X_1^* = (1, 0, 0)$.

- If $R_0 \leq 1$, then system (1) has no other equilibria.
- If $R_0 > 1$, there is a unique endemic equilibrium (E^*, L^*, A^*, X_2^*) where $X_2^* = (S_H^*, I_H^*, I_m^*)$, and

$$\begin{aligned} S_H^* &= \frac{b_H}{\beta_H + b_H} + \frac{\beta_H}{(\beta_H + b_H)R_0} \\ I_H^* &= \frac{d_m b_H}{\beta_m (\beta_H + b_H)} (R_0 - 1) \\ I_m^* &= \frac{b_H}{\beta_H + b_H R_0} (R_0 - 1) \end{aligned}$$

where (E^*, L^*, A^*) is the endemic equilibrium of the independent system (1).

System (3) is the coupling of the two subsystems (3a) and (3b), for which the coupling term is the function $s_L \frac{L(t)}{A(t)} I_m(t)$, that is system (3a) drives system (3b). Therefore, since the previous section was devoted to the study of the subsystem

(3a) corresponding to the population dynamic we only have to analyse the subsystem (3b),

$$\begin{cases} \frac{dS_H}{dt}(t) &= -(b_H + \beta_H I_m(t)) S_H(t) + b_H \\ \frac{dI_H}{dt}(t) &= \beta_H I_m(t) S_H(t) - (\gamma + b_H) I_H(t) \\ \frac{dI_m}{dt}(t) &= -\left(s_L \frac{L(t)}{A(t)} + \beta_m I_H(t)\right) I_m(t) + \beta_m I_H(t) \end{cases} \quad (7)$$

We use useful results given in [7] or [8], for our system.

Since N^* , the endemic equilibrium of subsystem (3a), is GAS for $r > 1$, then $\frac{L(t)}{A(t)} \rightarrow \frac{L^*}{A^*}$ as $t \rightarrow +\infty$. Therefore system (3b) is a three-dimensional asymptotically autonomous differential system with limit system:

$$\begin{cases} \frac{dS_H}{dt}(t) &= -b_H(1 - S_H(t)) - \beta_H I_m(t) S_H(t) \\ \frac{dI_H}{dt}(t) &= \beta_H I_m(t) S_H(t) - (\gamma + b_H) I_H(t) \\ \frac{dI_m}{dt}(t) &= -\left(s_L \frac{L^*}{A^*} + \beta_m I_H(t)\right) I_m(t) + \beta_m I_H(t) \end{cases} \quad (8)$$

First of all, note that the region of biological interest

$$\Omega = \{(S_H, I_H, I_m) \in \mathbb{R}_+^3 : 0 \leq S_H + I_H \leq 1, 0 \leq I_m \leq 1\}.$$

is positively invariant under the flow induced by (8), as the vector field on the boundary does not point to the outside of Ω .

Theorem IV.2. *If $R_0 > 1$, the endemic equilibrium is globally asymptotically stable in $\text{int}(\Omega)$.*

To prove this theorem we use some results about competitive systems, given in [9], [10] and stability of periodic orbits.

Let $D \subset \mathbb{R}^n$ be an open set, and $x \mapsto f(x) \in \mathbb{R}^n$ be a C^1 function defined in D . We consider the autonomous system in \mathbb{R}^n given by,

$$x' = f(x) \quad (9)$$

Definition IV.3. *We say that system (9) has the property of stability of periodic orbits, iff the orbit of any periodic solution $\gamma(t)$, if it exists, is asymptotically orbitally stable.*

The following theorem is the main tool to prove the global stability of the endemic equilibrium.

Theorem IV.4. *Assume that $n=3$, D convex and bounded. Suppose that (9) is competitive, persistent and has the property of stability of periodic orbits. If x_0 is the only equilibrium in $\text{int}(D)$, and if it is locally asymptotically stable, then it is globally asymptotically stable in $\text{int}(D)$.*

Now, let us apply those results to the GAS study of X_2^* . The proof of this theorem is similar to the one in [11]. In order to prove the persistence of system (8), we shall prove the following proposition.

Proposition IV.5. *On the boundary of Ω , system (8) has only one ω -limit point which is the equilibrium X_1^* . Moreover for $R_0 > 1$, X_1^* cannot be the ω -limit of any orbit in $\text{int}(\Omega)$.*

Proposition IV.6. *The equilibrium E_0 is globally asymptotically stable in Ω if $R_0 \leq 1$, and unstable if $R_0 > 1$.*

Proposition IV.7. *The system (8) has the property of stability of periodic orbits with asymptotic phase.*

Proof: A sufficient condition for a periodic orbit $\gamma = \{p(t) : 0 \leq t \leq \omega\}$ of (9) to be asymptotically orbitally stable is that the linear non-autonomous system,

$$y'(t) = \frac{\partial f^{[2]}}{\partial x}(p(t))y(t) \quad (10)$$

is asymptotically stable.

Equation (10) is called the *second compound equation* of (9) and $\partial f^{[2]}/\partial x$ is the *second compound matrix* of the jacobian matrix $\partial f^{[2]}/\partial x$ of f . ■

Theorem IV.8. *Consider system (8). If $R_0 > 1$, then $\Omega - \{(S_H, 0, 0) : 0 \leq S_H \leq 1\}$ is an asymptotic stability region for the endemic equilibrium X_2^* . Moreover all trajectories starting in the S_H -axis approach the disease-free equilibrium X_1^* .*

V. CONCLUSION

We have proposed models to describe the vector population dynamic and the chikungunya virus transmission to human population.

First of all, we have proposed model (1) to describe the vector population dynamic (*Aedes Albopictus*) which take into account auto-regulation phenomenon of eggs and larvae stages. We have shown that this model is well defined. For this model we found that,

$$r = \left(\frac{b}{s+d}\right) \left(\frac{s}{s_L+d_L}\right) \left(\frac{s_L}{d_A}\right)$$

is the threshold condition for the existence of the endemic state, where $\left(\frac{b}{s+d}\right)$, $\left(\frac{s}{s_L+d_L}\right)$ and $\left(\frac{s_L}{d_A}\right)$ are respectively eggs, larvae and adult growth rates. For $r > 1$, we proved using Lyapunov function that the endemic equilibrium is globally asymptotically stable.

Then we have proposed model (3) to describe the virus transmission to human population. We have found, in the case $r > 1$ which is the biological interesting case, the following threshold parameter,

$$R_0 = \frac{\beta_m \beta_H}{d_A(\gamma + b_H)} = \frac{\beta_m \beta_H}{S_L \frac{L^*}{A^*}(\gamma + b_H)}$$

which is a very important result for epidemiologists seeking to control a disease via the control of the vector population.

For $R_0 > 1$, there is a unique endemic equilibrium which is globally asymptotically stable. One of the next steps to have a more accurate model would be to take into account the effect of the control of the mosquito population. This could be done, for instance, destroying the breeding sites or introducing a sterile male mosquito population.

REFERENCES

- [1] W. A. Hawley, The biology of *Aedes albopictus*, *J Am Mosq Control Assoc Suppl* 1 (1988) 1–39.
- [2] M. C. Robinson, An epidemic of virus disease in southern province, Tanganyika territory, in 1952–1953, *Transactions of the Royal Society of Tropical Medicine and Hygiene* 49 (1) (1955) 28 – 32. doi:DOI: 10.1016/0035-9203(55)90080-8.
- [3] L. Esteva, C. Vargas, Analysis of a dengue disease transmission model., *Math Biosci* 150 (2) (1998) 131–151.
- [4] M. Derouich, A. Boutayeb, Dengue fever: Mathematical modelling and computer simulation, *Applied Mathematics and Computation* 177 (2) (2006) 528 – 544.
- [5] Y. Dumont, F. Chiroleu, C. Domerg, On a temporal model for the chikungunya disease: Modeling, theory and numerics, *Mathematical Biosciences* 213 (1) (2008) 80 – 91. doi:DOI: 10.1016/j.mbs.2008.02.008.
- [6] D. Moulay, M. Aziz Alaoui, M. Cadivel, A model for the chikungunya virus, *Journal of theoretical biology*, submitted.
- [7] H. Thieme, Convergence results and a Poincaré-Bendixson trichotomy for asymptotically autonomous differential equations, *Journal of mathematical biology* 30 (7) (1992) 755–763.
- [8] J. Zhang, Z. Ma, Global dynamics of an SEIR epidemic model with saturating contact rate, *Mathematical Biosciences* 185 (1) (2003) 15 – 32. doi:DOI: 10.1016/S0025-5564(03)00087-7.
- [9] H. L. Smith, Systems of ordinary differential equations which generate an order preserving flow. a survey of results, *SIAM Rev.* 30 (1) (1988) 87–113. doi:http://dx.doi.org/10.1137/1030003.
- [10] H. L. Smith, *Monotone Dynamical Systems : An introduction to the theory of competitive and cooperative systems*, American Mathematical Society, 1995.
- [11] M. Y. Li, J. S. Muldowney, Global stability for the SEIR model in epidemiology, *Mathematical Biosciences* 125 (2) (1995) 155 – 164. doi:DOI: 10.1016/0025-5564(95)92756-5.

Synchronization in inhibitory networks of bursting neurons

Igor BELYKH, Sajiya JALIL, Andrey SHILNIKOV

Signal Processing

Contents

A novel data pre-processing method on automatic determinig of sleep stages: k-means clustering based feature weighting Salih GÜNEŞ, Kemal POLAT, Şebnem YUSUNKAYA, Mehmet DURSUN	112
Spectral kurtosis and wavelets' entropy for transients' enhancement: application to termite detection Juan José GONZÁLEZ DE LA ROSA, Antonio MORENO-MUNOZ, Antolino GALLEGO, Rosa PIOTRKOWSKI, Enrique CASTRO	118
Modeling and simulation about rarefaction wave gun interior ballistics appling inertial breechblock Zhi JIANZHUANG, Zheng JIAN, Di CHANGCHUN	124
The effect of AR model degrees to early detection of rheumatoid arthritis disease Ali Osman ÖZKAN, Sadik KARA, Ali SALLI, Mehmet Emin SAKARYA, Serter GÜMÜŞ, Ayse ÜNAL, Salih GÜNEŞ	125

A Novel Data Pre-Processing Method on Automatic Determining of Sleep Stages: K-Means Clustering Based Feature Weighting

S. Güneş, K. Polat, Ş. Yosunkaya and M. Dursun

Abstract— Sleep scoring is a time consuming and difficult task conducted by sleep specialist. The main aims of this study are to propose a novel data pre-processing method called K-means clustering based feature weighting (KMCFW) and to develop an automatic recognition system for determining of sleep stages. In this paper, we have used a three stage hybrid system comprising: (i) feature extraction using Welch method from PSG (Polysomnogram) signals including EEG (Electroencephalogram) and chin EMG (Electromyogram), (ii) data pre-processing using KMCFW, and (iii) sleep stages classifying using C4.5 decision tree classifier. There are five sleep stages: Awake, REM (Rapid Eye Movement), N-REM (Non-Rapid Eye Movement) stage 1, N-REM stage 2, and N-REM stage 3. In order to determine the sleep stages, three all night PSG recordings were used for this study. Using alone spectral features belonging to EEG signal; decision tree has obtained classification accuracy of 37.84% on classification of sleep stages using ten fold cross validation. Sleep stages have been classified with accuracy of 41.85% using decision tree based on spectral features belonging to EEG and chin EMG signals. In weighted spectral features belonging to EEG signal with KMCFW, the classification accuracy of 92.40% has been achieved on sleep stages classification using decision tree. As for weighted spectral features belonging to EEG and chin EMG signals with KMCFW, sleep stages has been determined with accuracy of 93.39% using decision tree. These results have demonstrated that the proposed weighting method have a considerable impact on determining of sleep stages.

Index Terms— K-means clustering based feature weighting; Sleep staging; Decision tree; Polysomnography; EEG; EMG

I. INTRODUCTION

Sleep is a losing state as temporary, partial, and periodic in the form of that can be returned with various forced stimulus of the communication of organism with environment.

Manuscript received April 6, 2009. This work is supported fully by the TUBITAK Research Project under Grant No. 108E033. And also, this study is supported by the Scientific Research Projects of Selcuk University.

S. Güneş is with the Selcuk University, Electrical and Electronics Eng., 42035, Konya, Turkey (corresponding author to provide phone: +90 332 2232094; fax: +90 332 2410635; e-mail: sgunes@selcuk.edu.tr).

K. Polat is with the Selcuk University, Electrical and Electronics Eng., 42035, Konya, Turkey (e-mail: kpolat@selcuk.edu.tr).

Ş. Yosunkaya is with Selcuk University, Faculty of Medicine, Sleep Laboratory, Konya, Turkey (e-mail: syosunkaya@selcuk.edu.tr).

M. Dursun is with the Selcuk University, Electrical and Electronics Eng., 42035, Konya, Turkey (e-mail: mehmet_dursun@yahoo.com).

Sleep can be also defined the decreasing of motor activity, the decreasing of response with stimulus, and to be easy recycling as behavioral [1, 2].

Sleep scoring was made according to human sleep standard terminology and handbook prepared by Rechtschaffen and Kales (RK) in 1968 [2]. Scoring of sleep stages was done on the basis of RK standard (RKS) until recent dates and sleep stages in normal subject was divided into 5 stages including awake, N-REM 1 (non- Rapid Eye Movement-1), N-REM 2, N-REM 3, N-REM 4, and REM (Rapid Eye Movement). American Academy of Sleep Medicine (AASM) determined new rules in the scoring of sleep on the chairmanship of Dr. Iber Conrad. Nowadays, sleep staging is done according to these new rules. Sleep stages basically consists of two stages including N-REM (stage I, stage II, and stage III) and REM stages [3, 4, 5].

The process of sleep scoring consists of three steps as follows:

- a. There need epochs with 30 seconds to score the sleep stages.
- b. Each epoch is named with a sleep stage.
- c. If two stages take place in same epoch, it is called as that stage what stage is more than half of the epoch.

In literature, there are many works regarding to sleep stage scoring. The used systems are generally on the basis of extracting features obtained from EEG, EMG, and EOG and on classifying them into one of the sleep stages, while trying to obtain similar results as the experts of visual scoring [6]. Few studies among these studies have been explained as follows.

In study of Este'vez et al., an automated sleep scoring system has been demonstrated. Five patterns have been searched for: slow delta and theta wave predominance in the background EEG activity, presence of sleep spindles in the EEG, presence of rapid eye movements in an electro-oculogram, and the muscle tone in an electromyogram. Results on a test set have shown an overall accuracy of 87.7% between the automated system and the human expert [7].

Šušmáková et al. has researched the basic knowledge about classification of sleep stages from polysomnographic recordings. And also, they have reviewed and compared a large number of measures to find the suitable candidates for the study of sleep onset and sleep evolution. They obtained classification error of 23% on the most difficult decision problem, between

S1 and REM sleep by measures computed from electromyogram led by fractal exponent [6].

Zoubek et al. has focused on the problem of selecting relevant features extracted from human polysomnographic (PSG) signals to perform accurate sleep/wake stages classification. While they achieved an agreement of 71% with the whole database classification of two human experts using a simple set of features such as relative EEG powers in five frequency bands, 80% of agreement with the expert classification obtained using features extracted from the EEG, EOG and EMG signals [8].

In this paper, we have proposed a novel data pre-processing called k-means clustering based feature weighting and combined with C4.5 decision tree to classify the sleep stages including Awake, REM (Rapid Eye Movement), N-REM (Non-Rapid Eye Movement) stage 1, N-REM stage 2, and N-REM stage 3. Decision tree has obtained classification accuracy of 37.84% on classification of sleep stages with ten fold cross validation using frequency domain features belonging to EEG signal. Sleep stages have been classified with accuracy of 41.85% using decision tree based on frequency domain features belonging to EEG and chin EMG signals. In weighted frequency domain features belonging to EEG signal with KMCFW, the classification accuracy of 92.40% has been achieved on sleep stages classification using decision tree.

II. MATERIAL

A. Data

All night polysomnographic records were made by using VIASY trademark PSG device on the sleep laboratory of Meram Medicine Faculty of Selcuk University. Polysomnography device is a device that recorded electrophysical signals such as electroencephalograph (EEG), electrooculograms (EOG), Electromyography (EMG) etc. In this study, we studied on three male subjects and their ages are 56, 31, and 40, respectively. In automatic scoring of sleep stages, their PSG recordings including EEG and chin EMG were used. The average recording time was 6.22 h and total recording time was 18.67 h. Signals and their sampling frequencies that are the PSG recordings used by us are as follows:

- a) EEG (Channel: C4A1), sampling frequency: 128 Hz
- b) Chin EMG, sampling frequency: 256 Hz

Figure 1 presents the EEG and chin EMG with 30 seconds obtained from PSG device recorded on a subject. The sleep stages are divided into five stages including Awake, REM (Rapid Eye Movement), N-REM (Non-Rapid Eye Movement) stage 1, N-REM stage 2, and N-REM stage 3. The distribution of sleep epochs belonging to three subjects is shown in Table 1. And also, this table represents the reduced and full sleep stages as epoch. In the full sleep stages dataset, there are 2241 epochs

while there are 1000 epochs in the reduced sleep stages dataset. The causes of this reducing process are to balance the distribution of dataset and to prevent the over fitting in the learning and testing of sleep stages dataset. Thanks to this process, false classification results can be prevented.

TABLE I
THE DISTRIBUTION OF SLEEP
STAGES ON FULL DATASET AND REDUCED DATASET

Sleep Stages	Awake	N-REM-1	N-REM-2	N-REM-3	REM	Total
Full Dataset	287	119	1168	141	526	2241
Reduced Dataset	187	119	277	141	276	1000

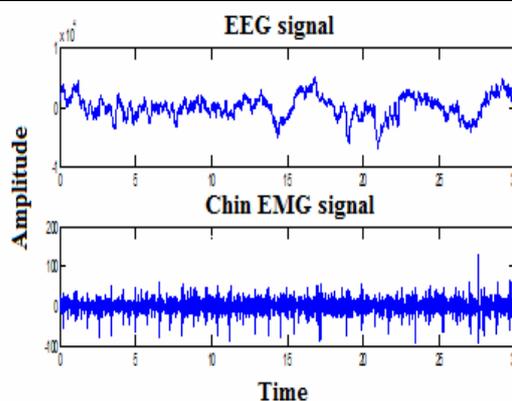


Fig. 1. Raw EEG and chin EMG signals with 30 seconds obtained from PSG device recorded on a subject

III. METHOD

In this paper, we have proposed a hybrid intelligent with three phases based on feature extraction, feature weighting, and classifier algorithm. Figure 2 shows the flowchart of proposed method.

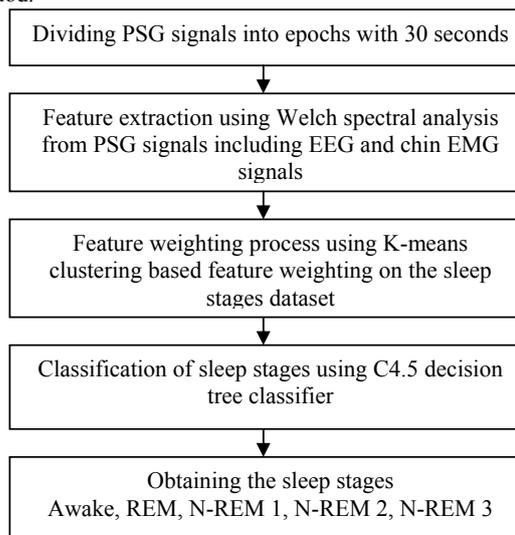


Fig. 2 The flowchart of proposed method

In scoring of sleep stages, PSG signals including EEG and chin EMG are divided into epochs with 30 seconds and labeled each epoch by sleep physician. Later, Welch FFT (Fast Fourier Transform) spectral analysis has been applied to each epoch comprising PSG signals and extracted salient features from frequency domain about sleep stages by means of Welch method. After feature extraction process, K-means clustering based feature weighting has been applied to sleep stages dataset with extracted features. Finally, sleep stages have been automatically scored using C4.5 decision tree classifier.

A. Welch Spectral Analysis Method

We have used Welch spectral analysis method to transform PSG signals from time domain to frequency domain. Welch method is explained as classical method based on FFT. Welch method is the second modification of periodogram spectral estimator, which is to window data segments before computing the periodogram [9-15]. If ready for use information on the signal composes of the samples $\{x(n)\}_{n=1}^N$, the periodogram spectral estimator is provided as follows:

$$\hat{P}_{PER}(f) = \frac{1}{N} \left| \sum_{n=1}^N x(n) \exp(-j2\pi fn) \right|^2 \quad (1)$$

where $\hat{P}_{PER}(f)$ is the estimation of periodogram. In Welch method, signals are divided into overlapping segments, each data segment is windowed, periodograms are calculated and then average of periodograms is found. $\{x_l(n)\}$, $l=1, \dots, S$ are data segments and each segment's length equals M . The overlap ratio is frequently chosen as 50% ($M/2$) [15]. The Welch spectrum estimate is given by:

$$\hat{P}_w(f) = \frac{1}{S} \sum_{l=1}^S \hat{P}_l(f) \quad (2)$$

$$\hat{P}_l(f) = \frac{1}{M} \frac{1}{P} \left| \sum_{n=1}^M v(n) x_l(n) \exp(-j2\pi fn) \right|^2$$

where $\hat{P}_l(f)$ is the periodogram estimate of l^{TH} segment, $v(n)$ is the data-window, M is window sequence. P is total average of $|v(n)|^2$ and given as $P = 1/M \sum_{n=1}^M |v(n)|^2$, $\hat{P}_w(f)$ is the Welch PSD estimate, M is the length of each signal segment and S is the number of segments [15].

Later, evaluation of $\hat{P}_w(f)$ at the frequency samples fundamentally demands the computation of the following discrete Fourier transform (DFT):

The FFT algorithm can calculate the Welch PSD. Variance of an estimator is one of the measures often used to characterize its performance. For 50% overlap and triangular window, variance for the Welch method is provided by [15];

$$\text{var}(\hat{P}_w(f)) = \frac{9}{8S} \text{var}(\hat{P}_l(f)) \quad (3)$$

where $\hat{P}_w(f)$ the Welch PSD is estimate and $\hat{P}_l(f)$ is the periodogram estimate of each signal interval [9-15].

In feature extraction from EEG and chin EMG signals, 129 data segments (windows) for each signal have been used and obtained a sleep stage dataset comprising 258 features in the end of Welch method for EEG and chin EMG signals.

B. Feature Reduction

In order to reduce the dimension of sleep stage dataset with 258 features, the statistical measures have been used. The used statistical features are minimum value, maximum value, standard deviation, and mean value belonging to each feature in sleep stage dataset. The dimension of sleep stage dataset is decreased from 258 to 8 features for EEG and chin EMG signals. The used statistical features are as follows:

TABLE II
THE USED STATISTICAL FEATURES

1.	Minimum value: $x_p = \min x(n) $
2.	Maximum value: $x_p = \max x(n) $
3.	Standard deviation:
	$X_{std} = \sqrt{\frac{\sum_{n=1}^N (x(n) - x_m)^2}{N-1}}$
4.	Mean value: $X_m = \sum_{n=1}^N x(n)/N$

where, $x(n)$, $n=1,2, \dots, N$ is a time series. N is the number of data points.

C. K-means Clustering Based Feature Weighting (KMCFW)

Clustering algorithms are used widely not only to collect similar or dissimilar data, but also useful for data compression and data reduction. The most used clustering algorithms are K-means clustering [16], fuzzy C-means clustering [17], the mountain clustering [18], and subtractive clustering [19]. In this paper, we have chosen K-means clustering as weighting process since this method is widely used in literature.

In K-means clustering based feature weighting method, at first the clusters of each feature are found using K-means clustering (KMC) and calculated the distance between its cluster and mean value of that feature. According to calculated distance, features are weighted.

The goal of feature weighting method is to map the features according to their distributions in a dataset and also transform from non-linearly separable dataset to linearly separable dataset. Feature weighting method works based upon principle that decreasing the variance in features forming dataset. Thanks to this weighting method, the similar data in same feature are gathered and the discrimination ability of classifier is increased.

In this study, a new weighting method (KMCFW) is proposed. The K-means clustering is briefly explained and then explained the proposed weighting method.

K-means clustering also known as C-means clustering has been applied to a variety of areas including image segmentation, speech data compression, data mining etc [20]. The working of KMC can be summarized as follows [21]:

Phase 1: Choose K initial cluster centers z_1, z_2, \dots, z_K randomly from the n points $\{X_1, X_2, X_3, \dots, X_n\}$.

Phase 2: Assign point $X_i, i=1,2,\dots,n$ to the cluster $C_j, j \in \{1,2,\dots,K\}$

if $\|X_i - z_j\| < \|X_i - z_p\|, p=1,2,\dots,K$ and $j \neq p$

Phase 3: Compute new cluster centers as follows

$$z_i^{new} = \frac{1}{n_i} \sum_{X_j \in C_i} X_j \quad i=1,2,\dots,K$$

where n_i is the number of elements belonging to the cluster C_i .

Phase 4: If $\|z_i^{new} - z_i\| < \epsilon, i=1,2,\dots,K$, then terminate. Otherwise continue from phase 2.

This weighting method works as follows: firstly the cluster centers are calculated using KMC method. After computing the centers of features, the ratios of means of features to their centers are calculated and these ratios are multiplied with data point of each feature. Figure 3 demonstrates the flowchart of KMCFW method.

D. C4.5 Decision Tree Classifier

Decision trees have been successfully used in solving problems related to machine learning and classifier systems. A decision tree is induced from sample training dataset and each sample is composed of feature values and class label. Decision trees are general classification method based on inductive inference. They can work with noisy data and missing data in dataset. Decision trees search in the hypothesis space that is fully explained. In constructing of decision tree, small trees are generally preferred to big trees [22, 23].

Each node in decision tree provides testing features belong to training set and each branch created from this node is suitable for a value of feature [23].

Decision trees are considered as a junction of disjunctions. C4.5 decision tree learning is a method for discrete-valued functions classifying, where a C4.5 decision tree depicts the learned function. Learned trees can be shown as sets of if-then rules. These learning methods are among the most popular of inductive inference algorithms and have been successfully applied to a broad range of tasks. C4.5 decision tree is a discovering method, hill climbing, not going backwards search

through the space of all available C4.5 decision trees. The objective of C4.5 Decision tree learning is to partition recursively data into subgroups. For more information on C4.5 decision tree learning, the readers can refer to [22, 23, 24].

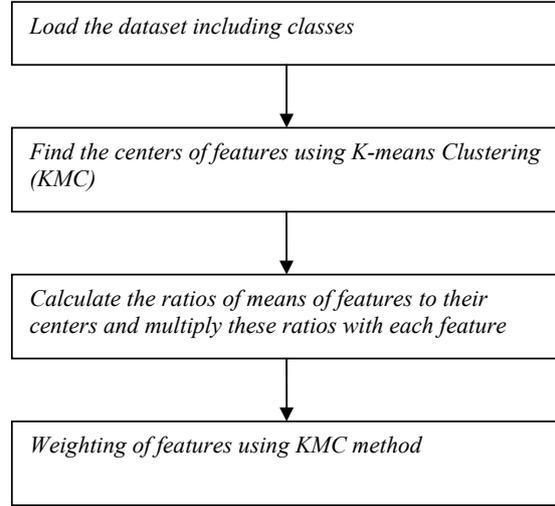


Fig. 3. The flowchart of KMCFW method

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, a novel feature weighting method based on K-means clustering was proposed and applied to automatic determining of sleep stages including awake, REM, N-REM 1, N-REM 2, and N-REM 3. Sleep scoring is a difficult and time consuming task for sleep clinicians. Therefore, the sleep staging process is a challenging work. First of all, the EEG and chin EMG signals were transformed from time domain to frequency domain using Welch method. The salient features were obtained about sleep stages by means of Welch method. In the second step, K-means clustering based feature weighting was applied to sleep stages dataset with extracted features. Finally, sleep stages have been automatically scored by using C4.5 decision tree classifier. Table 3 presents the results of only C4.5 decision tree and combination of C4.5 decision tree and KMCFW method on the determining of sleep stages. In training and testing of C4.5 decision tree classifier, 10 fold cross validation method has been used. And also, we have used the classification accuracy to evaluate the performance of proposed weighting method on the automatic determining of sleep stages.

Figure 4 presents the data distribution of sleep stages dataset including raw EEG frequency domain without KMCFW according to first three features (1st, 2nd, and 3rd). Figure 5 demonstrates the data distribution of sleep stages dataset including EEG frequency domain weighted by KMCFW according to first three features (1st, 2nd, and 3rd).

TABLE III
THE RESULTS OF ONLY C4.5 DECISION TREE AND COMBINATION OF C4.5
DECISION TREE AND KMCFW METHOD ON THE DETERMINING OF SLEEP STAGES

Method	Used signals	Number of Features	Classification Accuracy (%)
Decision tree classifier	EEG	4	37.84
Decision tree classifier	EEG and chin EMG	8	41.85
Combination of k-means clustering based feature weighting and decision tree classifier	EEG	4	92.40
Combination of k-means clustering based feature weighting and decision tree classifier	EEG and chin EMG	8	93.39

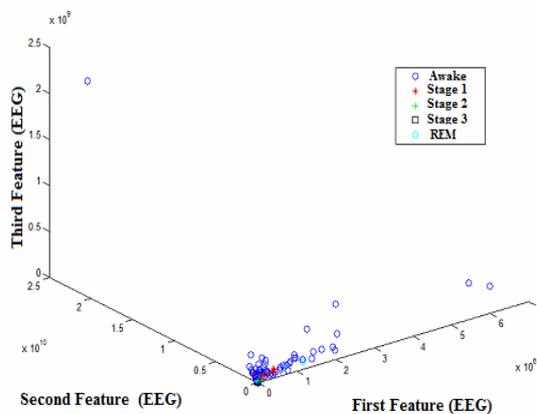


Fig. 4. The data distribution of sleep stages dataset including raw EEG frequency domain without KMCFW according to first three features (1st, 2nd, and 3rd)

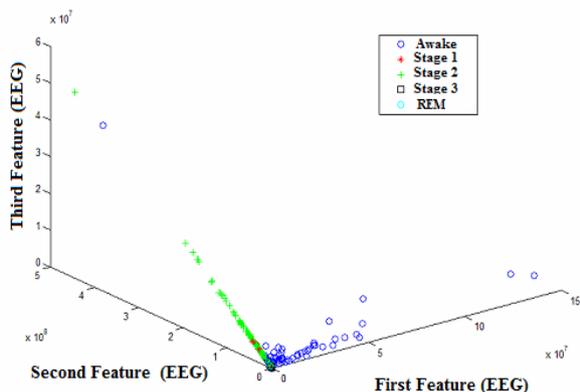


Fig. 5. The data distribution of sleep stages dataset including EEG frequency domain weighted by KMCFW according to first three features.

As can be seen from these figures, the linearity of sleep stages dataset was increased by means of KMCFW method. Thanks to KMCFW method, the distribution of non-linearly separable sleep stage dataset was transformed to a linear

separable distribution. These results have demonstrated that the proposed weighting method called KMCFW is an effective and superior method on determining of sleep stages.

V. CONCLUSION

In this paper, a novel feature weighting called K-means clustering based feature weighting was proposed for automatic scoring of sleep stages including awake, REM, N-REM 1, N-REM 2, and N-REM 3 stages. While decision tree has obtained worse results on automatic scoring of sleep stages without KMCFW, the combination of decision tree and KMCFW has achieved better results than only decision tree classifier. The advantage of this weighing method is that a non-linearly separable dataset can be transformed to a linearly separable dataset and in this way; the classification ability of classifier algorithms can be increased. The proposed method could be confidently used in automatic scoring of sleep stages.

ACKNOWLEDGMENT

This work is supported fully by the TUBITAK Research Project under Grant No. 108E033. And also, this study is supported by the Scientific Research Projects of Selcuk University.

REFERENCES

- [1] Polat, K., Yosunkaya Ş., Güneş S., "Comparison of Different Classifier Algorithms on the Automated Detection of Obstructive Sleep Apnea Syndrome", *Journal of Medical Systems*, 32(3) 2008, 243-250.
- [2] Rechtschaffen A, Kales A, editors. "A manual of standardized terminology, techniques and scoring system for sleep stages of human subject", Washington DC: US Government Printing Office, National Institute of Health Publication; 1968.
- [3] American Academy of Sleep Medicine Task Force. "Sleep related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research", *Sleep* 1999;22: 667—89.
- [4] AASM. "Sleep-Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research" *The Report of an American Academy of Sleep Medicine Task Force, SLEEP*, Vol. 22, No. 5, 1999.
- [5] Dursun, M., "Uyku evreleri", Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Semineri, 2008.
- [6] Šušmáková, K., Krakovská, A., "Discrimination ability of individual measures used in sleep stages classification", *Artificial Intelligence in Medicine*, 44(3), 2008, 261-277.
- [7] Este'vez PA, Held C, Holzmann C, Pe'rez C, Pe'rez JP, Heiss J, et al. Polysomnographic pattern recognition for automated classification of sleep-waking states in infants. *Medical and Biological Engineering and Computing* 2002;40(1): 105—13.
- [8] Zoubek, L., Charbonnier, S., Lesecq, S., Buguet, A., Chapotot, F., "Feature selection for sleep/wake stages classification using data driven methods", *Biomedical Signal Processing and Control*, 2(3), 2007, 171-179.
- [9] D. Evans, *Doppler Signal Analysis*, Ultrasound in Med. & Biol., Vol. 26, Supplement 1, pp. S13—S15, 2000
- [10] V. D. Saini, N. C. Nanda & D. Maulik, 'Basic Principal of ultrasound and Doppler effect', *Doppler echocardiography*, 1993.
- [11] B. Sigel, 'A brief history of Doppler ultrasound in the diagnosis of peripheral vascular disease', *Ultrasound Med. Biol.*, 1998 24:169—176.
- [12] M. Müller, P. Ciccotti, W. Reiche, T. Hagen, 'Comparison of color flow Doppler scanning, power Doppler scanning, and frequency shift for assessment of carotid artery stenosis', *J. Vasc. Surg.*, 2001 34: 1090—1095.

- [13] D.H. Evans, W.N. McDicken, R. Skidmore, J.P. Woodcock, 'Doppler Ultrasound: Physics', *Instrumentation and Clinical Applications*, 1989.
- [14] P.J. Vaitkus, R.S.C. Cobbold, K.W. Johnston, 'A comparative study and assessment of Doppler ultrasound spectral estimation techniques part II: methods and results', *Ultrasound Med. Biol.*, 1988 **14**: 673–688.
- [15] Latifoğlu, F., Polat, K., Kara, S., Güneş, S., " Medical diagnosis of atherosclerosis from Carotid Artery Doppler Signals using principal component analysis (PCA), k-NN based weighting pre-processing and Artificial Immune Recognition System (AIRS)", *Journal of Biomedical Informatics*, 41(1) 2008, 15-23.
- [16] B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1967, 1:281-297.
- [17] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, 1981, New York
- [18] Yager, R. R., Filev, D. P., Generation of fuzzy rules by mountain clustering. *IEEE Transactions on Systems, Man and Cybernetics*, (1994), 24, 209–219.
- [19] Chiu, S. L., Fuzzy model identification based on cluster estimation, *Journal of Intelligent and Fuzzy Systems*, (1994), 2.
- [20] Rui Xu and Donald Wunsch II, "Survey of Clustering Algorithms", *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 16(3) 2005, 645-678.
- [21] Guldemir, H., Sengur, A., "Comparison of clustering algorithms for analog modulation classification", *Expert Systems with Applications*, 30(4), 2006, 642-649.
- [22] Quinlan, J. R., *Induction of C4.5 Decision trees*, *Machine Learning*, 1986, 1, 81- 106.
- [23] Mitchell, M. T., *Machine Learning*, McGraw-Hill, Singapore, 1997.
- [24] Polat, K., Güneş, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, *Applied Mathematics and Computation*, 187(2) 2007, 1017-1026.

Spectral Kurtosis and Wavelets' Entropy for Transients' Enhancement: Application to Termite Detection

Juan José González de la Rosa, Member, *IEEE*, A. Moreno, Member, *IEEE*,
A. Gallego, R. Piotrkowski and E. Castro

Abstract—In this paper we present the operation results of a portable computer-based measurement equipment conceived to perform non-destructive testing of suspicious termite infestations. Its signal processing module is based in the Spectral Kurtosis (SK) and it can be applied to numerous kinds of transients' detection. We introduce the de-noising complement of the discrete wavelet transform (DWT) and we study how can it serves to reduce subjectiveness. The SK pattern allows the targeting of alarms and activity signals. The DWT complements the SK, by keeping the successive approximations of the termite emissions, supposed more non-Gaussian (less noisy) and with less entropy than the detail approximations. For a given mother wavelet, the maximum acceptable level, in the wavelet decomposition tree, which preserves the insects' emissions features, depends on the comparative evolution of the approximations details' entropies, and the value of the global spectral kurtosis associated to the approximation of the separated signals. The paper explains the detection criterion by showing different types of real-life recordings (alarms, activity, and background).

Index Terms—Acoustic Emission, Higher-Order Statistics, Insect detection, Spectral Kurtosis.

I. INTRODUCTION

BIOLOGICAL transients gather all the natural complexity of their associated sources, and the media through which they propagate. As a consequence, finding the most adequate method to get a complete characterization of the emission implies the selection of the appropriate model, which better explains the processes of generation, propagation and capture of the emitted signals. This description matches the issue of measurement termite activity.

This paper deals with the improved equipment whose previous prototype's performance, based in the time-frequency domain analysis of the kurtosis, was described in [1].

In this final version, the measurement method is mainly based in the interpretation of the spectral kurtosis graph, along with the wavelet analysis, which is thought as an aid. At the same time, we use a simple data acquisition unit, the sound card (maximum sampling speed at 44,100 Hz), which simplifies the hardware unit and the criterion of detection.

J.J.G. de la Rosa is with the University of Cádiz. Electronics-Research Unit PAIDI-TIC-168. EPSA-Av. Ramón Puyol S/N. E-11202-Algeciras-Spain. Tel-Fax: +34-956028020. Email: juanjose.delarosa@uca.es

A. Moreno is with the University of Córdoba. Research Unit PAIDI-TIC-168. Email: amoreno@uco.es

University of Granada-ESII. Research unit PAIDI-TEP-232 E-18071-Granada-Spain. Email: antolino@ugr.es

The instruments for plague detection are thought with the objective of decreasing subjectiveness of the field operator. On-site monitoring implies capturing the natural phenomenon of insect emissions with high accuracy. As a consequence it is imperative the use of a deep storage device, and high sensitive probes with a selective frequency response. These features make the price paid high, and do not guarantee the success of the detection.

Regarding the procedures, the prior detection methods are very much dependent on the detection of the excess of power in the signals; these are the so-called second-order methods. For example, the RMS calculation can only characterize the intensity, and does not provide information regarding the envelope of the signal nor the amplitude fluctuations. Another handicap of the second-order principle, e.g. the power spectrum, attends to the preservation of the energy during data processing. Consequently, the eradication of additive noise lies in filter design and sub-band decomposition, like wavelets and wavelet packets.

As an alternative to improve noise rejection and complete characterization of the signals, in the past ten years, a myriad of higher-order methods are being applied in different fields of Science and Technology, in scenarios involving signal separation and characterization of non-Gaussian measurements. Concretely, the area of diagnostics-monitoring of rotating machines is also under our interest due to the similarities of the signals to be monitored with the transients from termites. Many time-series of faulty rotating machines consist of more-or-less repetitive short transients of random amplitudes and random occurrences of the impulses.

This paper describes a method based in the SK (related to the fourth-order cumulant at zero lags) to detect infestations of subterranean termites in a real-life scenario (southern Spain). Wavelet decomposition is used as an extra tool to aid detection from the preservation of the approximation of the signal, which is thought to be more Gaussian than the details.

The interpretation of the results is focused on the peakedness of the statistical probability distribution associated to each frequency component of the signal, to get a measure of the distance from the Gaussian distribution. The SK serves as a twofold tool. First, it enhances non-Gaussian signals over the background. Secondly, it offers a more complete characterization of the transients emitted by the insects, providing the user with the probability associated to each frequency component.

The paper is structured as follows: in the following Section

If a review on termite detection and relevant HOS experiences sets the foundations. Then we make a brief report on the definition of kurtosis in Section III; we use an unbiased estimator of the SK, successfully used in [1]. Results are presented thereafter, and finally, conclusions are drawn.

II. SUBTERRANEAN TERMITES: DETECTION PROJECT TOWARDS HOS

Termite detection has been gaining importance within the research community in the last two decades, mainly due to the urgency of avoiding the use of harming termiticides, and to the joint use of new emerging techniques of detection and hormonal treatments, with the aim of performing an early treatment of the infestation. A localized partial infestation can be exterminated after two or three generations of the colony's members with the aid of these hormones, which stop chitin synthesis.

The primary method of termite detection consists of looking for evidence of activity. But only about 25 percent of the building structure is accessible, and the conclusions depend very much on the level of expertise and the criteria of the inspector [2], [1]. As a consequence, new techniques have been developed to remove subjectiveness and gain accessibility.

User-friendly equipment is being currently used in targeting subterranean infestations by means of temporal analysis of the vibratory data sequences. An acoustic-emission (AE) sensor or an accelerometer is fixed to the suspicious structure. This class of instruments is based on the calculation of the RMS value of the vibratory waveform. The RMS value comprises information of the AE raw signal power during each time-interval of measurement (averaging time). This measurement strategy conveys a loss of potentially valuable information both in the time and in the frequency domain [1]. A more sophisticated family of instruments makes use of spectral analysis and digital filtering to detect and characterize vibratory signals [3], [4]. Other second-order tools, like wavelets and wavelet packets (time-dependent technique) concentrate on transients and non-stationary movements, making possible the detection of singularities and sharp transitions, by means of sub-band decomposition.

Higher-order statistics are being widely used in several fields. The spectral kurtosis has been successfully described and applied to the vibratory surveillance and diagnostics of rotating machines [5], [6]. In the field of insect detection, the work published in [1] set the foundations of the present paper. The combined use of the SK and the time-domain sliding kurtosis showed marked features associated to termite emissions. In the frequency domain (sample frequency 64,000 Hz) three frequency zones were identified in the spectral kurtosis graph as evidence of infestation; two in the audio band (which will be also checked in the present paper) and one in the near ultrasound (roughly equal to 22 kHz). In the present paper the sample frequency was fixed to 44,100 Hz and the sound card was directly driven by MATLAB. Results are presented in the user interface, which is forwarded in Fig. 1; in this measurement situation, the time-raw data contains alarms an activity signals from termites. This is a clear example of positive detection.

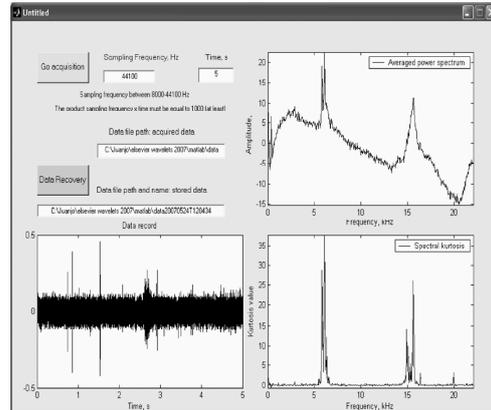


Fig. 1. The graphical user interface which presents the results to the field operator. The spectral kurtosis is in the bottom-right corner.

The developed virtual instrument also calculates and presents the power spectrum (up-right graph) and the raw data (bottom-left). The field operator adds therefore visual information to the classical audio-based criterion, which was by the way very subjective and very expertise-dependent.

III. KURTOSIS, SK AND DE-NOISING STRATEGY VIA WAVELETS

Kurtosis is a measure of the "peakedness" of the probability distribution of a real-valued random variable. Higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. This fact is used in this paper to detect termite emissions in an urban background. Kurtosis is more commonly defined as the fourth central cumulant divided by the square of the variance of the probability distribution, which is the so-called excess kurtosis, according to Eq. (1) [1], [7], [8]:

$$\gamma_{4,x} = E\{x^4(t)\} - 3(\gamma_{2,x})^2 = C_{4,x}(0, 0, 0) \quad (1)$$

Normalized kurtosis is defined as $\gamma_{4,x}/(\gamma_{2,x})^2$. We will use and refer to normalized quantities because they are shift and scale invariant.

Ideally, the spectral kurtosis is a representation of the kurtosis of each frequency component of a process (or data from a measurement instrument x_i). For estimation issues we will consider M realizations of the process; each realization containing N points; i.d. we consider M measurement sweeps, each sweep with N points. The time spacing between points is the sampling period, T_s , of the data acquisition unit.

A biased estimator for the spectral kurtosis for a number M of N -point realizations at the frequency index m , is given by Eq. (2):

$$\hat{G}_{2,X}^{N,M}(m) = \frac{M}{M-1} \left[\frac{(M+1) \sum_{i=1}^M |X_N^i(m)|^4}{\left(\sum_{i=1}^M |X_N^i(m)|^2 \right)^2} - 2 \right]. \quad (2)$$

This estimator is the one we have implemented in the program code in order to perform the data computation and it was also used successfully in [1].

We expect to detect positive peaks in the kurtosis's spectrum, which may be associated to termite emissions, characterized by random-amplitude impulse-like events. This non-Gaussian behavior should be enhanced over the symmetrically distributed electronic noise, introduced in the measurement system. Speech is perhaps also reflected in the SK, but not in the frequencies where termite emissions manifest. Besides, we assume, as a starting point, that non-Gaussian behavior of termite emissions is more acute than in speech. As a consequence, these emissions would be clearly outlined in the kurtosis spectrum. As a final remark, we expect that constant amplitude interferences are clearly differentiated due to their negative peaks in the SK.

To show the ideal performance of the estimator, which has been described in these lines, and also described in [1], we show an example based in synthetics. A mixture of six different signals has been designed. Each mixture is the sum of a constant-amplitude sine of 2 kHz, a constant-amplitude sine at 9 kHz, a Gaussian-distributed-amplitude sine at 5 kHz, a Gaussian-distributed-amplitude sine at 18 kHz, a Gaussian white noise, and a colored Gaussian noise between 12 and 13 kHz. Each mixture (realization or sample register) contains 1324 points. Negative kurtosis is expected for constant-amplitude processes, positive kurtosis should be associated to random-amplitudes and zero kurtosis will characterize both Gaussian-noise processes. A simulation has been made in order to show the influence of the number of sample registers (M) in the averaged results for the SK graph, and to test its performance. Fig. 2 shows a good performance because enough registers have been averaged (M=500).

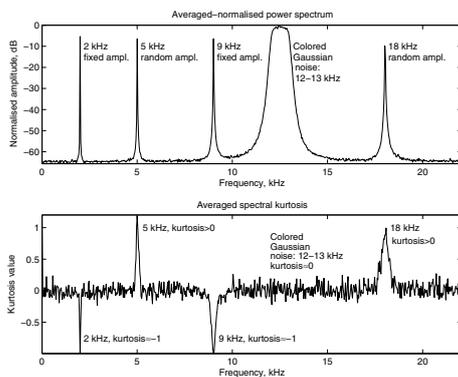


Fig. 2. Performance over a set of synthetics, for M=500 realizations.

The mother wavelet Daubechies 5 has been selected as most similar wavelet mother, because of the highest coefficients in the decomposition tree. Given the wavelet mother, to show the process of selecting the maximum decomposition level in the wavelet tree, we have adopted a criterion based on the calculation of Shannon's entropy (information entropy), which is a measure of the uncertainty associated with a random variable X ; this entropy denoted by $H(X)$, and defined by Eq.

(3):

$$H(X) := - \sum_{i=1}^N p(x_i) \log_{10} p(x_i), \quad (3)$$

where X is an N-outcome measurement process $\{x_i, i = 1, \dots, N\}$, and $p(x_i)$ is the probability density function of the outcome x_i .

We show this strategy via the following example, based on real-life data, which contain activity signals from termites, presented in Fig. 3. The lower sub-figure in Fig. 3 is the result of the de-noising performance at the 4th-decomposition level; using the global thresholding we keep the approximation signal. The entropy of the approximations and the details are compared for each level of comparison and shown in Fig. 4.

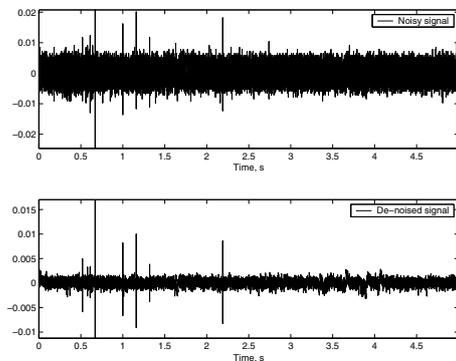


Fig. 3. An example of the de-noising performance with dB5 at level 4 for a sample register.

By looking at the graph of Fig. 4, at levels 3 and 4, the entropy of the approximations is less than the entropy of the details. So level 4 is in a sense, a point of inversion. No improvement is obtained for level 5, where the entropies are very similar.

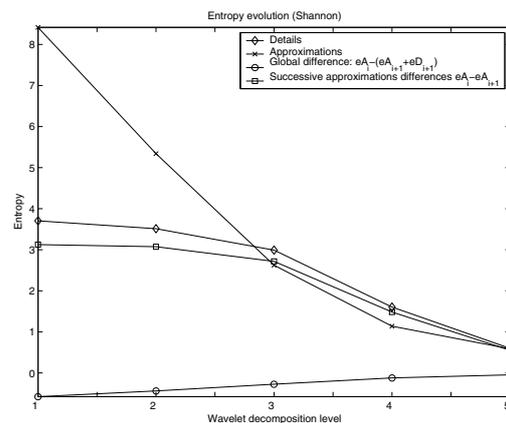


Fig. 4. Evolution of the entropy.

We can also see that the global difference of entropies increases towards zero, at level 5, as a complementary indication

that further decomposition will not suppose progress in denoising.

IV. EXPERIMENTAL RESULTS

A. The instrument and the measurement procedure

A piezoelectric probe-sensor (model SP-1L from *Acoustic Emission Consulting*) is used in the final version of the instrument, and was described in detail in [1]. The sensor is connected to the sound card of a lap-top computer and the acquisition is driven by MATLAB, via the Graphical User Interface (GUI).

The user interface was presented in Fig. 1, in which a clear detection case is shown, and both types of acoustic emissions (alarm and activity - feeding and excavating) are evidenced in the SK graph. The peak near 6 kHz outlines activity and the kurtosis near 16 kHz denotes generic activity - movement, feeding and excavating.

The operator can select the acquisition time and the sample frequency (maximum 44,100 Hz if the sound card is driven). In the bottom-right corner of Fig. 1, the spectral kurtosis graph is presented. The user can also examine the raw data (time-series) and the spectrum. After performing the acquisition, the instrument saves the acquired data, labeling the file with the date and exact hour. Additionally, the operator can recall the stored files; this option allows the operator to present a report to the customer.

The electronic transducer is presented in Fig. 5, along with its charge-to-volt conversion modulus (Integrated Circuit Piezoelectric; ICP interface), and the accessory to fix it in the wood that we used to test the sensor's performance in the lab. A bare waveguide has been used for insertion into soil.



Fig. 5. A photograph of the sensor SP-1L with the mounted accessories (drill bit) prepared to couple it into the a test-piece of wood, in our lab. On the left, the charge-to-volt converter according to the Integrated Circuit Piezoelectric (ICP) protocol.

The transducer SP-1L was used to record the data registers in the field experience, and the ICP unit was connected to the sound card of a lap-top computer, configuring an autonomous measurement unit. The sampling frequency was $F_s=44,100$ Hz for all the registers analyzed in this paper. The recording stage took place in a garden with evidence of infestation and the bare waveguide of the sensor was introduced in the lawn, over the suspicious zone.

Termite sounds from feeding are like sharp pops and crackles in the audio output. Hit rates of 25-100 per second are common in infestation locations¹.

We remember that the main goal of this signal-processing method is to reduce subjectiveness associated with visual or listening inspection of the registers. This means that in a noisy environment, it may be possible to ignore termite feeding activity even with an *ad hoc* sensor because, despite the fact that the sensor is capable of register these low-level emissions, the human ear can easily ignore them because emissions are buried in the speech, urban or environmental noise, or even may be confused by the steps or other sounds being propagated through the soil, in the moment of the measurement. It has been shown that filtering can only be used as a pre-processing tool, and that the real enhancement is developed using higher-order statistics [1].

Termite emissions are non-stationary, so the instrument treats data by ensemble averaging of the sample registers, following the indications in [9] (pp. 463-465). Each spectrum and spectral kurtosis graph presented in this section is the result of averaging the spectra of the sample registers, or realizations. As a final remark, acquired data is normalized according to the norm: $\|s\| = \left(\sum_{i=1}^N |s_i|^2 \right)^{1/2}$.

The virtual instrument does not include the wavelet analysis tool (it is thought to be easy to use by a field operator). Wavelet analysis are introduced here to complement the experimental results, in the sense of achieving further conclusions regarding signal separability and entropy criterion.

B. Operating cases

We present the signals out of the instrument display in order to be analyzed more precisely. A data acquisition time of 5 seconds and a sample frequency of 44,100 Hz were selected. So every time the user performs an acquisition (pressing the button "Go" in the interface) 220,500 points are stored. The software-engine is adjusted to calculate the averaged spectral kurtosis (SK) over a set of 220 realizations, each of them containing 1,000 points.

A number of 30 registers, each of them of 5 seconds length, have been analyzed. Each register is supposed to contain evidences of termite activity, where alarm impulses can appear spontaneously as a consequence of the defense mechanism of the insects. Among them a number of 5 were considered as false-positives.

Two couples of graphics (two measurement cases) have been selected in this paper to summarize the results: a clear detection case, and a doubtful situation. Fig. 6 presents a clear detection case, characterized by termite activity signals with almost non-audible alarms. Two peaks are clearly enhanced in the SK graph (near 6 kHz, and near 16 kHz), which matches the insect-activity frequency pattern associated to the infestation. If a frequency appears below 5 kHz, alarms (head-banging movements [10]) are supposed to be significant in amplitude.

¹Courtesy of Acoustic Emission Consulting, Inc. The definition of "hit" is related to the instrument AED2000, and refers to the times the voltage rms signal pass the threshold.

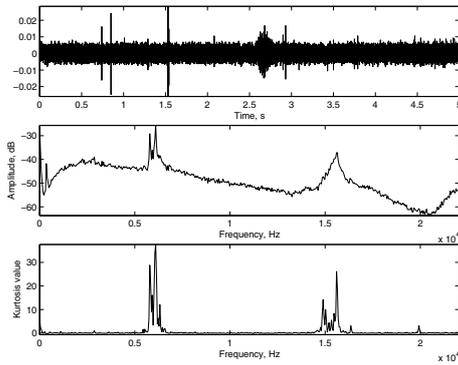


Fig. 6. A clear measurement of activity detection.

The de-noised data in the time domain are shown in the upper graph of Fig. 7. Applying the spectral kurtosis to the de-noised time-series, it is seen in the SK graph that all the frequency components are enhanced, specially those in the detection bands of interest. This fact confirms the presence of insects, and it is of special value in doubtful situations, when they are really needed.

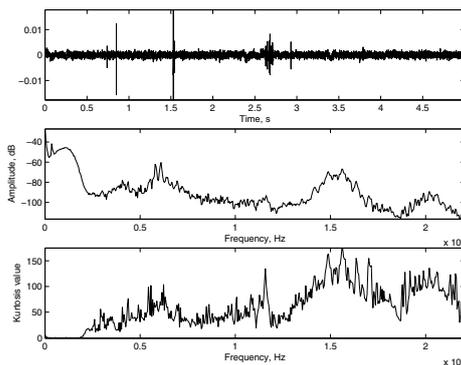


Fig. 7. De-noising results for data in Fig. 6. A general enhancement of the spectral kurtosis occurs.

In Fig. 8 a doubtful measurement case is presented. Activity evidence is outlined only near 6 kHz. Once, the wavelets have been applied (shown in Fig. 9), the enhancement near 6 kHz and 16 kHz confirms the detection.

Finally, in order to compare, we present a final couple of figures to show the measurement performance over background. In the time-domain, the impulses that are similar to *Dirac* deltas are associated to cracks or little movements of the sensor while attached in the ground.

This last couple of graphics clearly shows the difference between a possible detection case and a background pattern. There's no peaks in the SK pattern for both the original time-register and the de-noised register.

V. CONCLUSION

The kurtosis as a global indicator, considered as the average of the kurtosis computed for each individual frequency

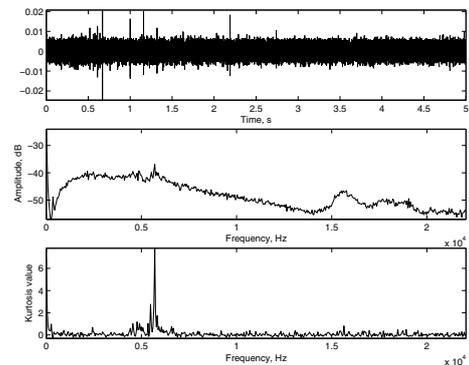


Fig. 8. A doubtful measurement situation. Only one peak appears and confirmation is required.

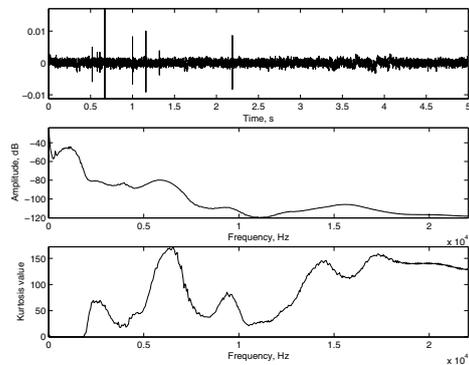


Fig. 9. De-noising results of data in Fig. 8.

component, is not a valid tool to target termite activity. This is due to the fact that no discrimination is made among the frequency bands of the emissions, which may be originated by different agents. A selective procedure is then required.

The key of the spectral kurtosis detection strategy used in this work lies in the potential enhancement of the non-Gaussian behavior of the emissions. If this happens, i.e. if an increase of the non-Gaussian activity (increase in the kurtosis, peakedness of the probability distribution) is observed-measured in concrete bands (near of 6 kHz and 16 kHz) of the spectral kurtosis graph, there may be infestation in the surrounding subterranean perimeter, where the transducer is attached.

Thus, assuming the starting hypothesis that the insect emissions may have a more peaked probability distribution than any other simultaneous source of emission in the measurement perimeter, we have design a termite detection strategy and a virtual instrument based in the calculation of the *4th*-order cumulants for zero time lags, which are indicative of the signals' kurtosis. Its frequency domain representation is the SK. The engine of the instrument is an estimator of the SK, which performs a selective analysis of the peakedness of the signal probability distribution. It has been shown that new frequency components gain in relevance in the spectral SK graphs.

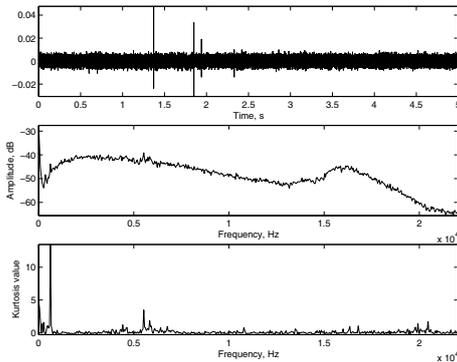


Fig. 10. Spectral kurtosis over a background sample register. No activity is seen in the surroundings of the characteristic frequencies.

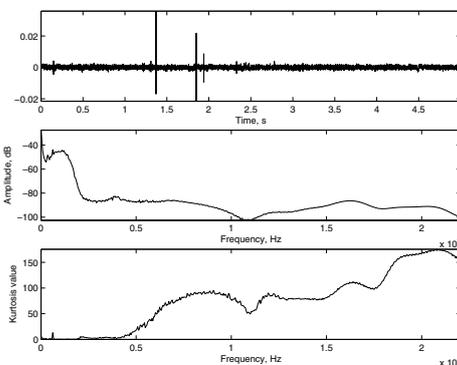


Fig. 11. De-noising of data in Fig. 10. No information is added.

The repeatability of the instrument is high and helps discriminate when false positives appear.

Regarding the benefits, the main goal of this signal-processing method is to reduce subjectiveness due to traditional visual or listening inspection of the registers. This means that in a noisy environment, it may be possible to ignore termite feeding activity even with an *ad hoc* sensor because, despite the fact that the sensor is capable of register these low-level emissions, the human ear can easily ignore them. The commercial version of the instrument is actually been planned to appear by a the Spanish Company Natural Connection and Consulting (*Konectia*).

ACKNOWLEDGEMENT

The authors would like to thank the *Spanish Ministry of Science and Education* for funding the project DPI2003-00878, where the different noise processes have been modeled and contrasted; and also for supporting the PETRI project PTR95-0824-OP dealing with plague detection using higher-order statistics. Our unforgettable thanks to the trust we have from the *Andalusian Government* for funding the excellency project PAI2005-TIC00155, where higher-order statistics are modeled and applied to plague detection and power quality analysis.

REFERENCES

- [1] J. J. G. De la Rosa and A. M. Muñoz, "Higher-order cumulants and spectral kurtosis for early detection of subterranean termites," *Mechanical Systems and Signal Processing* (Ed. Elsevier), vol. 22, no. Issue 1, pp. 279–294, February 2008, available online 1 September 2007.
- [2] W. P. Robbins, R. K. Mueller, T. Schaal, and T. Ebeling, "Characteristics of acoustic emission signals generated by termite activity in wood," in *Proceedings of the IEEE Ultrasonic Symposium*, 1991, pp. 1047–1051.
- [3] R. W. Mankin and J. R. Fisher, "Current and potential uses of acoustic systems for detection of soil insects infestations," in *Proceedings of the Fourth Symposium on Agroacoustic*, 2002, pp. 152–158.
- [4] R. W. Mankin, M. T. Smith, J. M. Tropp, E. B. Atkinson, and D. Y. Yong, "Detection of *Anoplophora glabripennis* (coleoptera: Cerambycidae) larvae in different host trees and tissues by automated analyses of sound-impulse frequency and temporal patterns," *Journal of Economic Entomology*, vol. 101, no. 3, pp. 838–849, 2008, forest Entomology.
- [5] J. Antoni, "The spectral kurtosis: a useful tool for characterising non-stationary signals," *Mechanical Systems and Signal Processing* (Ed. Elsevier), vol. 20, no. 2, pp. 282–307, february 2006.
- [6] —, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing* (Ed. Elsevier), vol. 20, no. 2, pp. 308–331, february 2006.
- [7] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, 1991.
- [8] A. K. Nandi, *Blind Estimation using Higher-Order Statistics*, 1st ed., A. K. Nandi, Ed. Boston: Kluwer Academic Publishers, 1999, vol. 1, no. 1.
- [9] J. Bendat and A. Piersol, *Random Data Analysis and Measurement Procedures*, 3rd ed., ser. Wiley Series in Probability and Statistics, N. Y. Wiley Interscience, Ed. Wiley Interscience, 2000, vol. 1, no. 1.
- [10] A. Röhrig, W. H. Kirchner, and R. H. Leuthold, "Vibrational alarm communication in the african fungus-growing termite genus *macrotermes* (isoptera, termitidae)," *Insectes Sociaux*, vol. 46, pp. 71–77, 1999.

Modeling and Simulation about Rarefaction Wave Gun Interior

Ballistics Applying Inertial Breechblock

ZHI Jianzhuang, ZHENG Jian, DI Changchun

Department of Guns Engineering
Shijiazhuang Mechanical Engineering College
Shijiazhuang, Hebei, PRC, 050003
zhijianzhuang@163.com

Abstract: During worldwide new military revolution, every military great power greatly attaches importance to improve mechanization weapon arm, while contending to study new information-based weapon system. During investigating the armament requirements for the future combat system (FCS), a new species of gun propulsion that dramatically reduces recoil momentum imparted to the gun is presented, first conceived by Eric Lee Kathe on 18 March 1999 as part of the Army After Next (AAN) project. It is termed RAVEN for RAREfaction waVE guN, and may usher in a new era of lightweight weaponry. The rarefaction wave gun (RAVEN) can achieve much low recoil momentum without decreasing the muzzle velocity of the projectile by means of suddenly opening nozzle while projectile traveling special location in barrel. Base on vent mechanics' work property of RAVEN, the paper research the modeling and simulation of RAVEN applying the inertial breechblock, and then compare the simulation result with same type orthodox gun. It is concluded that RAVEN can extremely reduce recoil impulse and the barrel temperature without effect the muzzle velocity under the same loading condition.

Guns remain heavy, despite advances in material technology, for two principal reasons. Their thermal mass is required to manage the heat generated during burst-fire. Also, the inertia of heavy guns aids in recoil –lighter guns are endowed with more recoil energy during firing than heavier guns. Therefore, RAVEN propulsion is an enabling technology to provide future war fighters with lightweight guns that impose less recoil burden and provide improved thermal management. This will allow the war fighter to engage with maximum firepower and to keep firing longer. RAVEN can achieve much low recoil momentum without decreasing the muzzle velocity of the projectile by means of suddenly opening breechblock while projectile traveling special location in barrel.

Two major different breechblock opening mode are discussed respectively, based on the launch mechanism of rarefaction wave gun (RAVEN). RAVEN which apply the redesigned newly inertial breechblock is emphatically modeled and simulated, and then compare the simulation result with same type orthodox gun. It is concluded that RAVEN can extremely reduce recoil impulse and the barrel temperature without effect the muzzle velocity under the same loading condition.

Keywords: Rarefaction wave gun, interior ballistics, inertial breechblock, simulation

References

- [1] Kathe, E. Sonic rarefaction wave recoilless gun system. USA: 09/631, 142, 8 Oct 2000.
- [2] Kathe, E. Inertial breechblock gun system. USA: 09/363,700, 11 Jun 2001.
- [3] Kathe, E. Rarefaction wave gun propulsion. Troy, NY: Rensselaer Polytechnic Institute, 2002.
- [4] TAN Lebin, HOU Baolin, CHEN Weimin. Gun recoil force reduction technology. Journal of gun launch & control, 2006(4), pp. 69-72
- [5] JIN Zhi-ming. Interior ballistics about gun. Beijing: Beijing institute of technology publisher, 2004:288-289.
- [6] Weng Chun-sheng, Wang Hao. Interior ballistics about computation. Beijing: National defense industry publisher, 2006:24-25.

The Effect of AR Model Degrees to Early Detection of Rheumatoid Arthritis Disease

A.O. Özkan, S. Kara, A. Sallı, M. E. Sakarya, S. Gümüş, A. Ünal and S. Güneş

Abstract— In this paper, we have investigated the effect of Burg AR spectral analysis model degrees to early detection of rheumatoid arthritis (RA) disease using Doppler signals belonging to right ulnar arterial obtained from patient and healthy subjects. This study focuses on the diagnosis of rheumatoid arthritis disease via the analysis of Doppler Signals' AR Burg power spectral density graphic with the aid of artificial neural network (ANN) on right ulnar arterial. In the feature extraction from Doppler signals belonging to right ulnar arterial, the AR Burg model degrees of 5, 10, 15, 20, and 25 were used. After feature extraction, multilayer feed forward ANN trained with a Levenberg Marquard (LM) back propagation algorithm was used to classify the Doppler signals belonging to right ulnar arterial whether the patient has RA disease or not (i.e. healthy). In classification of Doppler signals, the obtained classification accuracies from combination of AR Burg spectral analysis and ANN for AR model degrees of 5, 10, 15, 20, and 25 were 90%, 95%, 85%, 95%, and 85%, respectively. The integrations of AR Burg model degrees of 10 and 20 with ANN have achieved better results than other combinations of other AR model degrees with ANN on classification of RA disease using Doppler signals belonging to right ulnar arterial. This system can be useful for physicians to make the final decision without hesitation, on early diagnosis of RA disease.

Index Terms— Rheumatoid Arthritis Disease; Autoregressive spectral analysis; Burg Method; Artificial neural network; Back propagation algorithm

I. INTRODUCTION

Rheumatoid arthritis (RA) is a systemic chronic inflammatory disorder that mainly affects diarthrodial

Manuscript received April 6, 2009. This study is supported by the Scientific Research Projects of Selcuk University.

A. O. Özkan is with the Selcuk University, Technical Sciences Vocational School, Program of Industrial Electronics, 42003, Konya, Turkey (corresponding author to provide phone: +90 332 2232373; fax: +90 332 2410185; e-mail: alozkan@selcuk.edu.tr).

S. Kara is with Fatih University, Department of Electrical - Electronics Eng., 34100, Istanbul, Turkey (e-mail: skara@fatih.edu.tr).

A. Sallı is with the Selcuk University, Meram Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Turkey (e-mail: dralisalli@hotmail.com)

M. E. Sakarya is with the Selcuk University, Meram Faculty of Medicine, Department of Radiology, Konya, Turkey (e-mail: drsakarya@yahoo.com).

S. Gümüş is with the Selcuk University, Meram Faculty of Medicine, Department of Radiology, Konya, Turkey (e-mail: sertergum@yahoo.com).

A. Ünal is with the Selcuk University, Meram Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Konya, Turkey (e-mail: ayseunalenginar@mynet.com)

S. Güneş is with the Selcuk University, Electrical and Electronics Eng., 42035, Konya, Turkey (e-mail: sgunes@selcuk.edu.tr).

joints. It is characterized by inflammatory activity of synovium leading to the destruction of bone and joint cartilage along with periarticular structures like tendons and ligaments. It is the most common form of inflammatory arthritis and the world prevalence of RA might be around 0.3-1.2 %, in a female/ male ratio of 2-5/1. It is most common in patients at 40 - 70 years old and its incidence increases with age [1 – 3].

Disease activity and therapeutic response has predominantly based on clinical assessment and laboratory for serum markers of inflammation like erythrocyte sedimentation rate (ESR) or C reactive protein (CRP). Tender and swollen joint counts are essential for physical examination and evaluating disease activity. These are components of disease activity score 28 (DAS 28) which was developed for evaluating disease activity in RA. However, clinical evaluation of joint pain and swelling has not been sufficiently reliable [4]. Also direct radiography can be used for evaluating established erosions but gives us little information on synovial inflammation and early erosions [5]. However Color Doppler ultrasound (CDU) displays blood flow in the tissues and it can be a marker of the inflammatory response. So the amount of CDU activity in the inflamed synovium may be used to quantify the inflammatory activity in RA [6].

The Doppler effect, which is a resulting from interaction of the ultrasonic wave with moving red blood cells, has been extensively used to determine blood flow velocity [7]. Doppler systems are dependent on the principle that ultrasound, emitted by an ultrasonic transducer, is returned partly towards the transducer by the moving red blood cells, thereby inducing a shift in frequency proportional to the emitted frequency and the velocity along the ultrasound beam [8-10]. The Doppler shift is related to the flow velocity by

$$f_{Doppler} = \frac{2v}{\lambda} \cos\theta \quad (1)$$

where $f_{Doppler}$ is the mean frequency of the Doppler spectrum, v is the flow velocity, λ is the acoustic wavelength and θ is the Doppler angle [7]. Since the scatterers within the ultrasound beam does not usually move at the same speed, a spectrum of Doppler frequencies is investigated. By using spectrum analysis techniques, the variations in the shape of the Doppler spectra as a function of time are presented in the form of sonograms in order to acquire medical information [8-10].

Doppler Ultrasound detects the movement of red blood cells

in the vessels by the analysis of the change in frequency of the returning echoes. This information is converted into sound. Additionally, it is possible to delineate flow curves and to designate the direction of the blood flow [11].

In literature, the comparison of healthy and patient subjects having RA disease over Doppler Ultrasound images on RA disease and the resistive index (RI) and pulsatile index (PI) index comparisons calculated on these images have been generally conducted by researchers in recent years. Therefore, this study is a new study using Doppler ultrasound signals on the early diagnosis of RA disease. There are a few studies related to RA disease in literature.

Among these, Kiriş et al. has examined synovial vascularity and flow patterns in hand and wrist joints- metacarpophalangeal joints and ulnar styloid regions- of patients with RA using power Doppler ultrasonography (PDUS) and spectral Doppler analysis and to assess the accuracy of PDUS in detecting overall disease activity in RA patients [12].

Terslev et al. has evaluated the sensitivity and specificity of Doppler ultrasound in diagnosing arthritis in the wrist and hands, and, if possible, to define a cutoff level for this study ultrasound measures for inflammation, RI, and color fraction [13].

Varsamidis et al. has evaluated the use of spectral Doppler in the assessment of patients with RA. The study group consisted of 32 patients with RA followed up clinically and by ultrasound using the RI measurements in the patients' wrists as indicator of inflammation [14].

In this work, the effect of Burg AR spectral analysis model degrees to early diagnosis of RA disease using Doppler signals belonging to right ulnar arterial achieved from patient and healthy subjects has been searched. The detection of RA disease comprises of two phases: (i) feature extraction using AR Burg power spectral density (PSD) graphic from Doppler ultrasound signals taken from right ulnar arterial and (ii) the classification of RA disease as healthy and patient using ANN. The AR Burg model degrees of 5, 10, 15, 20, and 25 were used in the part of feature extraction from Doppler signals belonging to right ulnar arterial. Later, multilayer feed forward ANN trained with a LM back propagation algorithm was used to classify the Doppler signals belonging to right ulnar arterial whether the patient has RA or not.

II. MATERIAL

A. Hardware and Demographic Acknowledgments

The study included 24 patients with RA diseases, (2 men and 22 women with an age of 38 -70, mean: 48 years) and 16 healthy (1 man and 15 women with an age of 37 - 64, mean: 48 years).

Doppler signal acquisition was conducted by General Electric LOGIQ S6 Power Doppler Ultrasound Unit in the Radiology Department, Meram Faculty of Medicine, Selcuk

University. The system hardware was provided of Power Doppler Ultrasound unit that can work in the pulsed mode, linear ultrasound probe (12 MHz), input – output card and a laptop personal computer (Fig. 1). A laptop personal computer was used for storage, displaying and spectral analysis of the acquired Doppler data. Before Doppler data were recorded, a color and pulsed Doppler ultrasound examination of the right ulnar arterial was performed in order to exclude the presence of a hemodynamically significant stenosis. A linear ultrasound probe of 12 MHz was used to transmit pulsed ultrasound signals into the right ulnar arterial. Signals reflected from the arterial were recorded to derive out the Doppler shift frequencies. In all tests performed on the patients and healthy subjects, the insonation angle and the presetting of the ultrasound were kept fixed. The insonation angle was adjusted via electronic steering methods and manually in order to keep a constant value of 60 degrees on a longitudinal view. The sampling volume was placed within the center of the arterial. The amplification gain was carefully set to take a clean spectral output with minimized background noise on the spectral display [15-18]. The audio output of ultrasound unit was sampled at 44100 Hz and then sent to a laptop computer via an I/O card.

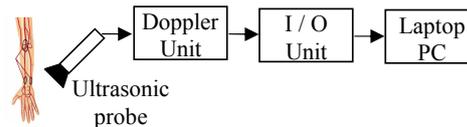
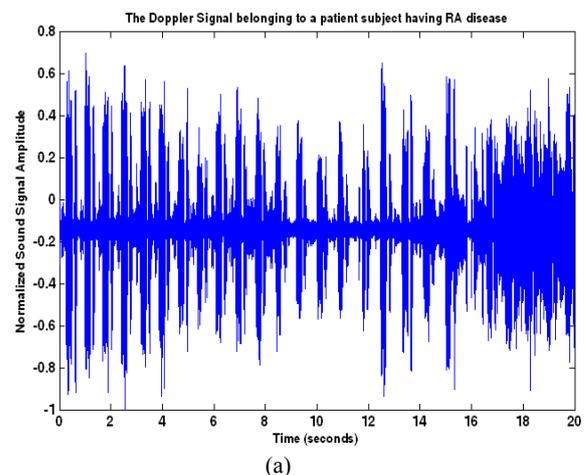


Fig. 1. Block diagram of the system hardware used to acquire Doppler data.



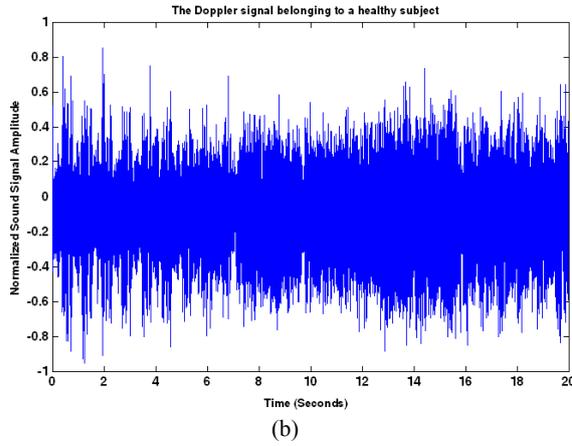


Fig. 2. Doppler signals for (a) a patient subject having RA disease and (b) a healthy subject on the right ulnar arterial

Figure 2 presents the Doppler signals for a patient subject having RA disease and a healthy subject on the right ulnar arterial. As shown in these figures, the difference between healthy Doppler signals and patient Doppler signals is clear. Transforming the Doppler signals from time domain to frequency domain using AR Burg method, RA disease has been diagnosed.

III. METHOD

In this paper, we have proposed a system which consists of three parts: (i) measurement of right ulnar arterial Doppler signals, (ii) feature extraction from right ulnar arterial Doppler signals using AR Burg spectral analysis method, (iii) classification of right ulnar arterial Doppler signals using neural network. Figure 3 shows the flowchart of proposed method. We have explained the following subsections below.

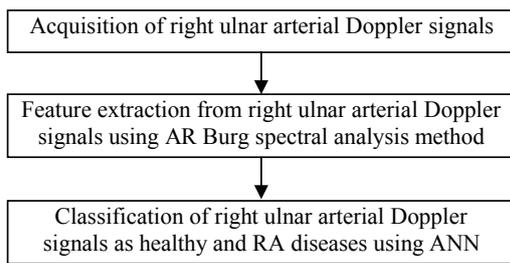


Fig. 3 The flowchart of proposed method

A. Feature Extraction using AR Burg Method

AR model is the most widely used model, first it is suitable for representing spectra with narrow peaks and secondly a number of linear equations need to be solved for finding the AR parameters. The Burg method is a least-squares optimization problem with the constraint from which the reflection coefficients obtained. The problem must satisfy

Levinson-Durbin recursion. The Burg method has three advantages. The Burg method has high frequency resolution, it results stable AR model and it uses Levinson-Durbin recursion [19-20].

An AR process model is defined as

$$v(n) = x(n) + a_1^* x(n-1) + \dots + a_p^* x(n-p) \quad (2)$$

where $\{x(n), x(n-1), \dots, x(n-p)\}$ is a realization of the process and $v(n)$ is a white-noise process with variance σ^2 .

Given the AR parameters $\{a_i\}$, the power spectrum density (PSD) of $\{x(n)\}$ is found as

$$\hat{P}_{AR}(f) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^p a_k^* e^{-j2\pi fk}\right|^2} \quad (3)$$

$$\min_{n=m}^{N-1} \left[|f_m(n)|^2 + |b_m(n)|^2 \right] \quad (4)$$

subject to constrain

$$a_m(k) = a_{m-1}(k) + \lambda_m a_{m-1}^*(m-k) \quad 1 \leq k \leq m-1, \quad 1 \leq m \leq p \quad (5)$$

which results as follows:

$$\lambda_m = \frac{-\sum_{n=m}^{N-1} f_{m-1}(n) b_{m-1}^*(n)}{\frac{1}{2} \sum_{n=m}^{N-1} \left[|f_{m-1}(n)|^2 + |b_{m-1}(n)|^2 \right]} \quad (6)$$

B. Classification of Right Ulnar Arterial Doppler Signals Using ANN

An ANN is a data processing system where data spreads parallel on. An ANN can determine its conditions and adjust itself to enable different responses by using inputs and desired outputs, which are provided to the system. The most charming thing about an ANN is that it works as an expert system that will lastly help the physicians with the decision processing about the existence of the disease. An ANN is trained with the available data samples to discover the association between inputs and outputs [20, 21].

A multilayer feed forward ANN trained with a LM back propagation algorithm was used to classify the Doppler signals belonging to right ulnar arterial as patient having RA disease or healthy. The advantage of using this type of ANN is the rapid execution of the trained network, which is especially advantageous in signal processing applications. ANN training is normally formulated as a non-linear least-squares problem [20].

Figure 4 shows the structure of ANN with LM used in our

experimental studies. In the input layer, hidden layer, and output layer, there are 129 features (neurons), 10 hidden neurons, and 2 outputs (healthy and patient), respectively.

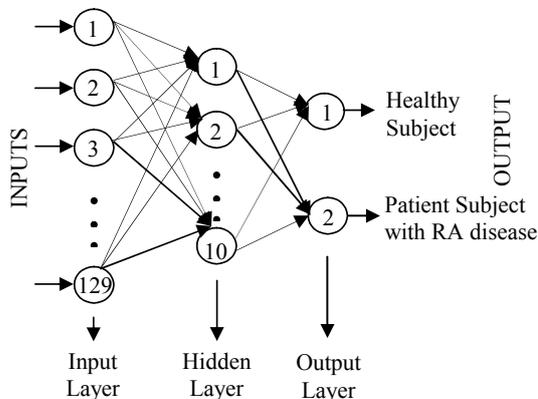


Fig. 4. The structure of ANN with LM

The back propagation algorithm is a widely used training procedure that adjusts the connection weights of a multilayer perceptron. As a matter of fact, the LM algorithm is a least-squares estimation algorithm based on the maximum neighborhood idea. A multilayer perceptron consists of three layers: an input layer, an output layer, and one or more hidden layers. Each layer is composed of a predefined number of neurons [22].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, the effect of the model degrees of Burg AR spectral analysis method to early diagnosis of RA disease has been examined. First, the Burg AR spectral analysis method was used to extract the relevant features from Doppler signals belonging to healthy and patient subjects having RA disease. In this part, we have used the various model degrees in Burg AR method and applied to Doppler signals. For each model degree, the power spectral density (PSD) values were obtained. And these obtained PSD values were applied to input of ANN to classify the Doppler signals as healthy and patient subject having RA disease.

In training and testing of ANN, the data partition of 50-50% train-test has been used. In our dataset, there are 24 patients with RA diseases and 16 healthy subjects. Totally, 40 subjects were used to diagnose the RA disease. In order to evaluate the performance of ANN models, the classification accuracy, ROC (Receiver Operating Characteristic) curves, sensitivity and specificity values have been used. Table 1 shows the obtained results for model degrees of 5, 10, 15, 20, and 25 in Burg AR method in the early diagnosis of RA disease. And also, we have given the obtained ROC curves for 5, 10, 15, 20, and 25 in Burg AR method and showed in Figure 5.

These results have shown that the model degrees of 10 and 20

in Burg AR method have obtained best classification results in classification of right ulnar arterial Doppler signals.

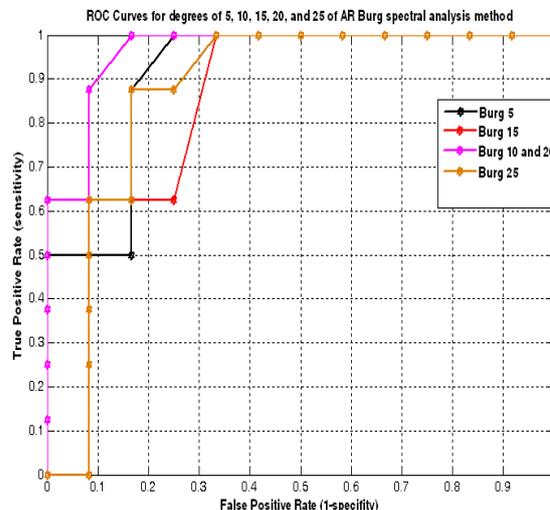


Fig.5. ROC curves for model degrees of 5,10,15,20 and 25 of AR Burg spectral analysis method on the early diagnosis of RA diseases

TABLE I
THE OBTAINED RESULTS FOR MODEL DEGREES OF 5, 10, 15, 20, AND 25 IN BURG AR METHOD IN THE EARLY DIAGNOSIS OF RA DISEASE FROM ANN CLASSIFIER

Model Degree in Burg AR method	Classification Accuracy (%)	Sensitivity (%)	Specificity (%)
5	90	100	80
10	95	92,30	100
15	85	80	100
20	95	92,30	100
25	85	80	100

V. CONCLUSION

In this paper, we have examined the effect of model degrees in Burg AR spectral analysis to early detection of RA disease using Doppler signals belonging to right ulnar arterial obtained from patient and healthy subjects. For model degrees of 5, 10, 15, 20, and 25 in Burg AR method, the RA disease has been diagnosed by using ANN classifier. In the obtained results, the best prediction accuracy has been achieved for model degrees of 10 and 20. This study is a new work related to diagnosis of RA disease using right ulnar arterial Doppler signals belonging to healthy and patient subjects. In future, we will increase the number of patient and healthy subjects to further evaluate proposed method. Therefore, this study is the preliminary study conducted by us. This system can be useful for physicians to make the final decision without hesitation, on early diagnosis of RA disease.

ACKNOWLEDGMENT

This work is supported by the Scientific Research Projects of

REFERENCES

- [1] D. M. Lee, M. E. Weinblatt ME, " Rheumatoid arthritis" *The Lancet* 2001;358: pp. 903-911.
- [2] S. Özgöçmen , H. Özdemir, A. Kiriş , Z. Bozgeyik and O Ardicoglu, "Clinical Evaluation and Power Doppler Sonography in Rheumatoid Arthritis: Evidence for Ongoing Synovial Inflammation in Clinical Remission", *Southern Medical Journal*, 101, 2008, :pp. 240-245.
- [3] L. Carmona , V. Villaverdeı , C. H. Garcia , J. Ballina , R. Gabriel and A. Laffon , "The prevalence of rheumatoid arthritis in the general population of Spain", *Rheumatology* 41, 2002, pp. 88-95.
- [4] E. Naredo, G. Bonilla, F. Gamero, J. Uson, L. Carmano and A. Laffon, "Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power doppler ultrasonography " *Ann Rheum dis*:64, 2005,pp. 875-881.
- [5] J. Teh, K. Stevens, L. Williamson, J. Leung and G. McNally, "Power Doppler Ultrasound of Rheumatoid Synovitis: quantification of therapeutic response", *The British Journal of Radiology*:76, 2003, pp. 875-879.
- [6] K. Ellegard, S. T. Pedersen, L. Terslev, B. D. Samsøe, M. Henriksen and H. Bliddal, "Ultrasound colour Doppler measurements in a single joint as measure of disease activity in patients with rheumatoid arthritis—assessment of concurrent validity", *Rheumatology* 48(3), 2009, pp. 254-257.
- [7] C. K. Yeh and P. C. Li, " Doppler angle estimation of pulsatile flows using AR modeling", *Ultrasonic Imaging* 24, 2002, pp. 135 -146
- [8] B. Sigel, "A brief history of Doppler ultrasound in the diagnosis of peripheral vascular disease", *Ultrasound Med. Biol.* 24(2),1998, pp. 169 - 176
- [9] D. H. Evans, W. N. McDicken, R. Skidmore and J. P. Woodcock, "Doppler Ultrasound: Physics, Instrumentation and Clinical Applications", Wiley, Chichester, 1989
- [10] İ. Güler, F. Hardalaç and E. D. Übeyli, "Determination of Behçet disease with the application of FFT and AR methods", *Comput. Biol. Med.* 32, 2002, pp. 419 - 434
- [11] W. A. Schmidt, " Doppler sonography in rheumatology", *Best Practice & Research Clinical Rheumatology*, Vol. 18, No:6, 2004, pp. 827 – 846
- [12] A. Kiriş, S. Özgöçmen, E. Kocakoç and Ö. Ardiçoğlu, " Power Doppler assessment of overall disease activity in patients with rheumatoid arthritis", *Journal of Clinical Ultrasound*, vol. 34, No.1,2006, pp. 5 -11
- [13] L. Terslev, P. von der Recke, S. Torp-Pedersen, M. J. Koeing and H. Bliddal, " Diagnostic sensitivity and specificity of Doppler ultrasound in rheumatoid arthritis", *The Journal of Rheumatology* 35:1, 2008, pp. 49 - 53
- [14] K. Varsamidis, E. Varsamidou, V. Tjetjis and G. Mayropoulos, " Doppler sonography in assessing disease activity in rheumatoid arthritis", *Ultrasound in Medicine & Biology*, vol.31, Issue 6, 2005, pp. 739 - 743
- [15] S. Kara, " A study of Mitral and tricuspid valve blood flows by autoregressive spectral analysis method and Doppler unit", Thesis of Doctorate, Institute of Science of Erciyes University Press, 1995
- [16] J. Joseph and J. M. Brown, "Introduction to Biomedical Equipment Technology", Prentice-Hall, Englewood Cliffs, New Jersey, 1993
- [17] P. Atikson, " Doppler ultrasound and its use in clinical measurement", Prentice-Hall, Englewood Cliffs, New Jersey, 1982
- [18] J. W. Arenson, " Real-time two-dimensional blood flow imaging using a Doppler ultrasound array", McGraw-Hill, New York, 1982
- [19] X. Wang and K.K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*", *The Journal of The Pattern Recognition Society* 36,2003, pp. 2429 - 2439
- [20] F. Dirgenali and S. Kara, " Recognition of early phase of atherosclerosis using principles component analysis and artificial neural networks from carotid artery Doppler signals", *Expert Systems with Applications* 31, 2006, pp. 643 - 651
- [21] P.K. Simpson, " Artificial neural systems", New York, Pergamon Press, 1989
- [22] İ. Güler and E. D. Übeyli, "Detection of ophthalmic artery stenosis by least-squares back propagation neural network", *Computers of Biology and Medicine*,32, 2003, pp. 333 – 343

Nonlinear Optimization

Contents

Nonlinear regression analysis of optimal conditions for a physics-chemical process Vyacheslav RUSANOV, Sergey AGAFONOV, Alexey DANEEV, Dmitry SHARPINSKIY, Timofey UDILOV	132
H_∞ optimal model reduction of complex systems using particle swarm optimization Maamar BETTAYEB, Reem SALIM	138
H2 optimal model reduction of dynamic systems with time-delay using particle swarm optimization Reem SALIM, Maamar BETTAYEB	144
Optimization algorithm for combined economic and emission dispatch with security constraints Nurettin ÇETINKAYA	150
Bandwidth optimization of a PIFA antenna using hybrid genetic algorithms Mohammad Riyad AMEERUDDEN, Harry C. S. RUGHOOPUTH	154

Nonlinear Regression Analysis of Optimal Conditions for a Physics-Chemical Process

V.A. Rusanov, S.V. Agafonov, A.V. Daneev, D.Yu. Sharpinskiy, T.V. Udilov
Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences

A technique of nonlinear mathematical programming good for grounding an optimal technological process of nitrogenization in a distributed environment of electrostatic field is proposed. The technique is based on the quadratic approximation for deviations of the vector argument of deviations of the vector argument of physics-chemical factors of metal working from some given regime of nitrogenization and imposes minimal requirements to experimental data in the process of identification of the mathematical model of the process of obtaining a nitrogenized layer.

Identification, Nitrogenization, Optimization, Regression Analysis

I. INTRODUCTION

A classical view to mathematical modeling implies a descriptive approach characteristic of a physicist: the functions bound up with natural phenomena are subject to definite universal principles (laws), and the problem is to discover them. But the practice of descriptive sciences is different. The central conception sooner presumes that mathematical modeling consists in following the principle: *the desired optimal model is simply the most exact model within the limits of a given admissible level of complexity or the least complex model, which approximates the (experimental) data observed with a precision up to a given admissible discoordination*. The idea of formalization of considerations of model's complexity, which relates to the theory of identification of systems, was investigated in [1,2]. From the viewpoint put forward by L. Ljung [3,4], the idea that identification algorithms (by all means) have interpretation in the language of optimal approximation, is the main one. In the present paper we essentially employed both of the indicated approaches, i.e. we have outline a combined methodology, which forms the ground of the procedure of optimal nonlinear approximation in the process of mathematical modeling of the process of nitrogenization of a mechanical part's surface to be processed under the conditions of effect of some inversive electrostatic field (with a non-stationary potential), within the frames of some linear-quadratic representation of vector regression equations.

This work was supported by Basic Research Program No. 22 of the Presidium of the Russian Academy of Sciences (Project No. 2.5), by the Russian Foundation for Basic Research (Project No. 05-01-00623), and by the Program "Leading Scientific Schools" (Project No. NSh-1676.2008.1).

II. STATEMENT OF THE PROBLEM OF SYNTHESIS OF OPTIMAL MULTI-DIMENSIONAL REGRESSION

In principle, static models of the type "input-output" may be obtained from dynamic ones by applying experimental stationary finite values (or, what is equivalent, for the zero frequency). Unfortunately, the dynamic model is generally linearized, what is inadmissible for the static model, when this model is to be used for the purpose of optimization within a substantial band. Furthermore, the static model must be more detailed than the dynamic one (optimization, which improves the productivity by some 1%, already represents a substantial interest from the application viewpoint), so, the structural-parametric identification of the multi-dimensional static nonlinear system of the type "input-output" in the absence of complete a priori understanding (knowledge) of the physics-mathematical principles of its functioning, a so called mathematical model of "black box", deserves an attentive deep consideration, especially when we have to ground the admissible level of complexity of the process under scrutiny.

From now on, R is the field of real numbers; R^n is an n -vector space over R (with the Euclidean norm denoted by $\|\cdot\|_R^n$); $M_{n,m}(R)$ is the space of all the $n \times m$ -matrices (i.e. the matrices of dimension $n \times m$) with the elements from R and with the Frobenius matrix norm $\|D\|_F := (\sum d_{ij}^2)^{1/2}$, $D = [d_{ij}]$ (what is equivalent to $D \in M_{n,m}(R) \Rightarrow \|D\|_F = (\text{tr } D^T D)^{1/2}$); as usually, the symbol $:=$ denotes the equality by definition; \det is a matrix determinant; $\text{tr } G := \sum g_{ii}$ is the trace of quadratic matrix G (the sum of its diagonal elements); $\langle \cdot \rangle^T$ is the operation of transposition of a matrix; E_n is a unit $n \times n$ -matrix; $\text{col}(a_1, \dots, a_n)$ is a column vector with real elements a_1, \dots, a_n .

A normal approach in the theory of identification of complex systems of the type of "input-output" methodologically consists in [5] *a priori* fixation of some partially parameterized class of stationary models and then, on the basis of fixed a posteriori data, to choose the parameters of the model's equations, which would minimize some formal criterion. In essence, this approach may be considered as application of the first method (denoted in the Introduction), in which "adjustment of the model's parameters" (under a fixed number of free coefficients in its equations) is conducted. In this case, the criterion is defined by the model's complexity chosen a priori. So, for the purpose of further consideration, let us identify a class of stationary static interconnected nonlinear systems of the type

“input–output”, which are described by the vector-matrix regression equation of the form

$$y=c+Au+\text{diag}[u^T B_1 u, \dots, u^T B_n u] \text{col}(1, \dots, 1) + \varepsilon(u); \quad (1)$$

$y \in R^n$ is the vector of system’s output signals, $u \in R^m$ is the vector of system’s assigning influences, $c \in R^n$, $A \in M_{n,m}(R)$, $B_i \in M_{m,m}(R)$, $B_i^T = B_i$ ($1 \leq i \leq n$) and $\text{diag}[\dots]$ is the diagonal $n \times n$ -matrix of corresponding bi-linear controlling influences $u^T B_i u$. As far as the vector function $\varepsilon(u)$ is concerned, we presume that the structure of its analytical representation is *a priori* unknown, but on the whole, it inexplicitly depends on the choice of the linear $c+Au$ and bi-linear $\text{diag}[u^T B_1 u, \dots, u^T B_n u]$ components of the input signal – because the nonlinear component $\varepsilon(u)$ of equation (1) may always be considered as a residual (“under-modeled”) term of the expansion of its right-hand side.

It is clear, the result y , predicted by the *linear-quadratic form* (LQF) $c+Au+\text{diag}[u^T B_1 u, \dots, u^T B_n u]$ of the right-hand side of equation (1), shall differ from the real signal, because the nonlinear law $\varepsilon(u)$ introduces some influence. On the other hand, as noted above, the analytical representation of the term $\varepsilon(u)$ depends on the choosing (fixation) of coefficients of the LQF. As a result, on the stage of identification, correction consists in varying the parameters of the LQF so that the results obtained, and those predicted on the basis of the LQF, would maximally coincide with each other. Obviously, new forecasts and parametric correction may then be conducted operatively (furthermore, additional information is used mainly for conducting partial or complete analysis of adequacy of the model on the basis of the latest current measurements). In other words, speaking more formally, the methodological paradigm of the a posteriori-optimal parametric synthesis of LQF shall provide for $\min \|\varepsilon(u)\|_R^n$ on the family of the representative sample of the field experiments conducted. When we proceed to the “language of formulas”, this paradigm acquires the form of the following optimization problem.

Statement of the problem of a posteriori-optimal parametric synthesis of LQF for the equation of nonlinear regression: find a vector-matrix solution $c, A, B_i, i=1, \dots, n$ bi-criterion problem

$$\left\{ \begin{array}{l} \min(\sum_{l=1, \dots, k} (\|y(l) - c - Au(l) - \text{diag}[u^T(l) B_1 u(l), \dots, u^T(l) B_n u(l)] \cdot \text{col}(1, \dots, 1)\|_R^n)^2)^{1/2}, \\ \min(\|c\|_R^n + \|A\|_R^2 + \sum_{i=1, \dots, n} \|B_i\|_R^2)^{1/2}, \end{array} \right. \quad (2)$$

$y(l) \in R^n$, $u(l) \in R^m$ are vectors of experimental data (here $y(l)$ is the “reaction” to the input influence $u(l)$), k is the number of experiments completed; noteworthy, there are no methodological constraints imposed on the value of k .

Remark 1. The first condition – $\min \sum \dots$ in the mathematical statement (2) guarantees – by the general sample of k field experiments – the optimal linear-quadratic approximation of the scrutinized physical process in terms of the nonlinear regression model (1); the second condition – provides (in the case of *non-uniqueness* of the solution for the

first $\min \sum \dots$) for parametric concretization of such a model with the property of the minimal matrix norm.

III. PARAMETRIC IDENTIFICATION OF THE LQF-STRUCTURE OF EQUATIONS OF NONLINEAR VECTOR REGRESSION

Let us relate the identification algorithm in the multi-criterion problem statement (2) for the interconnected stationary nonlinear system “input–output” of class (1) to the concept of *normal pseudo-solution* (or, what is equivalent, of canonical solutions by the method of least squares) for the system of linear algebraic equations.

Definition 1 [6, p. 501]. *Vector $x \in R^p$ is called the normal pseudo-solution of the system of linear equations $Dx=d$, $D \in M_{q,p}(R)$, $d \in R^q$. This vector has the smallest Euclidean norm $\|x\|_R^p$ among all the vectors, which make minimum the value of $\|Dx-d\|_R^q$.*

Let $D \in M_{q,p}(R)$ and D^+ be the inverted reciprocal (pseudo-inverse) Moore-Penrose matrix [6, p. 500] for matrix D . The asymptotic construction of the pseudo-inverse matrix has the following analytical form: $D^+ = \lim_{\tau \rightarrow 0} \{D^T(DD^T + \tau E_q)^{-1} \tau\}$. From now on, the mnemonic sign « $^+$ » denotes the operation of pseudo-inverting of the respective matrix.

Lemma 1 [7, p. 35]. *Vector $x=D^+d$ represents a normal pseudo-solution of the linear system $Dx=d$, $D \in M_{q,p}(R)$, $d \in R^q$.*

For the purpose of “interrelation” between the variables of input effects on the data of the general sample, let us denote by $\hat{u}(l)$ the $(1+m(m+3)/2)$ -vector, which has the following coordinate representation:

$$\hat{u}(l) := \text{col}(1, u_1(l), \dots, u_m(l), u_1(l)u_1(l), \dots, u_r(l)u_s(l), \dots, u_m(l)u_m(l)) \in R^{m(m+3)/2}, \quad 1 \leq r \leq s \leq m, \quad (3)$$

$$\text{col}(u_1(l), \dots, u_m(l)) := u(l) \in R^m, \quad 1 \leq l \leq k.$$

Let us call $U := [\hat{u}(1), \dots, \hat{u}(k)]^T \in M_{k, 1+m(m+3)/2}(R)$ the full matrix of experimental data related to input effects, respectively, $\beta_i := \text{col}(y_i(1), \dots, y_i(k)) \in R^k$ – the fill vector of experimental data related to output signal y_i ($i=1, \dots, n$). Next, orienting to the linear-parametric description of the coefficients for the nonlinear model of the type “input–output” for the output signal y_i , let us write down – due to system (1) – the linear-quadratic form of its regression equation

$$c_i + \sum_{1 \leq j \leq m} a_{ij} u_j + \sum_{1 \leq q \leq p \leq m} b_{iqp} u_q u_p, \quad (i=1, \dots, n). \quad (4)$$

Now introduce the $(1+m(m+3)/2)$ -vector of regression model’s parameters. Obviously, due to (4), any fixed set of n such vectors completely defines the representation of the LQF with respect to some “input–output” model of type (1): $z_i := \text{col}(c_i, a_{i1}, \dots, a_{im}, b_{i11}, \dots, b_{iqp}, \dots, b_{imm}) \in R^{1+m(m+3)/2}, \quad 1 \leq q \leq p \leq m$.

Proposition 1. *The optimization problem (2) has the solution*

$$z_i^* = U^+ \beta_i, \quad i=1, \dots, n; \quad (5)$$

here U is a complete matrix of experimental data related to

input effects, β_i is the full vector of experimental data related to output signal y_i ($i=1, \dots, n$).

Remark 2. Qualitative estimates *a*), *b*) from Corollary 1 depend mainly on the volume of a posteriori information (number of experiments *k*), i.e. if $k > 1+m(m+3)/2$, then, as a rule, realized is item *a*); if $k \leq 1+m(m+3)/2$ then it is quite probable that realized is item *b*).

IV. MODELING OF THE LINEAR-QUADRATIC STRUCTURE OF EQUATIONS OF VECTOR REGRESSION FOR THE PROCESS OF NITROGENIZATION

Without any loss of generality, in the capacity of the initial (zero) position of the vector of input control influences *u* it is possible to accept some empirically identified (from the general set of experimental data) point ω of space R^m ; obviously, in this case, coordinates u_1, \dots, u_m of vector *u* shall be considered as deviations with respect to the regime ω .

The process of nitrogenization in the environment of inversive electrostatic field in a series of field experiments ($k=12$) may be described in terms of the following variables:

vector $y = \text{col}(y_1, y_2, y_3) \in R^3$ of controlled characteristics of nitrogenization:

y_1 – Vickers surface hardness number 10^{-1} [HV],

y_2 – specific wear 10^{-1} [mg/cm²],

y_3 – depth of the nitrogenized layer 10^{-2} [mm];

vector $u = \text{col}(u_1, u_2, u_3, u_4) \in R^4$ of variations of the regime's parameters $\omega = \text{col}(\omega_1, \omega_2, \omega_3, \omega_4)$:

u_1 – variation (w.r.t. ω_1) of the degree of dissociation of ammonium 10^{-1} [%],

u_2 – variation (w.r.t. ω_2) of the temperature of the process 10^{-1} [°C],

u_3 – variation (w.r.t. ω_3) of the duration of the process 10^{-1} [h],

u_4 – variation (w.r.t. ω_4) of the voltage on the electrodes 10^{-3} [V].

TABLE I

EXPERIMENTAL DATA OF THE PROCESS OF OBTAINING THE NITROGENIZED LAYER ARE: $\omega_1=45\%$, $\omega_2=500^\circ\text{C}$, $\omega_3=25\text{ h}$, $\omega_4=1900\text{ V}$

Experiment	Assigning influences				Nitrogen layer parameters		
	u_1	u_2	u_3	u_4	y_1	y_2	y_3
№							
1	0	0	-1	0,4	80,3	6,0	14
2	1	5	-1	0,4	93,3	3,4	22
3- ω	0	0	0	0	97,4	13,1	17
4	1	0	0	3,4	84,7	12,2	18
5	0	5	0	3,4	79,2	10,3	28
6	1	0	-1	3,8	54,8	42,4	11
7	0	5	-1	3,8	87,0	11,9	25
8	1	5	0	0	89,4	3,5	33
9	0,4	3	-0,4	0,18	87,0	4,2	22
10	0,3	3,5	-0,3	0,16	92,0	3,8	24
11	0,2	4	-0,2	0,14	98,8	4,0	28
12	0	5	0	0,1	89,4	3,6	34

Note, direct application of analytical methods developed above, results in not very complex but bulky computations

(below the computation was conducted in the environment of MATLAB [8]); for example, according to Table 1, matrix *U* has the dimension of $k \times 1+m(m+3)/2=12 \times 15$, and the matrix pseudo-inverse with respect to U^+ has, respectively, the dimension of 15×12 .

Taking into account the solution of the parametric optimization problem (4)-(6) and equation of the model of linear-quadratic vector regression (which describes in terms of a multi-dimensional polynomial approximation the interconnected process of nitrogenization in the environment of inversive electrostatic field, which possesses the variation of the potential due to the parametric representation of the vector structure U^+ , and also, according to Table 1, of vectors β_i , $i=1, \dots, 3$) have the form:

$$\begin{aligned}
 y_1(u) &= 97,4 - 65,075u_1 - 3,706u_2 + 9,369u_3 + 5,991u_4 - 64,313u_1^2 \\
 &\quad + 25,339u_1u_2 + 11,2136u_1u_3 + 7,159u_1u_4 + 0,529u_2^2 \\
 &\quad - 8,346u_2u_3 - 6,161u_2u_4 - 8,607u_3^2 + 6,29u_3u_4 + 6,227u_4^2; \\
 y_2(u) &= 13,1 - 9,098u_1 - 2,232u_2 + 4,435u_3 - 2,235u_4 - 8,648u_1^2 \\
 &\quad + 3,531u_1u_2 - 15,604u_1u_3 + 5,491u_1u_4 + 0,067u_2^2 \\
 &\quad + 2,361u_2u_3 + 0,502u_2u_4 - 3,986u_3^2 - 5,336u_3u_4 + 0,5u_4^2; \\
 y_3(u) &= 17 + 0,398u_1 + 0,964u_2 + 1,388u_3 - 0,437u_4 + 0,226u_1^2 \\
 &\quad - 0,424u_1u_2 + 5,264u_1u_3 - 1,84u_1u_4 + 0,507u_2^2 + 0,091u_2u_3 \\
 &\quad - 0,772u_2u_4 - 1,56u_3^2 - 0,027u_3u_4 + 0,702u_4^2.
 \end{aligned} \tag{7}$$

Critical analysis of the “predicted efficiency” of the proposed model intended for nonlinear mathematical description of the physics-chemical properties of the process of nitrogenization expressed in terms of *quasi-linear vector-matrix* regression equations (1), i.e. by the system (7), allows to conduct the relative comparison of the latter three columns of Table 1 with the following table obtained due to (7).

TABLE II

Number in the forecast	Forecast for the nonlinear regression model		
№	$y_1(u)$	$y_2(u)$	$y_3(u)$
1	80,3	6,0	14
2	93,3	3,4	22
3	97,4	13,1	17
4	84,7	12,2	18
5	79,2	10,3	28
6	54,8	42,4	11
7	87	11,9	25
8	89,4	3,5	33
9	86,226	4,129	21,782
10	94,066	3,989	24,582
11	97,251	3,858	27,564
12	89,658	3,624	34,073

In the next section, we are going down to the multi-dimensional geometric investigation of “minimax” properties of solutions for the nonlinear vector regression, which describes electrostatic nitrogenization of the processed part surface, to the end of finding the regime of wear resistance and corrosion resistance for the geometry of its part. An

interesting trait of the analytical results obtained is their explicit algebraic dependence on the parameters of (7).

V. INTERPOLATION OF THE PHYSICS-TECHNOLOGICAL CHARACTERISTICS OF THE NITROGEN LAYER. OPTIMIZATION OF THE PROCESS OF NITROGENIZATION

After all, the main objective of mathematical modeling is to answer the question “How it can the scrutinized physical process proceed and how it must proceed actually under some external controlling influence?”. The answer to the second part of the question gives the solution of the optimization problem (9), while the answer to the first presumes the following fact:

Proposition 2. *The indicator of quality of nitrogenization $J_i(u) := y_i(u)$, ($i=1, \dots, n$) may have the internal maximum or minimum in the identified LQF-structure of equations of nonlinear regression only at point $u_i^* \in R^m$:*

$$u_i^* = -B_i^{-1} A^T e_i / 2, \quad (8)$$

$\{e_1, \dots, e_n\}$ – basis in R^n . Furthermore, when $u^T B_i u$ is a negative definite quadratic form, the indicator $J_i(u)$ has maximum at point (8); when $u^T B_i u$ is a positive definite quadratic form, the indicator $J_i(u)$ has minimum at u_i^* . In the case, when $u^T B_i u$ assume both positive and negative values, we encounter the stationary point of more complex type, i.e. the so called saddle point.

Corollary 2. *If matrix B_i is positive definite (similarly, negative definite) then the minimum (resp. maximum) value of the quality indicator $J_i(u)$ is $c_i e_i^T A B_i^{-1} A^T e_i / 4$.*

Now we can solve the analytical problem, which has been the stimulus to investigation of positiveness (or negativity) of quadratic forms from equation (7), i.e. to answer the question – when the stationary point (8) is the point of relative minimum, maximum or the saddle point. Speaking more formally, the problem of defining the positive (or negative) algebraic definiteness of the quadratic forms $u^T B_i u$ has been reduced to the geometric problem of rather general type – computing of eigenvalues λ_{ij} ($i=1, \dots, 3$; $j=1, \dots, 4$) of symmetric matrices B_i ($i=1, \dots, 3$):

- $\lambda_{11} = -67,5644$, $\lambda_{12} = -9,2743$, $\lambda_{13} = 1,8251$, $\lambda_{14} = 8,8491$, what speaks about the existence of a stationary saddle point for the goal functional $y_1(\cdot): R^4 \rightarrow R$;
- $\lambda_{21} = -14,7856$, $\lambda_{22} = -2,6697$, $\lambda_{23} = 0,362$, $\lambda_{24} = 5,0252$, what speaks about the existence of a stationary saddle point for the goal functional $y_2(\cdot): R^4 \rightarrow R$;
- $\lambda_{31} = -3,5248$, $\lambda_{32} = 0,0847$, $\lambda_{33} = 0,8665$, $\lambda_{34} = 2,4482$, what speaks about the existence of a stationary saddle point for the goal functional $y_3(\cdot): R^4 \rightarrow R$.

While combining previous results, the standard regime of nitrogenization, which provides for maximum hardness, wear resistance and the thickness of the physical structure of nitrogen layer of the processed surface of a mechanical part, let us relate them to the solution of the optimization problem of the following form

$$\max \{F(u) = r_1 J_1(u) + r_2 J_2(u) + r_3 J_3(u) : u \in R^4\}, \quad (9)$$

$$J_i(u) := y_i(u) \quad (i=1, \dots, 3)$$

where the weighting coefficients r_i , $i=1, \dots, 3$ of the goal functional $F(u)$ must be chosen, while proceeding from the considerations of proper expert assessment of the differentiated effect of the quality indicators $J_i(u)$, $i=1, \dots, 3$ [9]. We have considered the following weighting coefficients: $r_1=0,5$; $r_2=-0,3$; $r_3=0,2$; the sign «-» with the coefficient r_2 means that the problem statement (9) actually provides for relative minimization (!) of the parameter of specific wear y_2 (what is equivalent, displacement to the point of min $J_2(u)$ in the linear structure of functional $F(u)$). This allows us to write down the goal functional (9) in the following analytical form:

$$F(u) = 48,17 - 29,729u_1 - 0,99u_2 + 3,632u_3 + 3,579u_4 - 29,517u_1^2 + 11,526u_1u_2 + 11,341u_1u_3 + 1,564u_1u_4 + 0,346u_2^2 - 4,863u_2u_3 - 3,4u_2u_4 - 3,42u_3^2 + 4,74u_3u_4 + 3,1u_4^2. \quad (10)$$

In this case, parameters of the variations have constituted the following intervals (in terms of relative physics units): $u_1 = \pm 40\%$, $u_2 = \pm 50$ °C, $u_3 = \pm 5$ h, $u_4 = \pm 1000$ V.

Development of new techniques of alloying metals necessitates existence of an adequate mathematical model, which would be capable of predicting the reciprocal influence of different factors of the physics-chemical environment on the process of metal working, as well as revealing the influence of mechanical and geometric characteristics of the processed part's surface upon the results obtained. As far as the multi-factor process of nitrogenization is concerned, the mathematical model of optimization (9) gives such a possibility, i.e. the possibility to reveal the most critical parameters and give the defining directions of improving the exploited and developed technological installations intended for obtaining the nitrogenized layer. Proposition 2, and also formula (8), which allow to compute the geometric coordinates of the stationary point for the optimization problem (9), define (in terms of system (1)) the following highly efficient technological parameters of the regime of nitrogenization:

Proposition 3. *The stationary point $u^* \in R^4$ in the problem related to optimization of the regime of electrostatic nitrogenization (9) has the algebraic solution*

$$u^* = -(r_1 B_1 + r_2 B_2 + r_3 B_3)^{-1} (e_1 + e_2 + e_3)^T \text{diag} [r_1, r_2, r_3] A^T / 2, \quad (11)$$

in this case, the sufficient condition (that the given point ensures satisfaction of $\max \{F(u) : u \in R^4\}$) is the requirement that it is elliptic:

$$\det [b_{ij}]_q < 0, \quad q=1, \dots, 4 \quad (12)$$

or, what is equivalent, for the eigen-numbers λ_i of matrix $(r_1 B_1 + r_2 B_2 + r_3 B_3)$ we have $\lambda_i < 0$, $i=1, \dots, 4$; here $[b_{ij}]_q \in M_{q,q}(R)$ are the main sub-matrices [6, c. 30] of matrix $(r_1 B_1 + r_2 B_2 + r_3 B_3)$.

VI. DISCUSSION

Let us start from the remark that if condition (12) is not satisfied the stationary point (11) is possibly the saddle (hyperbolic) point of functional $F(u)$ and, consequently, additional analysis of coordinates (11) is required; when speaking more formally, the availability of the saddle point is guaranteed by the replacement – at least in one relation (but not in all relations) – of the inequality «<» from (12) with «>». In this case, a similar replacement of «<» with «≤» possible provokes the structure of the parabolic point.

Due to system (1) (or, what is equivalent, due to equation (10)) the stationary point (11) in the coordinate representation (of the vector-row) writes $u^{*T}=[0,1761 \quad 3,7794 \quad -0,5622 \quad 1,8787]$, or, the same, in terms of physical dimensions and “counting” from the regime ω , we have: $u^{*T}=[46,76 \% \quad 537,794 \text{ }^\circ\text{C} \quad 19,378 \text{ h} \quad -21,3 \text{ V}]$.

Let us show that the mathematical result (in particular, the coordinates of the stationary point of the regime of nitrogenization (11)) obtained above are in good correspondence with the logic of our physics related reasoning.

Since the eigen-numbers of matrix $(r_1B_1+r_2B_2+r_3B_3)$ are, respectively, $\lambda_1=-31,8762$, $\lambda_2=0,5298$, $\lambda_3=-3,276$, $\lambda_4= 5,1355$, this gives evidence that functional $F(u)$ has a stationary saddle point: $R^4 \rightarrow R$ for the weighting coefficients r_i , $i=1, \dots, 3$, chosen above.

According to (10), at the stationary point u^* obtained the functional $F(u)$ reaches its “max” with respect to variables u_1 and u_3 and “min”, respectively, with respect to u_2 and u_4 . The physical sense of this proposition implies the following: as far as the structure $F(u)$ is concerned, it is not possible to exceed (make larger) the degree of dissociation of ammonium by more than 46,76%, and the duration nitrogenization by more than 19,378 hours, and, furthermore, in this case, simultaneously, it is better not to decrease the temperature of the gas mixture below the level of 537.794°C, it is also better not to make the general potential of the electrostatic field smaller than 21.3 V. Otherwise, violation of these parameters shall provoke the reduction of efficiency of the process of nitrogenization in the aspect of reaching the technological indicator $F(u)$, which provides for the maximum surface hardness and the depth of the nitrogenized layer side by side with minimization of specific wear of the part processed.

If computed (predicted) coordinates of the stationary point (11) go beyond the confidence region of adequacy of the mathematical model (7) in virtue of some physics-technological factors-parameters, than it is necessary to conduct an additional practical experiment bound up with nitrogenization, which is “maximally close” to the coordinates (11) and introduce (in the capacity of the regime ω) the data of this experiment into the extended matrix of experimental data U , after what it is possible to conduct recomputation of all the stages of optimization of the process of obtaining the nitrogenized layer, which is described above (if there is the need, such an experiment and the process of identification of model (1) are to be repeated); this important improvement, in essence, methodologically extends the

standard [10] procedure of planning the experiment.

VII. CONCLUSION

We have described the process of constructing a nonlinear mathematical model of the type “input–output” for the process of nitrogenization in the distributed environment of electrostatic field. This model is used for technological computation of hardness parameters for the material of the metal part, whose surface is processed. It can be used for assessment of the specific mechanical wear, the depth of the nitrogenized layer, etc. This regression model uses the identified (on the basis of experimental data obtained) multi-dimensional quadratic equations, what allows the researcher to adequately describe the process of nonlinear diffusion in the process of “nitrogen-alloying” within a wide band of variations of i) the degree of dissociation of ammonium, ii) the temperature, iii) the duration of the process and iv) the electric voltage at the pair “anode–cathode”.

Deviations in the computed (predicted) values of the synthesized nitrogenized layer and experimental data revealed are hardly ever of principal character. This has given us the opportunity to propose an efficient mathematical technique (“a finite chain” of algebraic formulas) for computing optimal properties and parameters of nonlinear multi-factor regime of nitrogenization. The ideas explicated in the present paper may be developed in several directions of theoretical-applied investigations oriented to improvement of the algorithms of computing an optimal technology of nitrogenization in an electrostatic field proposed above, as well as to extending the frames of adequacy of regression equations of nitrogenization at the expense of additional investigation of the factors of its nonlinearity; these can be oriented to:

- determination and algorithmization of the procedure of choosing the weighting coefficients r_i , $1 \leq i \leq 3$ in (9), while proceeding from satisfaction of the algebraic conditions (12), which provide for the elliptic character of the stationary point of the goal functional (9);
- extension of the linear-quadratic form of regression equations (1) by the “Taylor expansion” of the vector function y of higher order;
- account (in the capacity of extended coordinates of the vector function y of the regression model) of such physics-mechanical parameters of the synthesized nitrogenized layer in the environment of some electrostatic field, such as the coefficient of dry friction for the surface processed and for the brittle nitrogenized layer obtained;
- constructing the process of identification of a nonlinear a posteriori–adaptive mathematical model of nitrogenization with an additional condition of presence of high-frequency electromagnetic field; determination (under such a problem statement) of high technological multi-factor parameters of the process of nitrogenization, and also obtaining optimal values for the length and the amplitude of the waves of electromagnetic oscillations.

REFERENCES

- [1] Caines P.E., "On the scientific method and the foundation of system identification." in *Modelling, Identification and Robust Control* (Byrnes C.I., Lindquist A., eds.), North Holland, Amsterdam, 1986, pp. 563–580.
- [2] Rissanen J., "Stochastic complexity and statistical inference". Unpublished manuscript, I.B.M. Research K54/282, San Jose, California, 1985.
- [3] Ljung L., Söderström T., *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Massachusetts, 1983.
- [4] Ljung L., "A non-probabilistic framework for signal spectra," in *Proc. 24th Conf. Decis. Control*, Ft Lauderdale, Florida, December, 1985, pp. 1056–1060.
- [5] Ljung L., *System Identification - Theory for the User*. Moscow: Nauka Publ., 1991 (in Russian).
- [6] Horn R.A., Johnson C.R., *Matrix Analysis*. Moscow: Mir. Publ., 1989 (in Russian).
- [7] Gantmacher F.R., *Theory of Matrices*. Moscow: Nauka Publ., 1988 (in Russian).
- [8] Andreyevsky B.R., Fradkov A.L., *Elements of Mathematical Modeling in Programming Environments MATLAB and SCILAB*. St. Petersburg: Nauka Publ., 2001 (in Russian).
- [9] Makarov I.M., Vinogradskaya T.M., Rubchinsky A.A., Sokolov V.B., *The Theory of Choosing a Decision and Decision Making*. Moscow: Nauka Publ., 1982 (in Russian).
- [10] Adler Yu.P., Markova E.V., Granovsky Yu.V. Planning, *Experiment in the Process of Finding Optimal Conditions*. Moscow: Nauka Publ., 1976 (in Russian).

H_∞ Optimal Model Reduction of Complex Systems Using Particle Swarm Optimization

Maamar Bettayeb

Electrical & Computer Engineering Department
University of Sharjah
Sharjah, United Arab Emirates
maamar@sharjah.ac.ae

Reem Salim

Electrical & Computer Engineering Department
University of Sharjah
Sharjah, United Arab Emirates
reemsalim@gmail.com

Abstract—Engineering physical systems can be represented using different mathematical modeling tools with finite dimensional models. However, the resulting models are usually complex and of high order. The complexity of these models imposes computational and implementation difficulties in simulation, analysis and control design. Several optimal model reduction methods that are mainly based on conventional analytical optimization techniques, either in time domain or in frequency domain, are available. However, these techniques do not usually lead to globally optimized reduced order models and, at best, converge to local optimal solutions. Thus, the use of an Evolutionary Algorithm, namely Particle Swarm Optimization (PSO) is adopted in this paper for its robustness in global optimization.

It is demonstrated in this work that, in the absence of analytic formulations to these challenging optimization problems, PSO leads to tractable solutions.

Keywords- *Model Reduction; Optimal Approximation; H_∞ Norm; Particle Swarm Optimization; Global Solution.*

I. INTRODUCTION

Most physical systems require complex high order mathematical models to well represent them. The complexity of these models imposes a lot of difficulties on simulation, analysis and control design.

Optimal model reduction reduces the order of the models representing the physical systems whilst ensuring high time and frequency response resemblance to the original high order models. Hence, reducing difficulties on simulation, analysis and control design.

Large number of techniques have been developed in the last few decades to approximate high order systems by low order models.

The model reduction problem is stated as follows: consider a general state space model representation of a single input single output (SISO) time-invariant linear continuous time system:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \quad (1)$$

where $x(t)$ is the state, $u(t)$ is the input, and $y(t)$ is the output. This state space model can be represented by the following n^{th} order transfer function:

$$G(s) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \dots + b_{n-1} s + b_n}{s^n + a_1 s^{n-1} + \dots + a_{n-1} s + a_n} \quad (2)$$

The aim of optimal model reduction is to obtain a reduced order state space model representation or a reduced order transfer function of the system that well represents the original system:

$$\begin{aligned} \dot{x}_r(t) &= A_r x_r(t) + B_r u(t) \\ y_r(t) &= C_r x_r(t) \end{aligned} \quad (3)$$

$$G_r(s) = \frac{d_1 s^{r-1} + d_2 s^{r-2} + \dots + d_{r-1} s + d_r}{s^r + c_1 s^{r-1} + \dots + c_{r-1} s + c_r} \quad (4)$$

In this paper, the H_{∞} Norm model reduction problem will be investigated using Genetic Algorithms (GA) and Particle Swarm Optimization (PSO).

Let $E(s) = G(s) - G_r(s)$, then the H_{∞} norm is defined as follows:

$$\begin{aligned} \|E\|_{\infty} &= \max |E(j\omega)| \\ &= \max \frac{\|E \cdot u\|_2}{\|u\|_2} \end{aligned} \quad (5)$$

Classical early model reduction techniques are summarized in [1]. The classical approach to model reduction dealt only with eigenvalues. However, in 1981, Moore published a paper presenting a revolutionary way of looking at model reduction by showing that the ideal platform to work from is that when all states are as controllable as they are observable [2]. This gave birth to “Balanced Model Reduction”, where the concept of dominance is no longer associated with eigenvalues, but rather with the degree of controllability and observability of a given state.

Moore’s approach aims at changing the form of the system’s state space model representation, by the use of a certain transformation matrix, into a balanced model with the transformed states being as controllable as they are observable, and ordered from strongly controllable and observable to weakly controllable and observable.

Since the output depends on both the controllability and the observability of a state, the states which are weakly controllable and observable will have little effect on the output, and thus, discarding them will not affect the output very much. This is what motivated Moore to develop his approach. Pernebo and Silverman [3] showed that the stability of this reduced model is assured if the original system was also stable. However, Moore's approach still suffered from steady state errors [1].

Hankel-norm reduction [4-7] on the other hand is optimal. It has a closed form solution and is computationally simple employing standard matrix software [1]. The singular values of the Hankel Matrix are called the Hankel Singular Values (HSV) of the system $G(z)$ and they are defined as follows [4-7]:

$$\sigma_i(G) = \lambda_i^{\frac{1}{2}}(PQ) \quad (6)$$

where P and Q are the controllability and observability Gramians respectively. The Hankel norm of a transfer function $G(z)$, denoted by $\|G\|_H$ is defined to be the largest HSV of $G(z)$ [4-7]:

$$\|G\|_H = \sigma_1 \quad (7)$$

The balanced model reduction realizations and the optimal Hankel-norm approximations changed the status of model reduction dramatically. Those two techniques made it possible to predict the error between the frequency responses of the full and the reduced order models [1].

Starting at 1992, Kavaranglu and Bettayeb [8-20] studied the H_∞ norm approximation of a given stable, proper, rational transfer function by a lower order stable, proper, rational transfer function. They found that the H_∞ norm model reduction problem can be converted into a Hankel norm model reduction problem, and therefore they based their approach on this finding [8].

A comparison between Hankel norm approximation and H_∞ norm model reduction in the H_∞ norm sense was conducted in [9]. Kavaranglu and Bettayeb found that the H_∞ approximation method can be much better or, in some cases, comparable to the Hankel norm approximation scheme.

Kavaranglu and Bettayeb then studied Hankel norm model reduction, and H_∞ approximation schemes where they explored some further properties related to the H_∞ norm [10]. In 1994, they presented a simple state-space suboptimal L_∞ norm Model reduction computational algorithm [11].

In 1995, Kavaranglu and Bettayeb developed a suboptimal computational scheme for the problem of constant L_∞ approximation of complex rational matrix functions, based on balanced realization for unstable systems [12]. They also derived an L_∞ error bound for unstable systems and obtained optimal solution for a class of symmetric systems [12].

In [13-15], they studied the L_∞ norm optimal simultaneous system approximation problem and explored various LMI based approaches to solve the simultaneous problem. On the other hand, L_∞ norm constant approximation of unstable systems was studied in [16].

Kavaranglu and Bettayeb [17] also presented an overview on H_∞ filtering, estimation, and deconvolution approaches, where they considered the problem of reduced order H_∞ estimation filter design. They then presented an iterative scheme for rational H_∞ approximation in 1995 [18].

Kavaranglu and Bettayeb also investigated L_∞ norm approximation of simultaneous multivariable systems by a rational matrix function with desired number of stable and unstable poles in [19].

A case study was presented in [20] where Sahin, Kavaranglu and Bettayeb applied four different model reduction schemes, namely, balanced truncation, singular perturbation balanced truncation, Hankel norm approximation, and H_∞ norm approximation; to a two-dimensional transient heat conduction problem.

Assunção et al. [21-22] addressed the H_∞ model reduction problem for uncertain discrete time systems with convex bounded uncertainties [21] and proposed a branch and bound algorithm to solve the H_2 norm model reduction problem for continuous time linear systems [22].

Ebihara et al. [23] noted that the lower bounds of the H_∞ Model Reduction problem can be analyzed by using Linear Matrix Inequality (LMI) related techniques, and thus, they reduce the order of the system by the multiplicity of the smallest Hankel Singular value which showed that the problem is essentially convex and the optimal reduced order models can be constructed via LMI optimization.

Wu et al. [24] investigated a frequency-weighted optimal H_∞ Model Reduction problem for linear time-invariant (LTI) systems. Their approach aimed to minimize the H_∞ norm of the frequency-weighted truncation error between a given LTI system and its lower order approximation. They proposed a model reduction scheme based on Cone Complementarity Algorithm (CCA) to solve their H_∞ Model Reduction problem.

In 2005, Xu et al. [25] studied H_∞ Model Reduction for 2-D Singular Roesser Models. However more recently in 2008 and 2009, Zhang et al. investigated the H_∞ Model Reduction problem for [26] a class of discrete-time Markov jump linear systems (MJLS) with partially known transition probabilities and for [27] switched linear discrete-time systems with polytopic uncertainties.

The main problem with the above analytical optimization techniques is that they result in non-linear equations in the parameters of the reduced order model. In order to solve those non-linear equations, one will have to go through computationally demanding iterative minimization algorithms, that suffer from many problems such as the choice of starting guesses, convergence, and multiple local minima, not to mention the huge amount of time it demand to reach a solution [1].

II. PARTICLE SWARM OPTIMIZATION

Researchers observed that some living creatures tend to perform swarming behavior. Examples of swarms include flocks of birds, schools of fish, herds of animals, and colonies

of bacteria. Such a corporative behavior has certain advantages as avoiding predators and increasing the chance of finding food, but it requires communication and coordinated decision making [28-31].

Particle Swarm Optimization simulates the behavior of bird flocking. When a group of birds are randomly searching food in an area, that has only one piece of food, all birds have no idea where the food is, but rather know how far the food is in each iteration; and thus tend to follow the bird that is nearest to the food.

Similarly, in PSO, each single solution is a particle (bird) in the search space. All particles have fitness values evaluated by the fitness function to be optimized, and have velocities which direct the flying of the particles.

The PSO algorithm is simple in concept, easy to implement and computationally efficient. The procedure for implementing a PSO is as follows [30]:

1. Initialize a population of particles with random positions and velocities on D dimensions in the problem space.
2. For each particle, evaluate the desired optimization fitness function in D variables.
3. Compare particles fitness evaluation with *pbest*. If current value is better than *pbest*, then set *pbest* equal to the current value, and p_i equals to the current position x_i in D-dimensional space. Where *pbest* is the best fitness value a particle has achieved so far.
4. Identify the particle in the neighborhood with the best success so far, and assign its position to the variable *G* and its fitness value to variable *gbest*.
5. Change the velocity and position of the particle according to the bellow equations [31]:

$$v_i(k+1) = \emptyset v_i(k) + c_1 \gamma_{1i} (p_i - x_i(k)) + c_2 \gamma_{2i} (G - x_i(k)) \quad (8)$$

$$x_i(k+1) = x_i(k) + v_i(k+1) \quad (9)$$

where:

- i is the particle index
- k is the discrete time index
- v is the velocity of the i^{th} particle
- x is the position of the i^{th} particle
- p is the best position found by the i^{th} particle (personal best)
- G is the best position found by swarm (global best, best of personal bests)
- γ_{1i} & γ_{2i} are random numbers on the interval [0, 1] applied to the i^{th} particle
- \emptyset is the inertial weight function
- c_1 & c_2 are acceleration constants

6. Loop to step 2 until a criterion is met, usually a sufficiently good fitness or a maximum number of iterations.

A decreasing inertial weight \emptyset of the following form is used in the PSO approach:

$$\emptyset = w_i + (w_f - w_i) \frac{k-1}{N-1} \quad (10)$$

where w_i and w_f are the initial and final inertial weights respectively, k is the current iteration (epoch) and N is the iteration (epoch) when the inertial weight should reach its final value [31]. The decreasing inertial weight is known to improve the PSO performance.

III. PREVIOUS STUDIES

Model reduction has caught the attention of many researchers in the past few decades. However, most of the existing work relies on tedious analytical solution methods. Minimal work has been done on H_∞ model reduction using Genetic Algorithms whereas no work at all has been done on H_∞ model reduction using Particle Swarm Optimization.

Tan et al. [32] developed a Boltzmann learning enhanced GA based method to solve L_∞ identification and model reduction problems, and obtain a globally optimized nominal model and an error bounding function for additive and multiplicative uncertainties. They used their GA to identify 2nd and 3rd order discrete nominal models for a 4th order discrete plant of an industrial heat exchanger. Comparing the frequency responses of the original plant with the two GA defined models; the GA results were proven to give a good fitting over the frequency range concerned and to outperform other techniques yielding the smallest L_∞ norm errors

It is the intent of our work to perform a comprehensive evaluation of GA and PSO for H_∞ optimal model reduction using several benchmark model reduction examples. Both time domain and frequency domain performances are considered in our work.

IV. SIMULATION AND RESULTS

The H_∞ norm has a finite lower bound. Consider the Hankel singular values of system $G(s)$ defined in eq. (6). In H_∞ Model Reduction [11], if the n^{th} order transfer function $G(s)$ is reduced into an r^{th} order transfer function $G_r(s)$, then:

$$\sigma_{r+1} \leq \|G - G_r\|_\infty \quad (11)$$

where σ_{r+1} is the $(r+1)^{st}$ HSV of $G(s)$. Thus, the H_∞ norm of the reduced order model can never be lower than the highest hankel singular value dropped.

However, it is convenient to mention here that the H_∞ lower bound is almost impossible to achieve; but if one results with a close enough H_∞ norm, this would be a good indication that an optimal solution was reached.

The first model to be reduced is the 4th order Wilson example [33]:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 0 & 0 & -150 \\ 1 & 0 & 0 & -245 \\ 0 & 1 & 0 & -113 \\ 0 & 0 & 1 & -19 \end{bmatrix} x(t) + \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix} u(t) \\ y(t) &= [0 \ 0 \ 0 \ 1] \end{aligned} \quad (12)$$

Hankel Singular Values:

$$\sigma_1 = 0.015938, \sigma_2 = 0.002724, \sigma_3 = 0.000127, \sigma_4 = 0.000008$$

Eigen Values: $[-1, -3, -5, -10]$

This system was reduced into a 2nd order model (lower bound Hankel singular value = 1.27×10^{-4}) using the following PSO settings:

$$\begin{aligned} \text{Swarm size} &= 100 \\ \text{Maximum velocity} &= 2 \\ \text{Acceleration Const. } c_1 &= 2 \\ \text{Acceleration Const. } c_2 &= 2 \\ \text{Initial inertia weight} &= 0.9 \\ \text{Final inertia weight} &= 0.1 \\ \text{Epoch when inertial weight at final value} &= 1000 \end{aligned}$$

and resulted with the following reduced order model:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} -0.9233 & 0.2375 \\ -1.117 & -2.766 \end{bmatrix} x(t) + \begin{bmatrix} 0.5446 \\ 1.087 \end{bmatrix} u(t) \\ y(t) &= [0.04847 \ -0.02771] x(t) \end{aligned} \quad (13)$$

Eigen Values: $[-1.0807, -2.6086]$

The following model represents the result of the H_∞ 2nd order model reduction using Genetic Algorithms (GA) obtained in [34]:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} -3.457 & -6.049 \\ 0.3239 & -0.2429 \end{bmatrix} x(t) + \begin{bmatrix} -0.2793 \\ 0.01254 \end{bmatrix} u(t) \\ y(t) &= [-0.05126 \ -1.428] x(t) \end{aligned} \quad (14)$$

Eigen Values: $[-1.0604, -2.6395]$

Table 1 compares the steady state errors and norms of the resulting 2nd order model to that reduced with H_∞ Model Reduction using GA in [34]. Figures 1 to 4 show the impulse responses, initial values, step responses and frequency responses of all models respectively.

TABLE I. WILSON SYSTEM SSE & NORMS OF REDUCED MODELS

	SS Error	H_∞ Norm
GA H_∞ model reduction [34]	2.144×10^{-6}	2.239×10^{-4}
PSO H_∞ model reduction	2.144×10^{-4}	2.144×10^{-4}

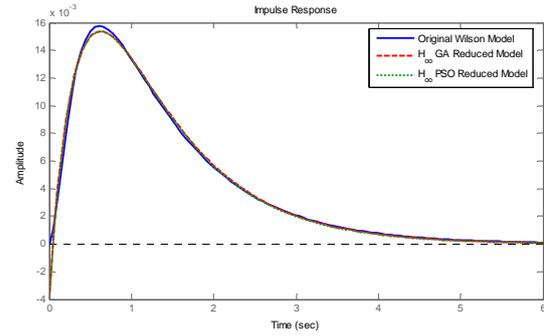


Figure 1. Impulse Response.

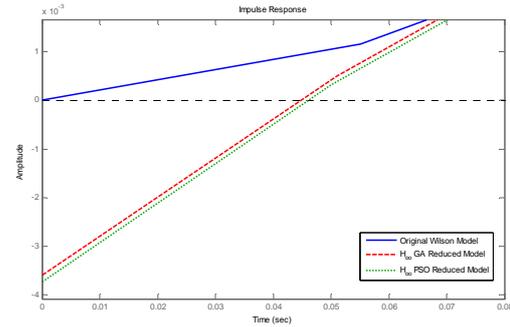


Figure 2. Initial Values.

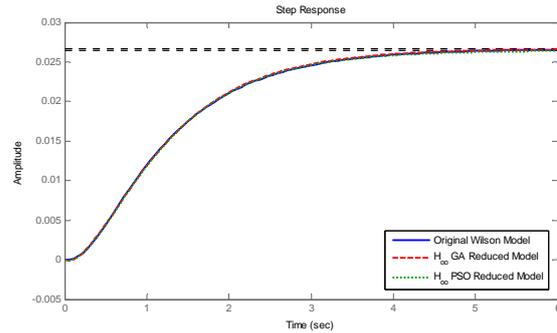


Figure 3. Step Response.

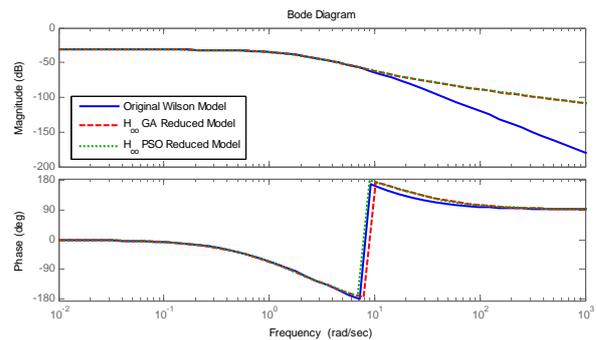


Figure 4. Frequency Response.

Note that the impulse responses and the step responses of the reduced order model well resemble the original system. However, due to the two missing states in the reduced order

models, they fail to resemble the frequency response of the original system in high frequencies. This tends to be acceptable since many practical systems operate in low frequencies.

The second system to be reduced is a 9th order Boiler System [35]:

$$A = \begin{bmatrix} -0.910 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -4.449 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -10.262 & 571.479 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -10.987 & 0 & 0 & 0 & 0 \\ 11.622 & 0 & 0 & 0 & 0 & -11.622 & -15.214 & 0 & 0 \\ 0 & 0 & 0 & 0 & -89.874 & 0 & 0 & 0 & -502.665 \end{bmatrix}$$

$$B = [-4.336; -3.691; 10.141; -1.612; 16.629; -242.476; -14.261; 13.672; 82.187]$$

$$C = [-0.422 \quad -0.736 \quad -0.00416 \quad 0.232 \quad -0.816 \quad -0.715 \quad 0.546 \quad -0.235 \quad -0.0806] \quad (15)$$

Hankel Singular Values:

$$\sigma_1 = 6.2115 \quad \sigma_2 = 0.8264 \quad \sigma_3 = 0.6770 \quad \sigma_4 = 0.0593$$

$$\sigma_5 = 0.0568 \quad \sigma_6 = 0.0188 \quad \sigma_7 = 0.0096 \quad \sigma_8 = 0.0031$$

$$\sigma_9 = 0.0007$$

Eigen Values: $[-0.91, -4.45, -10.26 \pm j571.48, -10.99, -15.21 \pm j11.62, -89.87, -502.67]$

This system was reduced into a third order system (lower bound Hankel singular value = 5.93×10^{-2}) using the same PSO settings, except the epoch when inertial weight at final value was set to 6000. The following model represents the result of 3rd order model reduction using PSO:

$$\dot{x}(t) = \begin{bmatrix} -24.01 & -13.71 & 26.67 \\ 22.65 & 10.85 & -22.19 \\ 2.5 & 9.653 & -22.01 \end{bmatrix} x(t) + \begin{bmatrix} 13.18 \\ -11.6 \\ 1.477 \end{bmatrix} u(t) \quad (16)$$

$$y(t) = [18.29 \quad 6.884 \quad -10.24] x(t)$$

Eigen Values: $[-0.8192, -17.1804 \pm j12.8069]$

and the following model represents the result of the GA approach of [34]:

$$\dot{x}(t) = \begin{bmatrix} -6.449 & -29.73 & 0.3856 \\ 3.129 & -21.71 & 10.69 \\ -4.918 & -6.326 & -6.661 \end{bmatrix} x(t) + \begin{bmatrix} 4.921 \\ -4.127 \\ 10.74 \end{bmatrix} u(t) \quad (17)$$

$$y(t) = [6.591 \quad 0.1677 \quad 10.71] x(t)$$

Eigen Values: $[-1.0423, -16.8889 \pm j13.0262]$

Table 2 and figures 5 to 8 compare the steady state errors, norms, impulse responses, initial values, step responses and frequency responses the resulting 3rd order model with the GA reduced order model of [34].

TABLE II. BOILER SYSTEM SSE & NORMS OF REDUCED MODELS

	SS Error	H _∞ Norm
GA H _∞ model reduction [34]	1.126×10^{-1}	1.127×10^{-1}
PSO H _∞ model reduction	1.275×10^{-3}	1.116×10^{-1}

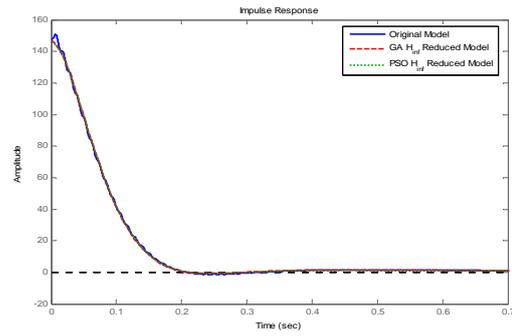


Figure 5. Impulse Response

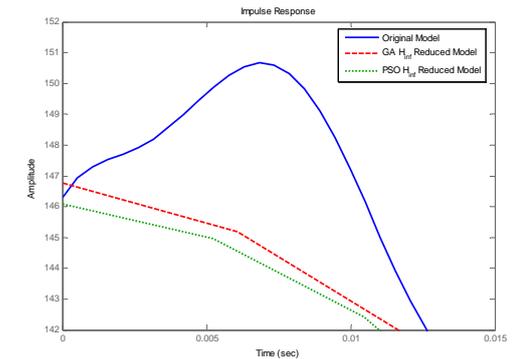


Figure 6. Initial Values

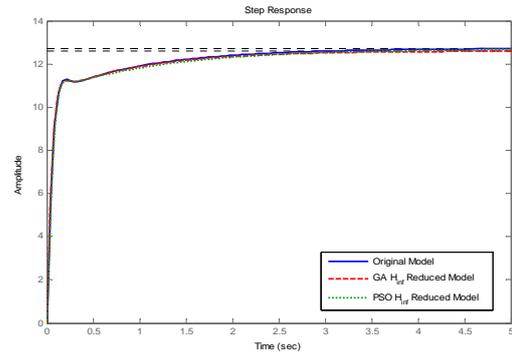


Figure 7. Step Response

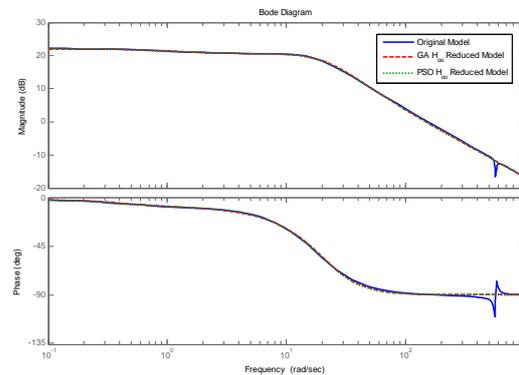


Figure 8. Frequency Response

Note that the impulse, step and frequency responses of the reduced order model well resemble the original system.

CONCLUSION

The use of Evolutionary Algorithms in optimal model reduction, such as Particle Swarm Optimization, provide robust search in complex spaces and helps solve complex optimal model reduction problems without the need to go through computationally demanding iterative non-linear mathematical procedures.

The PSO approach of this paper was compared to the GA approach of [34] and it was noted that the PSO outperforms the GA by resulting in better norms. Also, the simplicity of the computations in the PSO Algorithm in comparison to the GA Algorithm makes it much faster.

REFERENCES

- [1] U.M. Al-Saggaf, M. Bettayeb, "Techniques in Optimized Model Reduction for High Dimensional Systems", *Control and Dynamic Systems*, vol. 55, pp. 51-109, 1993.
- [2] B.C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction", *IEEE Trans. Automatic Control*, vol. AC-26, pp. 17-32, 1981.
- [3] L. Pernebo, L.M. Silverman, "Model Reduction via Balanced State Space Representation", *IEEE Trans. on Automatic Control*, vol. AC-27, no. 2, pp. 382-387, 1982.
- [4] L.M. Silverman, M. Bettayeb, "Optimal Approximation of Linear Systems", in *JACC*, vol. 2, p. FA 8-A, 1980.
- [5] M. Bettayeb, L.M. Silverman, M.G. Safonov, "Optimal Approximation of Continuous Time Systems", *Proc. 19th IEEE CDC, Albuquerque, New Mexico, 1980*.
- [6] M. Bettayeb, "Approximation of Linear Systems: New Approaches based on Singular Value Decomposition", *Ph.D. Thesis, University of Southern California, Los Angeles, 1981*.
- [7] K. Glover, "All Optimal Hankel Norm Approximations of Linear Multivariable systems and their L_∞ -error Bounds", *Int. J. Contr.*, vol. AC-27, pp. 1115-1193, 1984.
- [8] D. Kavranoglu, M. Bettayeb, "Characterization of the Solution to the Optimal H_∞ Model Reduction Problem", *Systems and Control Letters*, vol. 20, pp. 99-107, 1993.
- [9] M. Bettayeb and D. Kavranoglu, "Performance Evaluation of A New H_∞ Norm Model Reduction Scheme", *Proc. 1993 IEEE CDC, San Antonio, USA*, pp. 2913-2914, 1993.
- [10] D. Kavranoglu and M. Bettayeb, "Systems theory Properties and Approximation of Relaxation Systems", *Theme Issue "Control Theory and Its Applications"*, *Arab J. Science and Engg.*, vol. 18, no. 4, pp. 479-491, October, 1993.
- [11] D. Kavranoglu, and M. Bettayeb, "Characterization and Computation of the Solution to the Optimal L_∞ Approximation Problem", *IEEE Tran. Auto. Cont.*, vol. 39, pp. 1899-1904, September 1994.
- [12] D. Kavranoglu, and M. Bettayeb, "Constant L_∞ Norm Approximation of Complex Rational Functions", *Numerical Func. Analy. Opt.: An International Journal*, vol. 16, no. 1 & 2, pp. 197 - 217, 1995.
- [13] D. Kavranoglu, M. Bettayeb and M.F. Anjum, " L_∞ Norm Simultaneous Model Approximation", in *Special issue on LMIs in Systems and Control, International J. of Robust and Nonlinear Control*, vol. 6, pp. 999-104, November 1996.
- [14] D. Kavranoglu and M. Bettayeb, "Weighted L_∞ Norm Approximation with Given Poles Using LMI Techniques and Applications", *Proceedings of European Control Conference, Rome, Italy*, vol. 1, pp. 567-571, September 1995.
- [15] D. Kavranoglu and M. Bettayeb, "LMI Based Computational Schemes for H_∞ Model Reduction", *13th World Congress, IFAC, San Francisco, California, USA, June 1996*.
- [16] D. Kavranoglu, and M. Bettayeb, " L_∞ Norm Constant Approximation of Unstable Systems" *Proc. 1993 ACC, San Francisco, USA*, pp. 2142-2148, 1993.
- [17] M. Bettayeb and D. Kavranoglu, "Reduced Order H_∞ Filtering", *Proc. ACC 94, Baltimore*, pp. 1884-1888, June 1994.
- [18] M. Bettayeb and D. Kavranoglu, "An Iterative Scheme for Rational H_∞ Approximation", *Proceedings of ECCTD-95, Istanbul, Turkey*, pp. 905-908, August 1995.
- [19] D. Kavranoglu and M. Bettayeb, F. Anjum "Rational L_∞ Norm Approximation of Multivariable Systems" *Proceedings of the 34th IEEE CDC, New Orleans, Louisiana*, pp. 790-795, December 1995.
- [20] A.Z. Sahin, D. Kavranoglu, M. Bettayeb, "Model Reduction in Numerical Heat Transfer Problems", *Applied Mathematics and Computations*, vol. 69, pp. 209-225, 1995.
- [21] E. Assunção, P.L.D. Peres, " H_2 and/or H_∞ -Norm Model Reduction of Uncertain Discrete-time Systems", *Proceedings of the American Control Conference, San Diego, California, USA*, pp. 4466-4470, June 1999.
- [22] E. Assunção, P.L.D. Peres, "A Global Optimization Approach for the H_2 -Norm Model Reduction Problem", *Proceedings of the 38th Conference on Decision and Control, Phoenix, Arizona, USA*, pp. 1857-1862, December 1999.
- [23] Y. Ebihara, T. Hagiwara, "On H_∞ Model Reduction Using LMIs", *IEEE Transaction on Automatic Control*, vol. 49, no. 7, pp. 1187-1191, July 2004.
- [24] F. Wu, J. J. Jaramillo, "Computationally Efficient Algorithm for Frequency-Weighted Optimal H_∞ Model Reduction", *Asian Journal of Control*, vol. 5, no. 3, pp. 341-349, September 2003.
- [25] Huiling Xu, Yun Zou, Shengyuan Xu, James Lam, Qing Wang, " H_∞ Model Reduction of 2-D Singular Roesser Models", *Multidimensional Systems and Signal Processing*, vol. 16, no. 3, pp. 285-304, July 2005.
- [26] Lixian Zhang, Peng Shi, El-Kébir Boukas, Changhong Wang, " H_∞ model reduction for uncertain switched linear discrete-time systems", *Technical Communique, Elsevier Ltd.*, October 2008.
- [27] Lixian Zhang, El-Kébir Boukas, Peng Shi, " H_∞ model reduction for discrete-time Markov jump linear systems with partially known transition probabilities", *International Journal of Control*, vol. 82, no. 2, pp. 343-351, February 2009.
- [28] R. Eberhart, J. Kennedy, "A New Optimizer Using Particle Swarm Theory", *Sixth Intl. Symposium on Micro Machine and Human Science, Nagoya, Japan*, pp. 39-43, Oct. 1995.
- [29] J. Kennedy, R. Eberhart, "Particle Swarm Optimization", *Proc. IEE International Conference on Neural Networks*, pp. 1942-1947, 1995.
- [30] Y. Shi, "Particle Swarm Optimization", *IEEE Neural Networks Society*, pp. 8-13, Feb. 2004.
- [31] B. Birge, "PSOt – A Particle Swarm Optimization Toolbox for Use with MATLAB", *IEEE Swarm Intelligence Symposium Proceedings*, pp. 182-186, April, 2003.
- [32] K.C. Tan and Y. Li, " L_∞ Identification and Model Reduction Using a Learning Genetic Algorithm", *Proceedings of UKACC International Conference on Control 96, Exeter, UK, (Conf. Publ. No. 427)*, vol. 2, pp. 1125-1130, Sept. 1996.
- [33] D.A. Wilson, "Optimal Solution of Model Reduction Problem", *Proc. IEE*, vol. 117, 1970.
- [34] R. Salim and M. Bettayeb, " H_2 and H_∞ Optimal Model Reduction Using Genetic Algorithms", *Proceedings of the 3rd International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO'09), January 20 – 22, 2009, AUS, Sharjah, UAE*.
- [35] G. Zhao, N.K. Sinha, "Model selection in Aggregated Models", *Large Scale Systems*, P. 209-216, 1983.

H₂ Optimal Model Reduction of Dynamic Systems With Time-Delay Using Particle Swarm Optimization

Reem Salim

Electrical & Computer Engineering Department
University of Sharjah
Sharjah, United Arab Emirates
reemsalim@gmail.com

Maamar Bettayeb

Electrical & Computer Engineering Department
University of Sharjah
Sharjah, United Arab Emirates
maamar@sharjah.ac.ae

Abstract—The use of reduced order models of systems makes it easier to implement analysis, simulations and control designs. Several physical systems, including transportation systems, chemical and heat processes, computer and communication networks, have inherent time-delay and high order dynamics. The reduced order time-delay model is generally determined by minimizing a nonlinear quadratic cost function that has multiple local minima. Therefore obtaining global minima using nonlinear optimization techniques is very difficult.

To overcome these difficulties, we use Particle Swarm Optimization (PSO) to solve H₂ model reduction problems for high order dynamical systems with time-delay.

The performance of PSO is demonstrated with application examples.

Keywords- Model Reduction; Optimal Approximation; H₂ Norm; Time-Delay; Particle Swarm Optimization; Global Solution.

I. INTRODUCTION

Most physical systems are complex and their transfer function representations are of very high orders. In order to perform simulation, analysis or control design on those high order models, one will face many difficulties.

Therefore it is desirable to represent those physical systems with reduced order models that well resemble the original model in time and frequency domain.

However, many of those physical systems have inherently pure time delays. And in order to ensure high resemblance of the reduced order models to the original system we introduce time delay into the reduced order models.

The model reduction problem is stated as follows: consider a general nth order transfer function representation of a single input single output (SISO) time-invariant linear continuous time system:

$$G(s) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \dots + b_{n-1} s + b_n}{s^n + a_1 s^{n-1} + \dots + a_{n-1} s + a_n} \quad (1)$$

The aim of optimal model reduction with time delay is to obtain a reduced order transfer function representation of the

system that well represents the original system of the following form:

$$G_r(s) = \frac{d_1 s^{r-1} + d_2 s^{r-2} + \dots + d_{r-1} s + d_r}{s^r + c_1 s^{r-1} + \dots + c_{r-1} s + c_r} e^{-\tau s} \quad (2)$$

In this paper, the H₂ Norm model reduction problem with time-delay will be investigated using and Particle Swarm Optimization (PSO).

Let E(s) = G(s) – G_r(s), then the H₂ norm is defined as follows:

$$\|E\|_2 = \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} |E(j\omega)|^2 d\omega \right\}^{\frac{1}{2}} \quad (3)$$

The H₂ norm can also be represented in time domain as follows:

$$\|e\|_2 = \left\{ \int_0^{\infty} |y(\tau) - y_r(\tau)|^2 d\tau \right\}^{\frac{1}{2}} \quad (4)$$

where y(τ) and y_r(τ) are the impulse responses of the original model and the reduced order model respectively.

Classical early model reduction techniques are summarized in [1]. The classical approach to model reduction dealt only with eigenvalues. However, in 1981, Moore published a paper presenting a revolutionary way of looking at model reduction by showing that the ideal platform to work from is that when all states are as controllable as they are observable [2]. This gave birth to “Balanced Model Reduction”, where the concept of dominance is no longer associated with eigenvalues, but rather with the degree of controllability and observability of a given state. However, Moore’s approach still suffered from steady state errors [1].

Hankel-norm reduction [3-6] on the other hand is optimal. It has a closed form solution and is computationally simple employing standard matrix software [1].

The balanced model reduction realizations and the optimal Hankel-norm approximations changed the status of model reduction dramatically. Those two techniques made it possible to predict the error between the frequency responses of the full and the reduced order models [1].

Modern H₂ model reduction was studied in [7-14], for earlier classical least squares model reduction, see ([1] and

references therein). Hyland et al. [7] used optimal projection to derive H_2 reduced models. Yan et al. [8] proposed an H_2 optimal model reduction approach that uses orthogonal projections to reduce the H_2 cost over the stiefel manifold so that the stability of the reduced order model is assured.

Assunção et al. [9-10] addressed the H_2 and H_∞ model reduction problems for uncertain discrete time systems with convex bounded uncertainties [9] and proposed a branch and bound algorithm to solve the H_2 norm model reduction problem for continuous time linear systems [10].

In 2001, Huang et al. [11] proposed a globally convergent H_2 model reduction algorithm in the form of an ordinary differential equation. Then, in 2002, Marmorat et al. [12] proposed an H_2 approximation approach using shur parameters. An H_2 optimal model reduction case study is given in [13].

More recently, in 2007, Beattie et al. [14] proposed an H_2 model reduction technique, based on Krylov method, suitable for dynamical systems with large dimension.

Model Reduction with Time-Delay was studied in [15] where a procedure for approximating a high-order system with a rational transfer function by a low-order system with a rational transfer function together with a pure time delay. The procedure introduced a delay into the system output and computed the low-order transfer function using truncation of a certain balanced realization.

The main problem with the above analytical optimization techniques is that they result in non-linear equations in the parameters of the reduced order model. In order to solve those non-linear equations, one will have to go through computationally demanding iterative minimization algorithms, that suffer from many problems such as the choice of starting guesses, convergence, and multiple local minima, not to mention the huge amount of time it demand to reach a solution [1].

II. PARTICLE SWARM OPTIMIZATION

Researchers observed that some living creatures tend to perform swarming behavior. Examples of swarms include flocks of birds, schools of fish, herds of animals, and colonies of bacteria. Such a corporative behavior has certain advantages as avoiding predators and increasing the chance of finding food, but it requires communication and coordinated decision making [16-25].

Particle Swarm Optimization simulates the behavior of bird flocking. When a group of birds are randomly searching food in an area, that has only one piece of food, all birds have no idea where the food is, but rather know how far the food is in each iteration; and thus tend to follow the bird that is nearest to the food.

Similarly, in PSO, each single solution is a particle (bird) in the search space. All particles have fitness values evaluated by the fitness function to be optimized, and have velocities which direct the flying of the particles.

The PSO algorithm is simple in concept, easy to implement and computationally efficient. The procedure for implementing a PSO is as follows [16]:

1. Initialize a population of particles with random positions and velocities on D dimensions in the problem space.
2. For each particle, evaluate the desired optimization fitness function in D variables.
3. Compare particles fitness evaluation with $pbest$. If current value is better than $pbest$, then set $pbest$ equal to the current value, and p_i equals to the current position x_i in D-dimensional space. Where $pbest$ is the best fitness value a particle has achieved so far.
4. Identify the particle in the neighborhood with the best success so far, and assign its position to the variable G and its fitness value to variable $gbest$.
5. Change the velocity and position of the particle according to the bellow equations [25]:

$$\begin{aligned} v_i(k+1) &= \emptyset v_i(k) + c_1 \gamma_{1i} (p_i - x_i(k)) \\ &\quad + c_2 \gamma_{2i} (G - x_i(k)) \\ x_i(k+1) &= x_i(k) + v_i(k+1) \end{aligned} \quad (5)$$

where:

i	is the particle index
k	is the discrete time index
v	is the velocity of the i^{th} particle
x	is the position of the i^{th} particle
p	is the best position found by the i^{th} particle (personal best)
G	is the best position found by swarm (global best, best of personal bests)
γ_{1i} & γ_{2i}	are random numbers on the interval $[0, 1]$ applied to the i^{th} particle
\emptyset	is the inertial weight function
c_1 & c_2	are acceleration constants

6. Loop to step 2 until a criterion is met, usually a sufficiently good fitness or a maximum number of iterations.

A decreasing inertial weight \emptyset of the following form is used in the PSO approach:

$$\emptyset = w_i + (w_f - w_i) \frac{k-1}{N-1} \quad (7)$$

where w_i and w_f are the initial and final inertial weights respectively, k is the current iteration (epoch) and N is the iteration (epoch) when the inertial weight should reach its final value [25]. The decreasing inertial weight is known to improve the PSO performance.

III. PREVIOUS STUDIES

Model reduction has caught the attention of many researchers in the past few decades. However, most of the existing work relies on tedious analytical solution methods. Some work has been done on H_2 model reduction with time-

delay using Genetic Algorithms whereas no work at all has been done on H_2 model reduction with time-delay using Particle Swarm Optimization.

Yang et al. [26] proposed a GA based H_2 Norm model reduction approach for SISO continuous time systems that introduces time delay into the reduced order model.

Given an n^{th} order SISO time delay system with transfer function:

$$G(s) = \frac{\beta_1 s^{m-1} + \beta_2 s^{m-2} + \dots + \beta_m}{s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \dots + \alpha_n} \exp(-\tau s) \quad (8)$$

where its rational part is stable and strictly proper, they tried to find a strictly proper l^{th} order reduced model with the time delay:

$$\tilde{G}(s) = \begin{cases} \frac{d_1 s^{k-1} + d_2 s^{k-2} + \dots + d_k}{\prod_{i=1}^{\lfloor \frac{l}{2} \rfloor} (s^2 + a_i s + b_i)} \exp(-Ts) & (l \text{ even}) \\ \frac{d_1 s^{k-1} + d_2 s^{k-2} + \dots + d_k}{\prod_{i=1}^{\lfloor \frac{l}{2} \rfloor} (s^2 + a_i s + b_i) \cdot (s+c)} \exp(-Ts) & (l \text{ odd}) \end{cases} \quad (9)$$

Their cost function was defined as:

$$V = \int_{-\infty}^{+\infty} |G(j\omega) - \tilde{G}(j\omega)|^2 W(j\omega) d\omega \quad (10)$$

where $W(j\omega)$ is a frequency weighing function introduced to obtain better approximation over a pre-specified frequency range.

IV. SIMULATION AND RESULTS

Yang et al. [26] used the following 6th order academic example from [15] to test their GA approach:

$$G(s) = \frac{s^4 - 20s^3 + 180s^2 - 840s + 1680}{s^6 + 22.5s^5 + 231s^4 + 1310s^3 + 3960s^2 + 5040s + 1680} \quad (11)$$

Yang, Hachino and Tsuji reduced the above system into 1st order, 2nd order, 3rd order, and 4th order systems with time delay, and resulted with the following transfer functions respectively:

The 1st order reduced order model:

$$\tilde{G}_1(s) = e^{-1.2654s} * \frac{0.37279}{s+0.36011} \quad (12)$$

The 2nd order reduced order model:

$$\tilde{G}_2(s) = e^{-0.96755s} * \frac{-0.00029434s + 0.91327}{s^2 + 2.2827s + 0.91490} \quad (13)$$

The 3rd order reduced order model:

$$\tilde{G}_3(s) = e^{-0.666144s} * \frac{0.000016679s^2 - 0.73837s + 4.5827}{(s^2 + 3.5884s + 1.5133)(s + 3.0262)} \quad (14)$$

The 4th order reduced order model:

$$\tilde{G}_4(s) = e^{-0.34760s} * \frac{-0.0016844s^3 + 0.87950s^2 - 6.8437s + 21.542}{(s^2 + 2.3031s + 0.90710)(s^2 + 6.8039s + 23.716)} \quad (15)$$

Using PSO with the following settings we obtained the 1st order, 2nd order, 3rd order, and 4th order reduced order models of the same 6th order system:

Swarm size	=	100
Maximum Velocity	=	4
Acceleration Const. c_1	=	2
Acceleration Const. c_2	=	2
Initial inertia weight	=	0.9
Final inertia weight	=	0.1
Epoch when inertial weight at final value	=	10,000

The H_2 fitness function defined in (4) was computed using trapezoidal integration to perform the reduction in PSO.

The 1st order reduced order model:

$$\tilde{G}_1(s) = e^{-1.28s} * \frac{0.3459}{s+0.3045} \quad (16)$$

The 2nd order reduced order model:

$$\tilde{G}_2(s) = e^{-s} * \frac{-0.02225s + 0.9308}{s^2 + 2.3214s + 0.9363} \quad (17)$$

The 3rd order reduced order model:

$$\tilde{G}_3(s) = e^{-0.67s} * \frac{-0.01081s^2 - 0.6451s + 4.37}{s^3 + 6.32s^2 + 11.88s + 4.339} \quad (18)$$

The 4th order reduced order model:

$$\tilde{G}_4(s) = e^{-0.35s} * \frac{-0.007101s^3 + 0.7993s^2 - 6.462s + 20.4}{s^4 + 8.551s^3 + 38.48s^2 + 57.38s + 20.5} \quad (19)$$

Table I compares the H_2 Norms of the above eight reduced order models:

TABLE I. H_2 NORM OF REDUCED MODELS

	1 st order	2 nd order	3 rd order	4 th order
Yang et al.'s GA Approach	1.0330×10^{-1}	1.8286×10^{-2}	1.3084×10^{-2}	8.5880×10^{-3}
Proposed PSO Approach	9.9326×10^{-2}	1.8094×10^{-2}	1.2937×10^{-2}	8.2235×10^{-3}

Table I proves that the proposed PSO approach yields better results than Yang et al.'s proposed GA approach. Figures 1 to 12 show the impulse responses, step responses, and frequency responses of the eight reduced order models in comparison to the original 6th order system.

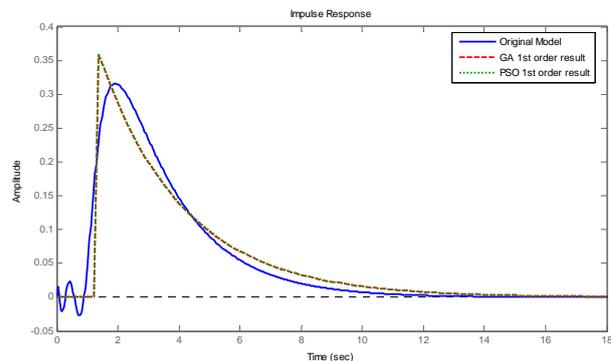


Figure 1. 1st order ROMs Impulse Responses.

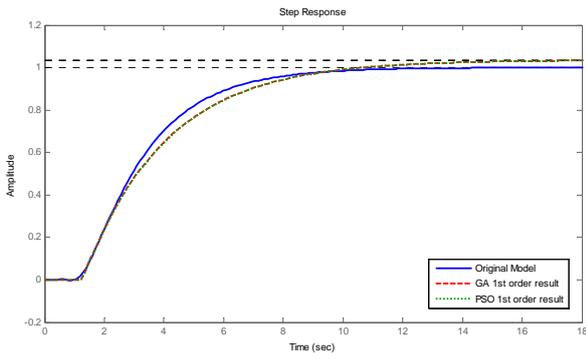


Figure 2. 1st order ROMs Step Responses.

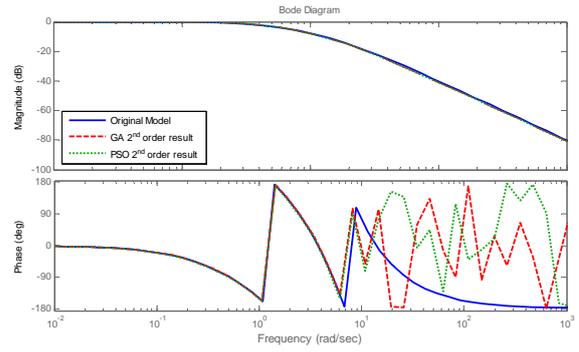


Figure 6. 2nd order ROMs Frequency Responses.

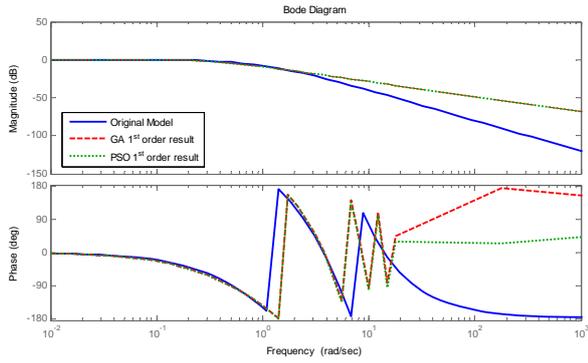


Figure 3. 1st order ROMs Frequency Responses.

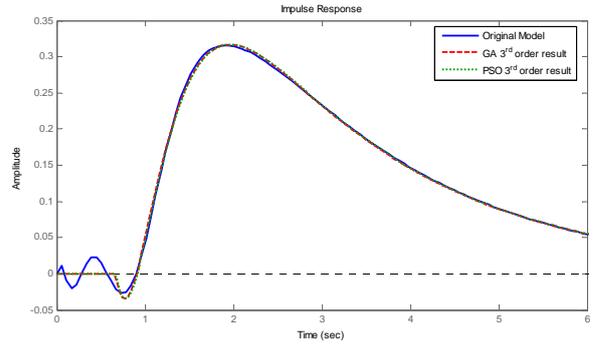


Figure 7. 3rd order ROMs Impulse Responses.

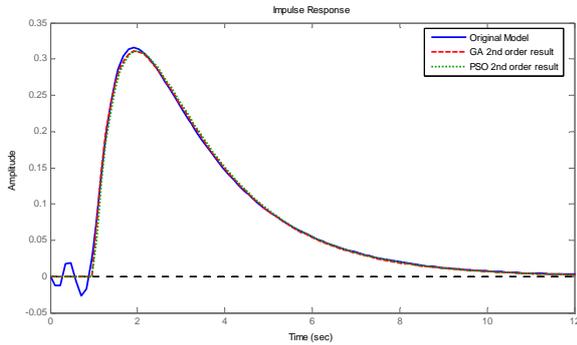


Figure 4. 2nd order ROMs Impulse Responses.

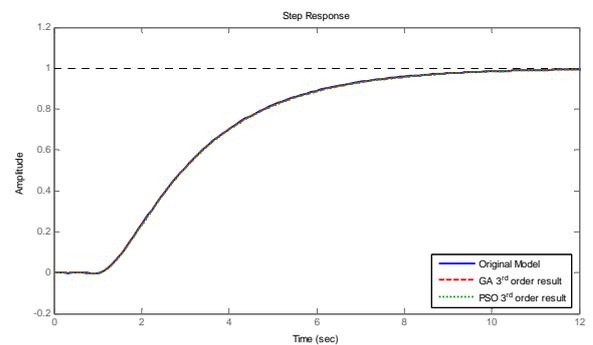


Figure 8. 3rd order ROMs Step Responses.

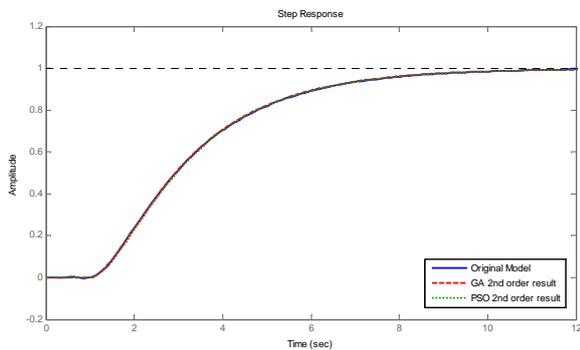


Figure 5. 2nd order ROMs Step Responses.

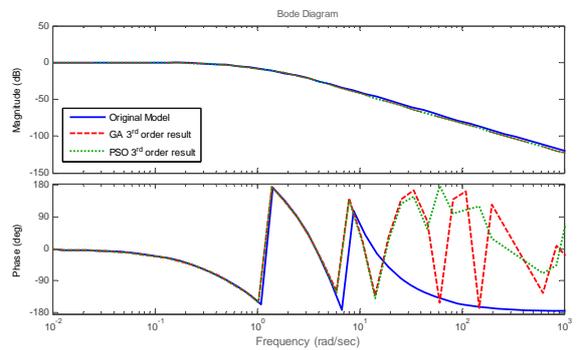
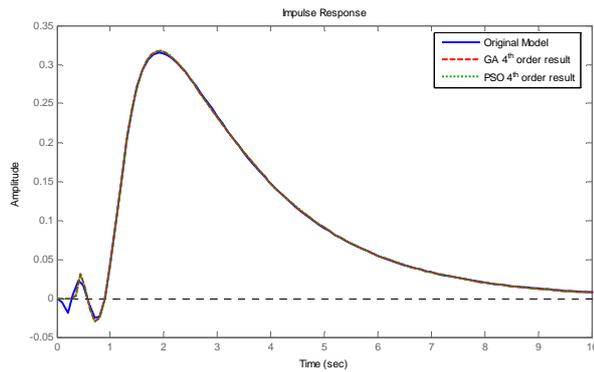
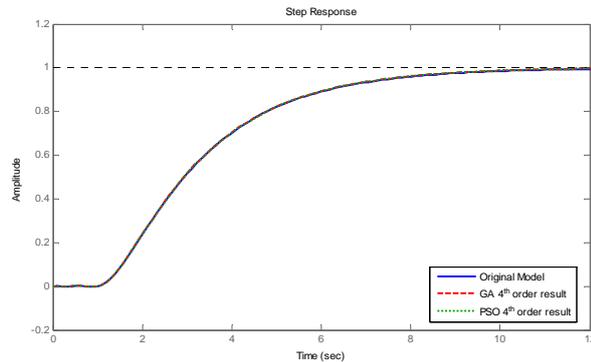
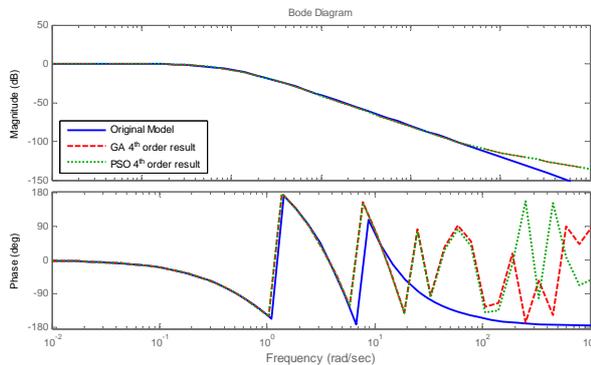


Figure 9. 3rd order ROMs Frequency Responses.

Figure 10. 4th order ROMs Impulse Responses.Figure 11. 4th order ROMs Step Responses.Figure 12. 4th order ROMs Frequency Responses.

From Figures 1 to 12, it is noted that the 4th order reduced model provides the best approximation to the original system amongst all the reduced order models. It is also shown in figure 3 from the Bode plot of the 1st order reduced model that the approximating error is quite large at low frequencies.

CONCLUSION

Particle Swarm Optimization proved to be a very strong optimization tool in the past few decades. In this paper PSO was used to find optimum reduced models for complex high

order SISO models using H_2 norm with time-delay. The PSO results were compared to the GA results of previous authors. PSO was found to outperform the GA resulting in better norms.

REFERENCES

- [1] U.M. Al-Saggaf, M. Bettayeb, "Techniques in Optimized Model Reduction for High Dimensional Systems", *Control and Dynamic Systems*, vol. 55, pp. 51-109, 1993.
- [2] B.C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction", *IEEE Trans. Automatic Control*, vol. AC-26, pp. 17-32, 1981.
- [3] L.M. Silverman, M. Bettayeb, "Optimal Approximation of Linear Systems", in *JACC*, vol. 2, p. FA 8-A, 1980.
- [4] M. Bettayeb, L.M. Silverman, M.G. Safonov, "Optimal Approximation of Continuous Time Systems", *Proc. 19th IEEE CDC, Albuquerque, New Mexico*, 1980.
- [5] M. Bettayeb, "Approximation of Linear Systems: New Approaches based on Singular Value Decomposition", *Ph.D. Thesis, University of Southern California, Los Angeles*, 1981.
- [6] K. Glover, "All Optimal Hankel Norm Approximations of Linear Multivariable systems and their L_∞ -error Bounds", *Int. J. Contr.*, vol. AC-27, pp. 1115-1193, 1984.
- [7] D.C. Hyland, and D.S. Bernstein, "The Optimal Projection Equations for Model Reduction and the Relationships Among the Methods of Wilson, Skelton and Moore", *IEEE Transaction on Automatic Control*, AC-30, 1201-1211, 1985.
- [8] W.-Y. Yan, J. Lam, "An Approximate Approach to H_2 Optimal Model Reduction", *IEEE Transaction on Automatic Control*, vol. 44, no. 7, pp. 1341-1357, July 1999.
- [9] E. Assunção, P.L.D. Peres, " H_2 and/or H_∞ -Norm Model Reduction of Uncertain Discrete-time Systems", *Proceedings of the American Control Conference, San Diego, California, USA*, pp. 4466-4470, June 1999.
- [10] E. Assunção, P.L.D. Peres, "A Global Optimization Approach for the H_2 -Norm Model Reduction Problem", *Proceedings of the 38th Conference on Decision and Control, Phoenix, Arizona, USA*, pp. 1857-1862, December 1999.
- [11] X.-X. Huang, W.-Y. Yan, K. L. Teo, " H_2 Near-Optimal Model Reduction", *IEEE Transaction on Automatic Control*, vol. 46, no. 8, pp. 1279-1284, August 2001.
- [12] J.-P. Marmorat, M. Olivi, B. Hanzon, R.L.M. Peeters, "Matrix Rational H_2 Approximation: A State-Space Approach Using Schur Parameters", *Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, Nevada, USA*, pp. 4244-4249, December, 2002.
- [13] R.L.M. Peeters, B. Hanzon, D. Jibeteau, "Optimal H_2 Model Reduction in State-Space: A Case Study", *Proceedings of the European Control Conference, Cambridge, UK*, September 2003.
- [14] C.A. Beattie, S. Gugercin, "Krylov-Based Minimization for Optimal H_2 Model Reduction", *Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, USA*, pp. 4385-4390, December 2007.
- [15] Y. Liu, B.D.O. Anderson, "Model Reduction with Time Delay", *IEE Proceedings D*, vol. 134, no. 6, pp.349-367, November 1987.
- [16] V. Gazi, K.M. Passino, "Stability Analysis of Swarms", *IEEE Transaction on Automatic Control*, vol. 48, no. 4, pp. 692-697, April 2003.
- [17] Y. Shi, "Particle Swarm Optimization", *IEEE Neural Networks Society*, pp. 8-13, Feb. 2004.
- [18] R. Eberhart, J. Kennedy, "A New Optimizer Using Particle Swarm Theory", *Sixth Intl. Symposium on Micro Machine and Human Science, Nagoya, Japan*, pp. 39-43, Oct. 1995.
- [19] J. Kennedy, R. Eberhart, "Particle Swarm Optimization", *Proc. IEE International Conference on Neural Networks*, pp. 1942-1947, 1995.

- [20] R. Mendes, J. Kennedy, J. Neves, "The Fully Informed Particle Swarm: Simpler, Maybe Better", *IEEE Trans. on Evolutionary Computation*, vol. 8, no. 3, pp. 204-210, June 2004.
- [21] X. Hu, R. Eberhart, Y. Shi, "Engineering Optimization with Particle Swarm", *IEEE Swarm Intelligence Symposium*, 53-57, 2003.
- [22] M. Clerc, J. Kennedy, "The Particle Swarm: Explosion, Stability, and Convergence in a Multi-Dimensional Complex Space", *IEEE Trans. Evolution. Comput.*, vol. 6, no. 1, pp. 58-73, 2002.
- [23] M.S. Voss, X. Feng, "ARMA Model Selection Using Particle Swarm Optimization and AIC Criteria", *15th Triennial World Congress, Barcelona, Spain: IFAC*, 2002.
- [24] M. Fleischer, "Foundation of Swarm Intelligence: From Principles to Practice", *Conference on Swarming: Network Enabled C4ISR, McLean, Virginia, Jan. 2003*.
- [25] B. Birge, "PSOt – A Particle Swarm Optimization Toolbox for Use with MATLAB", *IEEE Swarm Intelligence Symposium Proceedings*, pp. 182-186, April, 2003.
- [26] Z.J. Yang, T. Hachino, T. Tsuji, "Model Reduction with Time Delay Combining the Least-Squares Method with the Genetic Algorithm", *IEE Proceedings on Control Theory and Applications*, vol. 143, no. 3, pp. 247-254, 1996.

Optimization Algorithm for Combined Economic and Emission Dispatch with Security Constraints

N. Cetinkaya

Department of Electrical and Electronics Engineering, Selcuk University, Konya, 42075, Turkey

Abstract—A power system operation at minimum cost is no longer the only criterion for electrical power dispatch. Combined economic emission dispatch problem is obtained by considering both the economy and emission objectives with required constraints. This problem can be solving optimization techniques. Many optimization techniques are slow for such complex optimization tasks and are not suitable for on-line use. This paper presents an optimization algorithm, for solving security constrained combined economic emission dispatch problem, through the application of programming method. The proposed method has been tested on IEEE 30-bus test system and found to be suitable for on-line combined economic emission dispatch.

Index Terms—Economic dispatch, emission dispatch, combined economic emission dispatch.

I. INTRODUCTION

ECONOMIC load dispatch is one of the main functions of electrical power management system [1]. Electrical power system operation should be characterized by security, reliability and economy. The main objective of economic load dispatch (ELD) is to minimize the fuel cost while satisfying the required equality and inequality constraints.

Nowadays, a large part of energy production is done with thermal sources. Thermal electrical power generating is one of the most important sources of carbon dioxide (CO₂), sulfur dioxide (SO₂) and nitrogen oxides (NO_x) which create atmospheric pollution [2]. Emission control has received increasing attention owing to increased concern over environmental pollution caused by fossil based generating units and the enforcement of environmental regulations in recent years [3]. Numerous studies have emphasized the importance of controlling pollution in electrical power systems [4-14].

Which is the best for optimal solution? Economic load dispatch (ELD), emission dispatch (ED) or combined economic emission dispatch (CEED). To find the correct answer to this question, a good power management strategy is required. Several optimization techniques such as lambda iteration, linear programming (LP), non-linear programming (NLP), quadratic programming (QP) and interior point method (IPM) are employed for solving the security constrained economic dispatch and unit commitment problem [15]. Among these

methods, the lambda iteration method has been applied in many software packages due to its ease of implementation and used by power utilities for ELD [16]. Most of the time, alone lambda method does not find optimal solution because of power system constraints. Therefore, the lambda method is used in conjunction with other optimization techniques.

The solution of ELD problem using genetic algorithm required large number of iterations/generations when the power system has large number of units. In order to minimize the number of generations and avoid the loss of useful chromosome for further generation micro genetic algorithm (MGA) was developed [17].

Combined economic and emission dispatch (CEED) has been proposed in the field of power generation dispatch, which simultaneously minimizes both fuel cost and pollutant emissions. When the emission is minimized the fuel cost may be unacceptably high or when the fuel cost is minimized the emission may be high.

In literature as environmental economic dispatch or economic emission dispatch, many algorithms are used to solve CEED problem. Literature [18] proposed a cooling mutation technique in EP algorithm to solve CEED problem for nine units system. Literature [19] validated EP algorithm to solve optimal power flow problem with quadratic and sine component cost functions.

Proposed methods in [20, 21] convert a multiobjective problem into a single objective problem by assigning different weights to each objective. This allows a simpler minimization process but does require the knowledge of the relative importance of each objective and the explicit relationship between the objectives usually does not exist.

In this study two objectives considered are minimizing both fuel cost and environmental impact of emission by using programming based algorithm.

II. PROBLEM FORMULATION

A. Economic Dispatch

The ELD problem is to find the optimal combination of power generation that minimizes the total fuel cost while satisfying the total demand and power system constraints. The fuel costs for power generation units should be defined. The total fuel cost function of ELD problem is defined as follows:

$$F_T = \sum_{i=1}^n f_i(P_i) = \sum_{i=1}^n (a_i P_i^2 + b_i P_i + c_i) \quad (1)$$

$$F_T = F_1 + F_2 + \dots + F_K = \sum_{i=1}^n F_i \quad (2)$$

$$P_T = P_1 + P_2 + \dots + P_K = \sum_{i=1}^n P_i \quad (3)$$

where $f_i(P_i)$ is the cost of i th generator in $\$/h$; P_i is the power output of generator i in MW; a_i , b_i and c_i are the cost coefficients of the i th generator.

The total fuel cost function including valve point loading of ELD problem is defined as follows:

$$F_T = \sum_{i=1}^n ((a_i P_i^2 + b_i P_i + c_i) + |d_i \sin(e_i (P_{i,\min} - P_i))) \quad (4)$$

where F_T is the total fuel cost of electrical power generation in $\$/h$; $P_{i,\min}$ is min. power constraint for i th unit in MW; d , e are the fuel cost coefficients of the i th generating unit reflecting the valve-point effect.

B. Emission Dispatch

The solution of ELD problem will give the amount of active power to be generated by different units at a minimum fuel cost for a particular demand. But the amount of emission or is not considered in pure ELD problem. The amount of emission from a fossil-based thermal generator unit depends on the amount of power generated by the unit. Total emission generated also can be approximated as sum of a quadratic function and an exponential function (5) of the active power output of the generators. The emission dispatch problem can be described as the optimization of total amount of emission release defined by as:

$$E_T = \sum_{i=1}^n ((\alpha_i P_i^2 + \beta_i P_i + \gamma_i) + \xi \exp(\zeta_i P_i)) \quad (5)$$

where E_T is total amount of emission (lb/h); α_i , β_i , γ_i , ξ_i and ζ_i are coefficient of generator emission characteristics.

The total emissions of SO₂, CO₂ and NO_x are represented as follows:

$$E_{SO_2} = \sum_{i=1}^n E_{SO_2i} = \sum_{i=1}^n (\alpha_{SO_2i} P_i^2 + \beta_{SO_2i} P_i + \gamma_{SO_2i}) \quad (6)$$

$$E_{CO_2} = \sum_{i=1}^n E_{CO_2i} = \sum_{i=1}^n (\alpha_{CO_2i} P_i^2 + \beta_{CO_2i} P_i + \gamma_{CO_2i}) \quad (7)$$

$$E_{NO_x} = \sum_{i=1}^n E_{NO_xi} = \sum_{i=1}^n (\alpha_{NO_xi} P_i^2 + \beta_{NO_xi} P_i + \gamma_{NO_xi}) \quad (8)$$

C. Combined Economic and Emission Dispatch

The economic dispatch and emission dispatch are two different problems. Emission dispatch can be included in conventional economic load dispatch problems by the addition of emission cost to the normal dispatch cost. In this method different types of emissions are modeled as a cost in addition to the fuel cost. Actually, CEED problem have two objectives. But CEED can be converted into single objective optimization

problem by introducing a price penalty factor h ($\$/lb$) as follows:

$$h = \frac{F_T(P_{i,\max}) / P_{i,\max}}{E_T(P_{i,\max}) / P_{i,\max}} \quad (9)$$

where $P_{i,\max}$ is max. power constraint for i th unit in MW.

$$\text{Minimize } \phi_T = F_T(P) + h E_T(P) \quad (10)$$

where ϕ_T is the total operational cost of the system subject to the required constraints.

$$\text{Minimize } \phi_T = w_1 F_T(P) + w_2 h E_T(P) \quad (11)$$

where w_1 and w_2 are weight factors. The weight factors w_1 and w_2 have many implications. For $w_1 = 1$ and $w_2 = 0$ the solution will yield results for pure economic dispatch. For $w_1 = 0$ and $w_2 = 1$, results for pure emission dispatch and for $w_1 = w_2 = 1$ results for combined economic emission dispatch can be obtained. The problem can be formulated other form [3, 22] as:

$$\phi_T = w F_T(P) + (1 - w) h E_T(P) \quad (12)$$

D. Equality constraint

$$\sum_{i=1}^n P_i - P_D - P_L = 0 \quad (13)$$

where P_D is the total power demand and P_L is the total transmission loss.

The transmission loss P_L can be calculated by using B matrix technique and is defined by as:

$$P_L = \sum_{i=1}^n \sum_{j=1}^n P_i B_{ij} P_j \quad (14)$$

where B_{ij} 's are the elements of loss coefficient matrix B .

E. Inequality constraints

The cost is minimized with the following generator capacities and active power balance constraints as;

$$P_{i,\min} \leq P_i \leq P_{i,\max} \quad (15)$$

where, $P_{i,\min}$ and $P_{i,\max}$ are the minimum and maximum power generation by i th unit respectively.

$$P_m \leq P_{m,\max} \quad m = 1, \dots, nl \quad (16)$$

where P_m is magnitude of the line flow in m th line, nl is number of lines.

III. PROPOSED ALGORITHM FOR COMBINED ECONOMIC AND EMISSION DISPATCH

Mathematical calculations and comparisons to be done very quickly with Delphi, that's why the proposed algorithm in this paper is written in Delphi programming language.

At first power system data and required constraint must be entered into the program. ELD and ED are solved separately. Suitable generating unit's powers are leaved the total demand power after making the ELD and ED. According to the rest of total demand power, ELD and ED are made solution. When the total power demand and required constraints are suitable for all power system, CEED is done. The total fuel cost and emission are calculated with together in CEED step. When the

convergence is done the problem will be solved. There is a convergence on the change of circumstances of cost increase or cost decrease. Then, generating unit's powers are saved.

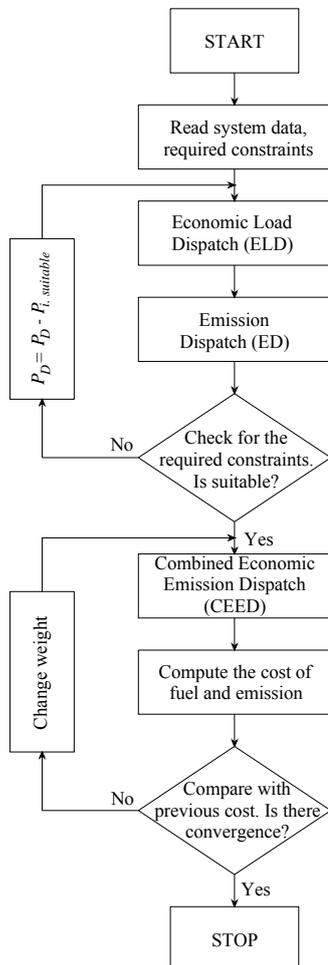


Fig. 1. Flowchart for the proposed CEED method

IV. SIMULATION RESULTS AND DISCUSSION

The proposed combined economic and emission dispatch (PCEED) was solved for IEEE 30 bus and 6 generating unit system.

The cost coefficients and power generation limits for the test system are given in Table I. The NO_x emission coefficients are given in Table II. Test system data are taken from [23].

TABLE I
Generator cost coefficients for IEEE 30 bus system

Unit	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>P_{min}</i>	<i>P_{max}</i>
1	100	200	10	15	6,283	0,05	0,5
2	120	150	10	10	8,976	0,05	0,6
3	40	180	20	10	14,784	0,05	1
4	60	100	10	5	20,944	0,05	1,2
5	40	180	20	5	25,133	0,05	1
6	100	150	10	5	18,48	0,05	0,6

TABLE II
Generator emission coefficients for IEEE 30 bus system

Unit	α	β	γ	ξ	ζ
1	6,490	-5,554	4,091	2,0e ⁻⁴	2,857
2	5,638	-6,047	2,543	5,0e ⁻⁴	3,333
3	4,586	-5,094	4,258	1,0e ⁻⁶	8,000
4	3,380	-3,550	5,326	2,0e ⁻³	2,000
5	4,586	-5,094	4,258	1,0e ⁻⁶	8,000
6	5,151	-5,555	6,131	1,0e ⁻⁵	6,667

TABLE III
Results of best fuel cost for the PCEED and 3 approaches

Unit	PCEED	[23]	[29]	[30]
1	0,1098	0,1281	0,1086	0,1168
2	0,2998	0,2702	0,3056	0,3165
3	0,5244	0,5552	0,5818	0,5441
4	1,0160	1,0053	0,9846	0,9447
5	0,5240	0,4544	0,5288	0,5498
6	0,3598	0,4453	0,3584	0,3964
Best Cost	600,18	606,66	607,807	608,245
Emission	0,2537	0,2207	0,22015	0,21664

TABLE IV
Results of best emission for the PCEED and 3 approaches

Unit	PCEED	[23]	[29]	[30]
1	0,3918	0,3713	0,4043	0,4113
2	0,4603	0,4665	0,4525	0,4591
3	0,5252	0,5642	0,5525	0,5117
4	0,3810	0,3650	0,4079	0,3724
5	0,5467	0,5223	0,5468	0,5810
6	0,5528	0,5783	0,5005	0,5304
Cost	644,40	648,01	642,603	647,251
Best Emis.	0,2125	0,1945	0,19422	0,19432

TABLE V
CEED solutions

Unit	PCEED	[23]	[29]	[30]
1	0,2245	0,17613	0,2594	0,2699
2	0,3324	0,28188	0,3848	0,3885
3	0,5682	0,54079	0,5645	0,5645
4	0,7066	0,76963	0,7030	0,6570
5	0,5917	0,65019	0,5431	0,5441
6	0,4269	0,44569	0,4091	0,4398
Cost	611,635	612,35	616,069	618,686
Emission	0,22915	0,20742	0,20118	0,19940

The fuel costs of PCEED are found to be better than other methods which are shown in Table III, IV and V. The maximum differences between the PCEED fuel costs with the other method's costs are 8.065, 3.61 and 7.051 respectively. But the emission is not the best. The differences between the best emissions with PCEED are 0.03706, 0.01828 and 0.02975 respectively. PCEED solutions for IEEE 30 bus test system are given in Fig. 2.

The PCEED method is used to simulate the cases on an Intel (R) Core (TM)2 Duo T7300 2GHz laptop computer with 1 GB RAM. Computation time is only about 0.6 s.

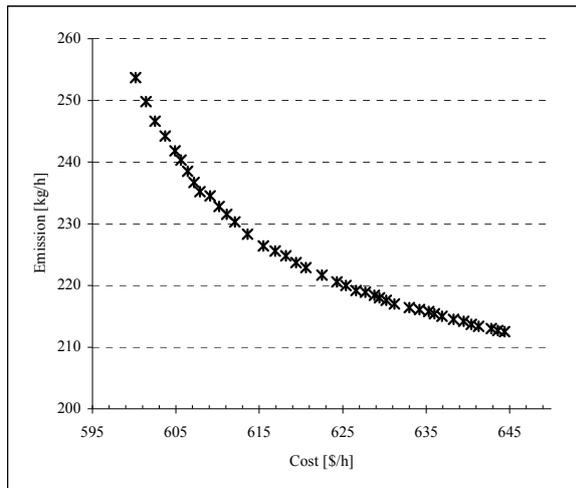


Fig. 2. PCEED solutions for IEEE 30 bus test system

V. CONCLUSION

The proposed multiobjective computer programming based algorithm has been developed to solve CEED problem. The performance of the proposed algorithm is demonstrated for IEEE 30 bus, 6 unit test system. The performance of proposed algorithm is compared with literature [23, 29 and 30]. The results showed that the PCEED method is well suited for obtaining minimum fuel cost. The differences between the emissions are not very important values. Calculation time of the PCEED algorithm is very less according to genetic algorithm and neural network applications in literature. As a result, PCEED algorithm is acceptable and applicable for CEED problem solution.

ACKNOWLEDGMENT

This work is supported by the Coordinatorship of Selçuk University's Scientific Research Projects. (P. N.: 09701253)

REFERENCES

- [1] P. Venkatesh, K. Y. Lee, "Multi-Objective Evolutionary Programming for Economic Emission Dispatch problem," *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008 IEEE, pp. 1-8.
- [2] M. Basu, "A simulated annealing-based goal-attainment method for economic emission load dispatch of fixed head hydrothermal power systems," *Electrical Power and Energy Systems*, 27, 2005, pp. 147-153.
- [3] S.D. Chen and J.-F. Chen, "A direct Newton-Raphson economic emission dispatch," *Electrical Power and Energy Systems*, 25, 2003, pp. 411-417.
- [4] L. Xuebin, "Study of multi-objective optimization and multi-attribute decision-making for economic and environmental power dispatch," *Electric Power Systems Research*, 79, 2009, pp. 789-795.
- [5] L. Wang, C. Singh, "Reserve-constrained multiarea environmental/economic dispatch based on particle swarm optimization with local search," *Engineering Applications of Artificial Intelligence*, 22, 2009, pp. 298-307.
- [6] K. T. Chaturvedi, M. Pandit, L. Srivastava, "Hybrid neuro-fuzzy system for power generation control with environmental constraints," *Energy Conversion and Management*, 49, 2008, pp. 2997-3005.
- [7] Khaled Zehar, Samir Sayah, "Optimal power flow with environmental constraint using a fast successive linear programming algorithm: Application to the algerian power system," *Energy Conversion and Management*, 49, 2008, pp. 3361-3365.
- [8] K. T. Chaturvedi, M. Pandit and L. Srivastava, "Modified neo-fuzzy neuron-based approach for economic and environmental optimal power dispatch," *Applied Soft Computing*, 8, 2008, pp. 1428-1438.
- [9] L. Wang and C. Singh, "Balancing risk and cost in fuzzy economic dispatch including wind power penetration based on particle swarm optimization," *Electric Power Syst. Research*, 78, 2008, pp. 1361-1368.
- [10] L. Wang and C. Singh, "Stochastic economic emission load dispatch through a modified particle swarm optimization algorithm," *Electric Power Systems Research*, 78, 2008, pp. 1466-1476.
- [11] C. Palanichamy and N. S. Babu, "Analytical solution for combined economic and emissions dispatch," *Electric Power Systems Research*, 78, 2008, pp. 1129-1137.
- [12] M. Basu, "Dynamic economic emission dispatch using nondominated sorting genetic algorithm-II," *Electrical Power and Energy Systems*, 30, 2008, pp. 140-149.
- [13] L. Wang and C. Singh, "Environmental/economic power dispatch using a fuzzified multi-objective particle swarm optimization algorithm," *Electric Power Systems Research*, 77, 2007, pp. 1654-1664.
- [14] C.-L. Chiang, "Optimal economic emission dispatch of hydrothermal power systems," *Electrical Power and Energy Systems*, 29, 2007, pp. 462-469.
- [15] P. Somasundaram and K. Kuppusamy, "Application of evolutionary programming to security constrained economic dispatch," *Electrical Power and Energy Systems*, 27, 2005, pp. 343-351.
- [16] A.J. Wood and B.F. Wollenberg, "Power Generation, Operation and Control," Newyork, Wiley, 1984.
- [17] P. Venkatesh, R. Gnanadass, and N. P. Padhy, "Comparison and Application of Evolutionary Programming Techniques to Combined Economic Emission Dispatch With Line Flow Constraints," *IEEE Trans. Power Syst.*, vol. 18, pp. 688-697, May 2003.
- [18] K.P. Wong and J. Yuryevich, "Evolutionary-programming based algorithm for environmentally constrained economic dispatch," *IEEE Trans. Power Syst.*, vol.13, pp. 301-306, May 1998.
- [19] J. Yuryevich and K.P. Wong, "Evolutionary programming based optimal power flow," *IEEE Trans. Power Syst.*, vol.14, pp. 1245-1250, Nov. 1999.
- [20] J. Zahavi, L. Eisenberg, "Economic environmental power dispatch," *IEEE Trans. Syst. Man, Cybern. SMC-5*, 1975.
- [21] J. Cadogan, L. Eisenberg, "SO₂ emission management for electric power systems," *IEEE Trans. PAS*, 96 (2), pp. 393-401, 1977.
- [22] M. Muslu, "Economic dispatch with environmental considerations: tradeoff curves and emission reduction rates," *Electric Power Systems Research*, 71, 2004, pp. 153-158.
- [23] S. Hemamalini, Sishaj P. Simon, "Emission constrained economic with valve point effect using particle swarm optimization," *TENCON 2008. IEEE Region 10. Conference*, 19-21 November 2008 page(s):1-6.
- [24] S.-L. Chen, M.-T. Tsay and H.-J. Gow, "Scheduling of cogeneration plants considering electricity wheeling using enhanced immune algorithm," *Electrical Power and Energy Systems*, 27, 2005, pp. 31-38.
- [25] S. Balakrishnan, P.S. Kannan, C. Aravindan and P. Subathra, "On-line emission and economic load dispatch using adaptive Hopfield neural network," *Applied Soft Computing*, 2, 2003, pp. 297-305.
- [26] R. Baharathi, M.J. Kumar, D. Sunitha and S. Premalatha, "Optimization of Combined Economic and Emission Dispatch Problem - A Comparative Study," *The 8th International Power Engineering Conference, IPEC 2007*, pp. 134-139.
- [27] Prasanna T.S. and Somasundaram P., "Fuzzy Mutated Evolutionary Programming based algorithm for combined economic and emission dispatch," *TENCON 2008 - 2008. TENCON 2008. IEEE Region 10. Conference*, pp. 1-5, 19-21 Nov. 2008.
- [28] Yalcinoz T., Köksöy O., "A multiobjective optimization method to environmental economic dispatch," *International Journal of Electrical Power & Energy Systems*, 2007, 29 (1), pp. 42-50.
- [29] M. Abido, "Environmental/Economic power dispatch using multiobjective evolutionary algorithms," *IEEE Trans. Power Syst.*, vol. 18, No.4, pp. 1529-1537, Nov 2003.
- [30] M.A. Abido, "A novel multiobjective evolutionary algorithm for environmental/economic power dispatch," *Electric Power System Research*, vol. 65, pp. 71-81, 2003.

Bandwidth optimization of a PIFA Antenna using Hybrid Genetic Algorithms

Mohammad Riyad Ameerudden

*Department of Electronics and Communication Engineering
University of Mauritius
Réduit, Mauritius
ameerudden_riyad@infosys.com*

Harry C. S. Rughooputh

*Department of Electrical and Electronic Engineering
University of Mauritius
Réduit, Mauritius
r.rughooputh@uom.ac.mu*

Abstract - Nowadays, the development of mobile communications and the miniaturization of radio frequency transceivers are experiencing an exponential growth, hence increasing the need for small and low profile antennas. As a result, new antennas have to be developed to provide larger bandwidth and this, within small dimensions. The challenge which arises is that the gain and bandwidth performances of an antenna are directly related to its dimensions. The objective is to find the best geometry and structure giving best performance while maintain the overall size of the antenna small.

This paper presents the bandwidth optimization of a PIFA (Planar Inverted-F Antenna) antenna in order to achieve an optimal bandwidth in the 2 GHz band, using optimization techniques based upon Genetic Algorithms (GA). This work involves the Binary Coded Genetic Algorithms (BCGA) as the main optimization engine and the Real Coded Genetic Algorithms (RCGA) as an alternative. The optimization process has been enhanced by using a Hybrid Genetic Algorithm by Clustering. The different PIFA models are evaluated using the finite-difference time domain (FDTD) method.

I. INTRODUCTION

The Planar Inverted-F Antenna (PIFA) is the most widely used antenna owing to its low profile, ease of fabrication, and high efficiency. PIFAs are nowadays being experimented for Multiple Input Multiple Output (MIMO) antenna arrays as well [1]. High gain of antennas, which is an important characteristic, may be attained through proper design and structure. However, there are many constraints, like the overall size, that prevent engineers from designing such antennas. To achieve optimal performance from the antenna, the geometry has to be optimized. This work describes the performance evaluation and optimization of the PIFA antenna via Modeling methods and Optimization techniques.

A. Modeling methods

A variety of three-dimensional full-wave methods are available for electromagnetic propagation in space. Of these, three particular methods have become the most popular:

- the Finite Element Method (FEM)
- the Transmission Line Matrix (TLM)
- the Finite Difference Time Domain (FDTD) method

The application of these methods requires the use of powerful computers and delivers good approximation of electric and magnetic field propagation. To evaluate the performance of the antenna and observe the three-dimensional propagation of the electric and magnetic fields, the FDTD method was used.

B. Optimization techniques

Traditional optimization techniques can be classified into two distinct groups: direct and gradient-based methods [2]. Direct search methods use only objective function and constraint values to guide the search strategy, whereas gradient-based methods use the first and second-order derivatives of the objective function and constraints to guide the search process. Since derivative information is not used, the direct search methods are usually slow, requiring many function evaluations for convergence. For the same reason, they can also be applied to many problems without a major change of the algorithm. On the other hand, gradient-based methods quickly converge to an optimal solution, but are not efficient in non-differentiable or discontinuous problems.

Because of the nonlinearities and complex interactions among problem variables, the search space may have more than one optimal solution. When solving these problems, there is no escape if traditional methods get attracted to any of these locally optimal solutions.

The technique which can fortunately alleviate some of the above difficulties is the Genetic Algorithm (GA) technique which can constitute an efficient optimization tool. In this work, the GA used has also been modified for fitness evaluation by clustering.

II. METHODOLOGY

The method used in this work involves the modeling of the PIFA using the FDTD technique whereby bandwidth of the antenna is calculated. The bandwidth is optimized by varying parameters in the antenna and then allowing the optimization tool to converge to the optimal solution, that is, the best bandwidth performance. Further experimentation has been done to analyze the convergence towards the best solution.

A. Implementation of FDTD

FDTD starts by discretizing a three-dimensional space into rectangular cells, which are called Yee Lattice [3]. The Yee lattice is specially designed to solve vector electromagnetic field problems on a rectilinear grid. The grid is assumed to be uniformly spaced, with each cell having edge lengths Δx , Δy and Δz . Fig. 1 shows the positions of fields within a Yee cell.

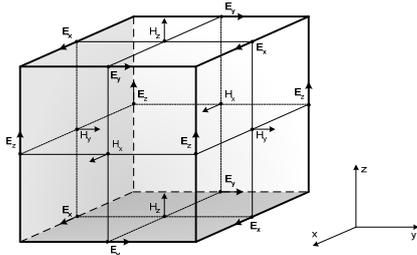


Fig. 1 An FDTD cell or Yee cell showing the positions of electric and magnetic field components

Every \mathbf{E} component is surrounded by four circulating \mathbf{H} components. Likewise, every \mathbf{H} component is surrounded by four circulating \mathbf{E} components. In this way, the curl operations in Maxwell's equations can be performed efficiently.

Arrays must be used to represent the discrete space into a high-level programming language. One-dimensional space is represented by a 1D array, similarly 2D and 3D discrete spaces are represented by 2D and 3D arrays respectively. However, the discrete fields must satisfy Gauss's laws to avoid spurious charge which can corrupt the solution [4].

The cell size is an important aspect of FDTD. It has to be properly chosen for the well-running of the simulation. A wrongly chosen cell size may cause the simulation to display rough and imprecise curves. The cell size is usually determined by calculations involving the frequency of the excitation pulse. In this project, frequencies in the 2 GHz band are studied and since no dielectric has been used, the velocity of the propagating waves has been chosen to be the same as the velocity of free-space that is 300,000km/s. In accordance to the wavelength used, a cell size of 2mm has been used in this work for better resolution of the simulated fields.

1) FDTD onto PIFA

After having discretized the computational space and time, the FDTD has to be applied to the PIFA in order to simulate the propagating \mathbf{E} -fields and \mathbf{H} -fields. The structure of the PIFA varies according to the different context in which it is used. This work deals only with the basic geometry of a

PIFA which normally consists of a ground plate, a radiating plate and a feeding wire.

To excite the PIFA with a wide range of frequencies, a Gaussian pulse implemented as soft source has been used as the excitation source. The source is represented by the feeding wire of the PIFA. The structure to be modeled is shown in Fig. 2 below.

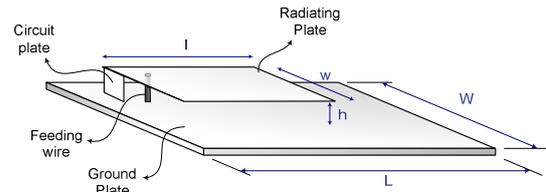


Fig. 2 Geometry of the PIFA to be modeled

2) Absorbing Boundary Condition

Various absorbing boundary conditions have been used for truncating the FDTD mesh in this project, and among those, the Higdon boundary condition, the Dispersive boundary condition and the Mur's second-order boundary condition provide minimal reflection. However the Higdon boundary and Dispersive boundary do not provide significant attenuation over the frequency range of interest. Thus, because of its effectiveness, the Mur's second order boundary condition has been used in this work.

3) Voltage Standing Wave Ratio

The Voltage Standing Wave Ratio (VSWR) is the key to obtaining the bandwidth of the PIFA and thus, the key to achieve the objective of this project. In order to obtain the VSWR, the input impedance of the PIFA has first to be determined. The generalized input or line impedance can be simply calculated using the line voltage and current at a fixed point on the transmission line. These are obtained by Fourier transforming the time-dependent voltages and currents. Using the input impedance calculated, the S_{11} parameter can be evaluated and consequently the VSWR is calculated as

$$VSWR = \frac{1 + |S_{11}|}{1 - |S_{11}|} \quad (1)$$

4) Performance evaluation

VSWR is calculated for several frequencies in the 2GHz band, ranging from 1.9GHz to 2.5GHz. A graph of VSWR against frequencies can be plotted to observe the parabolic shape of the curve. The performance of the antenna is then evaluated by determining the bandwidth from the range of frequencies where the VSWR is less than 2, which is recommended by most telemetry applications.

B. Implementation of the GA

GA is applied to the whole FDTD process which acts as the main component for the fitness evaluation. GA begins its search with a random set of solutions, analyses the solutions and selects the best ones to afterwards converge to the optimal solution, which will result to the best bandwidth performance. The set of solutions was first coded in binary string structures and Binary-Coded GA was used for this purpose. Then Real-Coded GA was used for improvement in convergence and precision to the optimal solution. The GA was then modified to a hybrid version using Clustering technique.

1) Binary-Coded GA

The basic block of the genetic algorithm is the chromosome. Each chromosome is composed of genes described as a binary sequence of zeros and ones. Each gene is associated with a parameter to be optimized. For the PIFA dealt in this project, three parameters were used: the coordinates of the feeding wire, ' f_x ' and ' f_z ', and the height ' h ' of the radiating plate as shown in Fig. 2. For the parameters f_x and f_z , only four discrete possible values were assigned to them and thus, only 2 bits were used to represent each of them. For the height h , only 1 bit was enough to quantify it as only two discrete possible values were assigned to it. Hence, 5 bits represent each chromosome associated with the PIFA and in any case, each chromosome describes a particular geometry.

The population strings are represented as shown in Fig. 3 below:

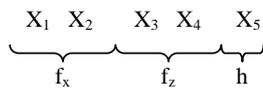


Fig. 3 Population string known as the chromosome

where X_i is a binary digit (0 or 1) and i taking values from 1 to 5. As illustrated in Fig. 3, the first 2 bits represent the parameter f_x , the next 2 bits the parameter f_z and the last bit the height h of the radiating plate.

The population used for the GA is of size 10. Those 10 strings are chosen at random out of all the possible combinations of the parameters and form the population. Each string is then decoded and evaluated.

To start the GA simulation, the initial population is created at random, each string is evaluated and the three GA operators are used. The operators used are the proportionate selection reproduction operator, the single-point crossover operator with a probability of 1 and the bit-wise mutation operator with a probability of 0.01.

Each string is decoded, mapped and evaluated. The evaluation process involves the FDTD method mentioned in the previous section. This explains the long simulation time of

this project: For each chromosome, the GA evaluates the fitness of the string by applying the FDTD method to evaluate the performance of the antenna with the predefined parameters f_x , f_z and h . It should be noted that using all possible values for the parameters, the computational time to end up with an optimal solution may go up to months of simulation. Consequently, the parameters have been restricted to few discrete values, thus reducing the simulation time to 3 to 5 weeks depending on the processing speed of the computer.

2) Real-Coded GA

Handling continuous search space with binary coded GA has several difficulties. Real coded GA represents parameters without coding, which makes representation of the solutions very close to the natural formulation of many problems. Real-world optimization problems often involve a number of characteristics, which make them difficult to solve up to a required level of satisfaction [5]. After experimenting the BCGA, RCGA was experimented to compare the convergence and precision of the optimization process.

In RCGA, decision variables are used directly to form chromosome-like structure. Chromosome represents a solution and population is a collection of such solutions. The operators modify the population of the solution to create new one.

For implementing the RCGA in order to solve problems developed in this model, the following basic components are considered: Parameters of GA, Representation of chromosomes, Initialization of chromosomes, Evaluation of fitness function, Selection process, Genetic operators like crossover [6] and mutation [7].

The real-coded GA depends on some parameters like population size, maximum generation number, probability of crossover and probability of mutation. According to genetics, the probability of crossover is always greater than that of mutation. Generally, the probabilities of crossover and mutation are taken as 0.75 to 0.9 and 0.05 to 0.2 respectively.

3) Clustering GA

GA can sometimes get stuck on sub optimal solution without any progress to the real optimal solution. Previous work has shown that this might be caused by the fact that GA optimizes the whole population out of a whole search space. One of the possible solutions to this problem is to maintain a population size as large as possible. However, maintaining large population involve high cost to evaluate each individual. Therefore to reduce the cost of evaluation and accelerate the convergence the Hybrid Clustering GA [8] is applied in this work. Fig. 4 illustrates the concept behind the conventional GA and the modified clustered GA.

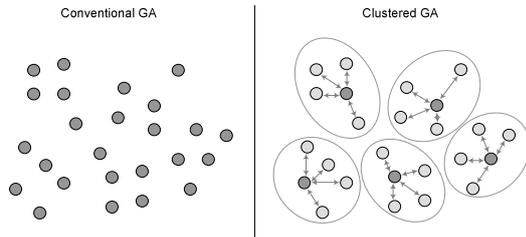


Fig. 4 Conventional GA vs. Clustered GA

As illustrated in Fig. 4, clustering is a simple method of grouping the population into several small groups, called as clusters [9]. The algorithm evaluates only one representative for each cluster. The fitness of other individuals are estimated from the representatives' fitness. Using this method, large population can be maintained with relatively less evaluation cost. One of the important factors to take into consideration for clustering is the similarity measure. This is commonly achieved using distance measures such as Euclidean distance, City block distance and Minkowski distance [10]. Computation of the distance is generally done using equation

$$d_y = d(X_i, X_j) = m \sqrt{\sum_{k=1}^p |x_{ik} - x_{jk}|^m} \quad (2),$$

where $m=1$, $m=2$ and $m \geq 3$ for City block distance, Euclidean distance and Minkowski distance respectively.

There exist other clustering techniques namely the Hierarchical clustering, Overlapping clustering and Partitional clustering. A hybrid GA with clustering based on the k-means algorithms [11] from Partitional clustering had been used in the presented work because of its applicability and flexibility of specifying the number of clusters required.

III. EXPERIMENTAL RESULTS

A series of tests were carried out throughout the work to check whether the implementation of the FDTD was appropriate to evaluate the performance of the PIFA. These tests were carried out using different boundary conditions, different excitation pulses and different computational space size.

The PIFA was excited using a Gaussian waveform of frequency ranging from 1.9GHz to 2.5GHz. The feeding point, that is, the source location was varied by adjusting the parameters f_x and f_z . The height of the radiating plate from the ground plate was also varied by changing the value of another parameter 'h'.

The mesh was tested with various ABCs for comparison purposes. The Mur's Absorbing Boundary Condition has been observed to provide "reflection less" boundary over broad

spectrum and for all angles of incidence and polarization. Moreover, it achieves minimal in computational cost and memory requirement.

The FDTD mesh size was defined large enough for the waves to propagate smoothly. A very large mesh size would obviously give better approximation of the fields propagation since the reflection from the boundaries would be farther from the source. However, a very large mesh size would automatically increase the simulation time considerably.

The bandwidth was evaluated from the input impedance and different VSWR values obtained for the set of frequencies used. A graph of VSWR against frequencies, as shown in Fig. 7, is plotted to show how the bandwidth is obtained.

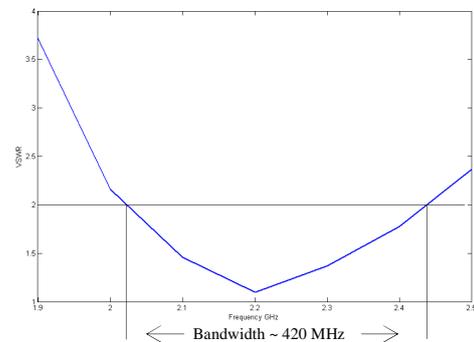


Fig. 7 Graph of VSWR v/s frequency

The BCGA and RCGA has shown to be very good optimization methods. However, it has been observed that both may get stuck to sub optimal solution. To improve the optimization engine, the hybrid GA by clustering was used. The distance from the representatives has been used in the clustered GA. The representative is evaluated and fitness of each representative is calculated. The fitness for the remaining individuals is calculated from the fitness of the representatives in proportion to their respective distance. Using the Euclidean distance as in equation (2) with $m=2$, the distance measure and fitness of remaining individuals can thus be estimated.

Table I shows the comparison of the different GA techniques used in this work after 100 iteration of the optimization. The experiment has been repeated 5 times. The figures in the table show that the conventional GAs may tend towards sub optimal zones and may also get stuck in the wrong zones. BCGA in this case has the worst results in terms of convergence to real optimal zone. RCGA comes to the second position with a significant improvement on the convergence while the hybrid GA by clustering is the clear winner and has proved to have better chance of converging to the optimal solution.

TABLE I
COMPARISON OF GAS OPTIMIZATION – 100 ITERATIONS

	BCGA	RCGA	Hybrid
1	1.0000	1.0000	1.0000
2	0.9093	1.0000	1.0000
3	0.9093	1.0000	1.0000
4	1.0000	0.9996	1.0000
5	0.9093	1.0000	1.0000

IV. CONCLUSION

A. FDTD Outcome

The results show that at different positions of the feeding wire of the PIFA, different waveforms are generated and therefore different values of the input impedance and of the VSWR are obtained. The simulated results also show how the E-fields in between the radiating plate and the ground plate are highly concentrated and how they attenuate and fade out as they propagate in the surroundings.

The FDTD technique was found to be a very powerful tool for the analysis of electromagnetic propagation. However, the FDTD method is limited by the amount of computational memory storage required, which depends on the complexity of the problem structure. The total storage requirement for a given computation can be determined by considering that each three-dimensional FDTD cell requires six real number storage places for the six field components, and an additional large storage for the number of iterations.

B. GA Outcome

The bandwidth obtained from the simulation is approximately 420 MHz. This is the optimal solution generated by the GA. The ground and radiating plates' dimensions were set to 50x26mm and 22x14mm respectively. The values of the parameters achieving this particular bandwidth are $f_x = 3$ cells (6mm), $f_z = 3$ cells (6mm) and $h = 4$ cells (8mm). The results could be enhanced if the population size of the GA was bigger and if the number of discrete values used for the parameters were larger. However, as mentioned earlier, this would cause the simulation to last much longer.

The RCGA has proved to converge better to the optimal solution. However, owing to the large number of combination using real values, the RCGA took more time to simulate.

The Hybrid GA by clustering, on the other hand, has shown to converge faster to the optimal solution. Population size could be increased to some extent without affecting the performance of the optimization using the Hybrid GA by clustering.

C. Improving Outcome

The FDTD simulation has been observed to take really very long before generating the required output. As such, most of the time is spent in the simulation and the future work regarding the FDTD is to optimize the FDTD logic such that processing is faster.

The BCGA has proved to be a very good optimizing tool and if used properly, it may serve to solve various problems of search and optimization. However, the optimization does not always converge to the optimal solution and sometimes get diverted to some other sub-optimal solution. RCGA on the other hand has proved to be a very good convergence tool. Coupled with the clustering method, the GA has shown to be a very powerful optimization tool.

The next step is to test the simulation and optimization using the following:

- Optimized FDTD technique.
- Real-Coded GA with different combination of operators.
- Binary String Fitness Characterization.

ACKNOWLEDGEMENT

I would like to thank the Tertiary Education Commission (TEC) of Mauritius for sponsoring my post graduate research work at the University of Mauritius. I am also grateful to my wife, Shehzanah, whose understanding and ability to put things back in perspective at critical times have been very helpful.

REFERENCES

- [1] Y. Gao, X. Chen, Zhinong Ying, and C. Parini, "Design and performance investigation of a dual-element PIFA array at 2.5 GHz for MIMO terminals", IEEE Transactions on Antennas and Propagation, vol. 55, no. 12, 2007.
- [2] K. Deb, "Optimization for engineering design: Algorithms and examples", Prentice-Hall, Delhi, 1995.
- [3] Yee K. S., "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", IEEE Trans. Antennas Propagat., vol. AP-14, pp. 302-307, May 1966.
- [4] Gedney and Maloney, "Finite Difference Time Domain modeling and applications", FDTD Short Course, Mar. 1997.
- [5] Deb K., "Single and Multi-Objective Optimization Using Evolutionary Algorithms", KanGAL Report No. 2004002, 2004.
- [6] Deb K, Anand A. & Joshi D., "A computationally Efficient Evolutionary Algorithm for Real Parameter Optimization", KanGAL report: 2002003, 2002.
- [7] Z. Michalewicz, "Genetic Algorithms + Data Structures = Evolution Programs", Springer-Verlag, New York, 1992.
- [8] Seront, G. and Bersini, H., "A new GA-local search hybrid for continuous optimization based on multi level single linkage clustering," Proc. of GECCO-2000, pp.90-95, 2000.
- [9] Gose, E., Johnsonbaugh, R. and Jost, S., "Pattern Recognition and Image Analysis", Prentice Hall PTR, 1996.
- [10] Kandel, A., "Fuzzy Techniques in Pattern Recognition", A Wiley-Interscience Publication, 1982.
- [11] Maulik U. and Bandyopadhyay S., "Genetic algorithm-based clustering technique", Journal of Pattern Recognition Society, 1999.

Heuristics, Decision and Optimization

Contents

Artificial neural network training models in prediction of concrete compressive strength using Euclidean normalization method Akdemir BAYRAM, Salih GÜNEŞ, A GENÇ	160
Concept clustering on interval concept lattice Wen ZHOU, Yan ZHAO, Wei LIU, Dong XU, Zongtian LIU, Bofeng ZHANG, Mingjun XIN	166
Dagda, a load-balanced middleware to distribute complex systems simulations Antoine DUTOT, Damien OLIVIER, Guilhelm SAVIN	171
GeoCells model : European Structural Funds and regional interactions, which convergences for the European regions? Bernard ELISSALDE, Patrice LANGLOIS, Dominique GOYAT	175
Hybrid approaches for stock price prediction Wen QINGHUA, Yang ZEHONG, Song YIXU	202
A new method to improve network exchanges of multimedia stream Bechir ALAYA, Claude DUVALET, Bruno SADEG	206

Artificial Neural Network Training Models in Prediction of Concrete Compressive Strength Using Euclidean Normalization Method

B. Akdemir, S. Güneş, A. Genç

Abstract— Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of ingredients. Several studies have shown that concrete strength is not only related to water-to-cement ratio, but also is influenced by the other additive materials. The concrete compressive strength is a complex non-linear regression problem for construction engineering. It is highly difficult to predict the concrete strength due to non-linearity. For nonlinear problems, artificial neural network is widely used to solve and easy to create for any nonlinear problem. Although it is easy to apply, there are some parameters to set carefully to obtain more accuracy. Selection of training method, transfer function, number of nodes and hidden layer are crucial to solve the problem. According to these parameters, normalization method, kind of preprocessing way, is important to improve the accuracy of the solution. Euclidean based normalization method uses Euclid distance to normalize dataset through the attribute. In this study, artificial neural network training methods were employed using euclidean normalization method on prediction of concrete compressive dataset. The best result for concrete compressive dataset was obtained using Levenberg-Marquardt training algorithm. To compare the ANN training methods, mean square error and average deviation have been used as performance measure using two-fold cross validation. Euclidean normalization method composed with artificial neural network with Levenberg-Marquardt training algorithm has better results than the other training methods, namely gradient descent training method, gradient descent with momentum training method, gradient descent momentum and adaptive learning rate training method and gradient descent with adaptive learning rate training method. For proposed method, Euclidean normalization merged with artificial neural network with Levenberg-Marquardt training algorithm, the obtained results for mean square error and average deviation are 0.000197 and 4.564, respectively.

Index Terms— Artificial neural network; Training method; Euclidean normalization; Concrete compressive strength; Levenberg-Marquardt

B. Akdemir is with the Selcuk University, Electrical and Electronics Eng., 42035, Konya, Turkey (corresponding author to provide phone: +90 332 2232042; fax: +90 332 2410635; e-mail: bayakdemir@selcuk.edu.tr).

S. Güneş is with the Selcuk University, Electrical and Electronics Eng., 42035, Konya, Turkey (e-mail: sgunes@selcuk.edu.tr).

A. Genç is with the Selcuk University, Statistical Department, Science Faculty, 42035, Konya, Turkey (e-mail: agenc@selcuk.edu.tr).

I. INTRODUCTION

Concrete has high importance in civil engineering. It is widely used for many kinds of structures. Day by day different structures have been designed and constructed. Earthquake is the most important obstacle to be overcome and taken into account by the experts. In order to simulate behavior of the structure, concrete compressive strength must be known. But it can not be guessed easily due to its ingredients and processes [1]. Concrete compressive dataset has been obtained from data repository of University of California, Irvine [2]. Concrete compressive strength is a nonlinear problem and artificial neural network (ANN) is a way to present a solution [3,4]. ANN has a huge background and many studies have proven that it can present solutions on nonlinear problems [5-7]. ANN has a simple and comprehensible structure. Although a few parameters are present to be adjusted in ANN, it requires expert knowledge and extreme care must be given. ANN has five different training methods widely used in literature. Five common training methods are Levenberg-Marquardt training algorithm, gradient descent training algorithm, gradient descent with momentum training algorithm, gradient descent momentum and adaptive learning rate training algorithm and gradient descent with adaptive learning rate training algorithm abbreviated as trainlm, traingd, traingdm, traingdx, traingda, respectively.

In addition to training algorithms, transfer functions, number of nodes and number of layers must be determined before the processing of ANN. In this study, transfer function was chosen as logsig transfer function for all training algorithms. Number of nodes and hidden layer were chosen as 20 and 1, respectively.

Before the processing of ANN, Euclidean normalization method was performed to normalize concrete compressive dataset. Euclidean normalization method uses Euclid distance to normalize the dataset [8,9]. Euclidean distance is a way to measure similarity and accuracy between predicted and measured value.

In this paper, we have discussed different artificial neural network training methods using real concrete compressive dataset to evaluate the concrete compressive dataset.

Firstly, concrete compressive dataset was performed using Euclidean base normalization method (EBNM) to normalize the dataset as pre-processing. After the normalization, normalized

dataset, which is concrete compressive strength, was applied to ANN using different training algorithms.

In this paper, we have discussed different training algorithms for ANN application using concrete compressive dataset. Artificial neural network with Levenberg-Marquardt training algorithm has given the best results in terms of mean squared error (MSE) and average deviation (AD). MSE and AD values are 0.000197 and 4.564, respectively. Levenberg-Marquardt training algorithm has given approximately 8.36 times better MSE result and 4.33 times better AD result than that of the closest training algorithm, which is gradient descent momentum and adaptive learning rate training algorithm.

II. MATERIAL

The concrete compressive strength is a highly nonlinear function of age and ingredients. Concrete compressive dataset was obtained from data repository, which belongs to University of California, Irvine. The dataset includes 1030 real observation and nine attributes. Concrete compressive dataset has real numbers and no missing value. Table 1 presents the attributes details belonging to concrete compressive dataset.

TABLE I
CONCRETE COMPRESSIVE DATASET DETAILS

Attributes	Unit	Minimum	Maximum	Average
Cement	Kg/m ³	540	102	281.2
Blast Furnace Slag	Kg/m ³	359.4	0	73.9
Fly Ash	Kg/m ³	200.1	0	54.2
Water	Kg/m ³	247	121.8	181.6
Superplasticizer	Kg/m ³	32.2	0	6.2
Coarse Aggregate	Kg/m ³	1145	801	972.9
Fine Aggregate	Kg/m ³	992.6	594	773.6
Age	Day	365	1	45.7
Concrete compressive strength	Mega pascal (MPa)	82.6	2.3	35.8

Concrete compressive strength not only belongs water to cement ratio but also belongs to other materials which is used in the mix. There are six ingredients in the concrete except cement and water. Computing of the concrete compressive strength is not easy due to large number of inputs and nonlinearity of the concrete structure.

III. METHOD

In this paper, we have proposed Euclidean normalization method merged with ANN to predict the concrete compressive strength. Proposed method starts with Euclidean based normalization method (EBNM). After the normalization, normalized concrete compressive dataset was applied to ANN using different training algorithms to obtain the best accuracy.

Mean square error and average deviation were used as performance criteria after the two fold cross validation. Mean square error and average deviation are widely used in literature

to evaluate performance of the system [10, 11].

Two fold cross validation is a common method to present any solution validity [12, 13].

Figure 1 presents the proposed method.

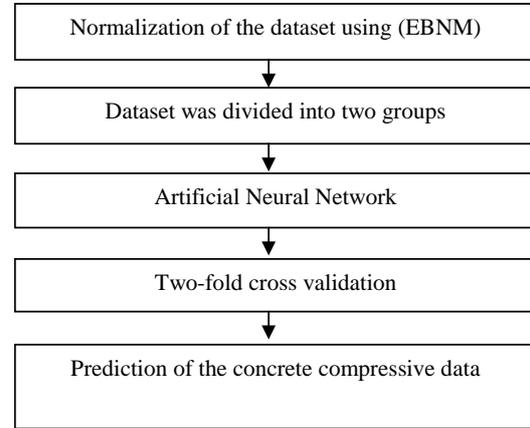


Fig. 1 The flowchart of proposed method

Proposed method includes Euclidean based normalization method and artificial neural network. Firstly, dataset is normalized according to Euclid distance. In second and third stages, artificial neural network is performed and two fold cross validation is applied to improve proposed methods' validity.

A. Euclidean Based Normalization Method

Euclidean based normalization uses Euclid distance to normalize the dataset. Euclidean distance is widely used to measure the similarity and distance between predicted and current value [14]. For every attribute, there is an Euclid distance to compute the new values. Euclidean normalization can be calculated by Equation 1.

$$a' = \frac{a}{\sqrt{a_1^2 + a_2^2 \dots a_{n-1}^2 + a_n^2}} \quad (1)$$

where a' is the new computed values normalized and a_1, a_2, \dots, a_n are member of the same attributes of the dataset, $a \in \mathfrak{R}^n$. After the normalization, obtained matrix and raw matrix have the same dimension.

In addition to training methods, Euclidean based normalization method has been compared to Tchebychev Norm to make a comparison.

B. Artificial neural Network and training algorithm

Artificial neural network is one of the most famous artificial intelligence methods studied in literature, which has learning ability. It has over fifty years of background and there are many studies on ANN, in which it has been employed to solve nonlinear problems [15]. ANN uses nodes to calculate the output. Performing of ANN resembles neurons in brain. There are different transfer functions in the nodes to simulate the

neurons. In this study, logsig transfer function was chosen for all training algorithms and logsig transfer function was not changed through the experiment. Learning ability of ANN was hidden in the neurons and the weights of the net. As depicted in Figure 2, multi layer perceptron (MLP) ANN structure, which is a common configuration, includes one input, one hidden and one output layer [16].

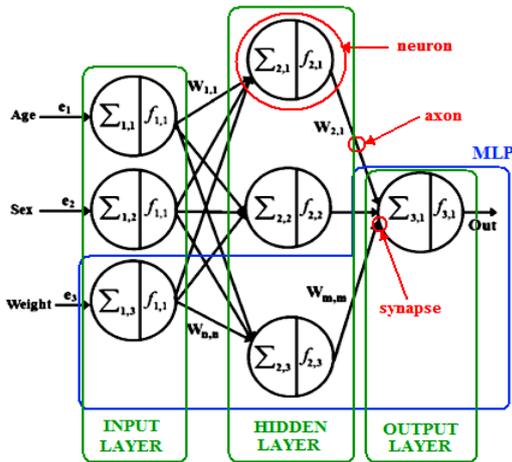


Fig. 2 Structure of ANN

This is miniature of an ANN. The back propagation algorithm is widely used to adjust connection weights and bias values using training.

Each MLP layer is formed by a number of predefined neurons. The neurons in the input layer can be explained as a buffer which distributes the input signals x_i to following neurons in the hidden layer without decaying the signal. Each neuron j in the hidden layer sums the input signals x_i after weighting them with the strengths of the respective connections w_{ij} from the input layer, and computes its output y_j as a function f of the sum as given in Equation 2.

$$y_i = f\left(\sum w_{ij}x_i\right) \quad (2)$$

where f is the activation function which is needed to transform the weighted sum of all signals influencing a neuron. Although there are many activation functions for ANN application due to different data set groups, *logsig* is a very common activation function. All activation functions have different transfer curves that may be threshold, linear tangent etc. In the end, the output neuron in the output layer can be calculated similarly [16-17].

The training parameters and structure of the MLP was set as *logsig* and tried 10 times each for 2000 epoches. In this study, the experimental studies were performed on the MATLABTM 6.5 environment.

In this study, five training algorithms, namely Levenberg-Marquardt training algorithm, Gradient descent training algorithm, Gradient descent with momentum training algorithm, Gradient descent momentum and adaptive learning rate training algorithm, were used.

Gradient descent search method is listed below. Gradient descent search method uses derivatives to reach the target. Gradient can be computed by Equation 3.

$$\nabla E(w) = \partial E / \partial w_0, \dots, \partial E / \partial w_n \quad (3)$$

Fig. 3 Presents computing of gradient descent

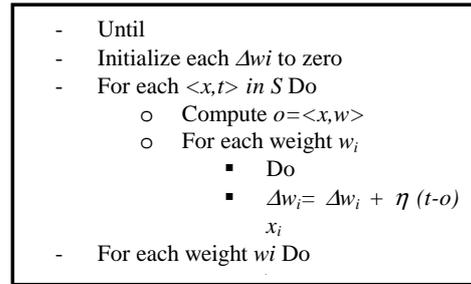


Fig. 3 Gradient descent flow chart

Gradient descent training algorithm (traingd):

Traingd is a network training function that updates weight and bias values according to output error. This learning method is the basic form of back propagation method.

Gradient descent with adaptive learning rate training algorithm (traingda):

For standard gradient descent, the learning rate is held constant throughout training. The performance of the algorithm is very sensitive to the exact setting of the learning rate. If the learning rate is set too high, the algorithm may oscillate and become unstable. If the learning rate is too small, the algorithm takes too long to reach the solution. It is not practical to determine the optimal setting for the learning rate before training. Traingda is a network training function that updates weight and bias values according to adaptive learning rate.

Gradient descent with momentum training algorithm (traingdm):

Another technique that can help the network out of local minima is the use of a momentum term. This is probably the most popular extension of the back propagation algorithm; it is hard to find cases where this is not used. With momentum m , the weights can be updated at a given time t using Equation 4.

$$\Delta w_{(i,j)}(t) = \mu_i \delta_i y_i + m w_{(i,j)}(t-1) \quad (4)$$

Traingdm is a network training function that updates weight and bias values according to gradient descent with momentum.

Gradient descent momentum and adaptive learning rate training algorithm (traingdx): Traingdx is a network training function that updates weight and bias values according to error surface using momentum and an adaptive learning rate.

Levenberg Marquardt training algorithm (trainglm): Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. Levenberg-Marquardt uses approximated Hessian matrix computed from Jacobian Matrix. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in Equation 5 Newton-like update [18]:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (5)$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size.

In Levenberg-Marquardt method, the change (Δ) in the weights (w_i) are obtained by solving Equation 6 [19].

$$\alpha \Delta = -\frac{1}{2} \nabla E \quad (6)$$

where E is the mean-squared network error and can be computed by Equation 7.

$$E = \frac{1}{N} \sum_{k=1}^N [\bar{y}(x_k) - \bar{d}_k]^2 \quad (7)$$

N is the number of samples, $\bar{y}(x_k)$ is the network output corresponding to the sample x_k , and \bar{d}_k is the desired output for the sample.

Adjusting the μ changes the behavior of the Levenberg-Marquardt between Newton method and gradient descent method.

Five training methods have been performed using real concrete compressive strength data set to evaluate ANN performance. The obtained performance results have been given in Section IV.

C. Two-Fold cross validation

Two-fold cross validation is widely used to improve validity of the any proposed system [18]. The data set is divided into two subsets as 50% training and 50% testing and the two-fold cross validation method is two times. Each time, one of the one subsets is used as the test set and the other subset is used a training set. Then the average error across two trials is computed. The advantage of this method is that it matters less how the data gets divided and every data is used test and train.

D. Performance Criteria

There are several ways to measure performance of any system. In this study, mean squared error and average deviation were used as performance criteria after the two fold cross validation.

Mean squared error (MSE) is a widely used method to evaluate the performance of any system. As the name suggests,

MSE error can be explained as a measure of the center distribution quality. MSE is average of the squares of the distances between target and the predicted values [19]. MSE can be calculated by Equation 8.

$$MSE = \frac{1}{n} \sum_{i=1}^n (T_{(i)} - P_{(i)})^2 \quad (8)$$

Average deviation (AD) Average deviation can be explained as a percentage of the average error, as shown in Equation 9 [20].

$$AD = \frac{1}{n} \sum_{i=1}^n \frac{|T_{(i)} - P_{(i)}|}{P_{(i)}} \times 100 \quad (9)$$

In order to demonstrate the performance of proposed method, two-fold cross validation was used. Dataset was divided into two groups as 50% training and 50% testing. After training and testing was performed in ANN, testing and training datasets were re-located and ANN was performed again.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, Euclidean based normalization merged with artificial neural network was discussed for different ANN training methods to predict concrete compressive strength. Concrete compressive data set has real observation. For training methods, Levenberg Marquardt has given the best results among the five training methods to predict the concrete compressive strength. Figure 4 presents error graphics of the Levenberg Marquardt training method for concrete compressive dataset. For the best fitting, the obtained results must be closer to diagonal line.

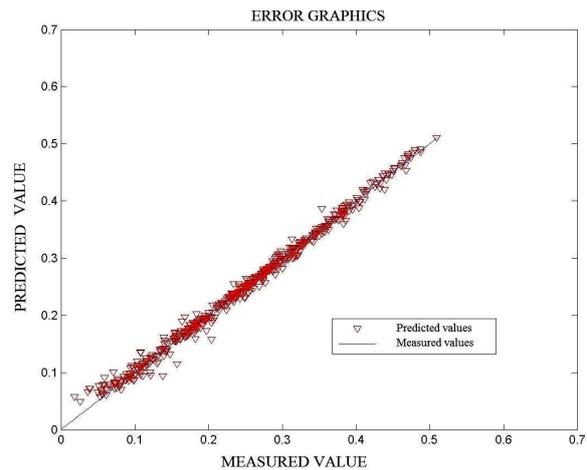


Fig. 4 Error graphics of the Levenberg-Marquardt training method

Figure 5 presents outputs for all inputs in the normalized axis.

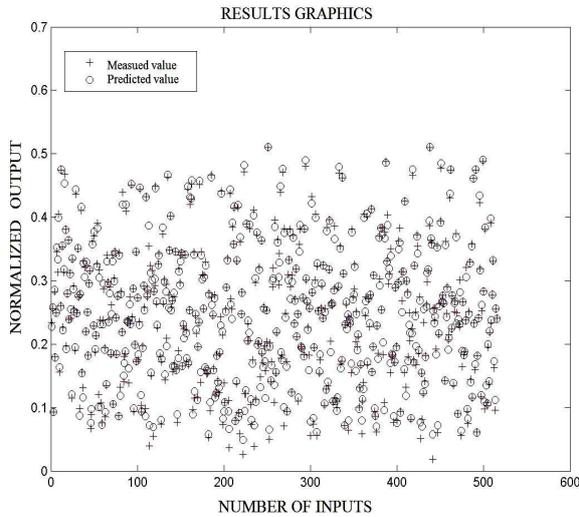


Fig. 5 The results graphics of Levenberg-Marquardt training method

Table 2 presents five training methods and their mean squared error and average deviations for Euclidean normalization based ANN performing. In addition to Euclidean based normalization method.

It can be seen in the Table 2, Levenberg-Marquardt training method has the best results among all the training methods. MSE and AD values for Levenberg-Marquardt training method are 0.000197 and 4.564, respectively.

TABLE II
THE PERFORMANCE RESULTS OF THE DIFFERENT TRAINING METHODS FOR THE PREDICTION OF CONCRETE COMPRESSIVE STRENGTH USING EUCLIDEAN NORMALIZATION

Training method	Mean Squared Error	Average Deviation
Levenberg Marquardt	0.000197	4.564
Triangd	0.033742	66.58
Triangdm	0.052518	89.32
Triangda	0.002474	23.34
Triangdx	0.001648	19.77

Table 3 presents same five training methods and their mean squared error and average deviations for Tchebytchev normalization to make a comparison.

TABLE III
THE PERFORMANCE RESULTS OF THE DIFFERENT TRAINING METHODS FOR THE PREDICTION OF CONCRETE COMPRESSIVE STRENGTH USING TCHEBYTCHEV NORMALIZATION

Training method	Mean Squared Error	Average Deviation
Levenberg Marquardt	0.888465	-12.729
Triangd	0.931183	-26.836
Triangdm	0.811167	-30.465
Triangda	0.774674	-17.45
Triangdx	0.677321	-25.504

From the Table 3, MSE and AD values belonging to concrete compressive strength data set which is normalized using Tchebytchev norm are worse than Euclidean normalization method.

V. CONCLUSION

In this paper, five common ANN training methods were discussed to predict concrete compressive dataset using Euclidean normalization method. Among the five methods, Levenberg-Marquardt training method has the best results in terms of mean squared error and average deviation. The obtained results of MSE and AD are 0.000197 and 4.564, respectively. In addition to Euclidean normalization method, Tchebytchev norm was used as normalization and compared to Euclidean normalization for same training methods and same data set, concrete compressive strength data set. The obtained results of MSE and AD are 0.888465 and -12.729, respectively.

The test results have shown that Levenberg-Marquardt training method for ANN has the best result for predicting concrete compressive dataset and it encourages us to use Levenberg-Marquardt training method to predict the concrete compressive dataset.

Acknowledgments. This study has been supported by Scientific Research Project of Selcuk University

REFERENCES

- [1] Noorzaei J., Hakim S.J.S., Jaafar M.S., Thanoon W.A.M. "Predicting the compressive strength and slump of high strength concrete using neural network", *International Journal of Engineering and Technology*, Vol. 4, No. 2, 2007, pp. 141-153.
- [2] <http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>, Last Access: 21 May 2009.
- [3] I-Cheng Yeh, "Modeling of strength of high performance concrete using artificial neural networks," *Cement and Concrete Research*, Vol. 28, No. 12, 1998, pp. 1797-1808
- [4] I-Cheng Yeh, "Analysis of strength of concrete using design of experiments and neural networks," *Journal of Materials in Civil Engineering*, ASCE, Vol.18, No.4, pp.597-604 (2006).
- [5] Dipietro L., Sabatini A.M., Dario P. Artificial neural network model of the mapping between electromyographic activation and trajectory patterns in free-arm movements. *Medical & Biological Engineering & Computing*, 41, 2003, pp:124-132.
- [6] Akdemir B., Oran B., Gunes S., Karaaslan S., Prediction of Aortic Diameter Values in Healthy Turkish Infants, Children, and Adolescents by Using Artificial Neural Network, *Journal of Medical Systems*, Accepted: 31 July 2008 (Article In Press).
- [7] RH Stevens, AC Lopo and P Wang, Artificial neural networks can distinguish novice and expert strategies during complex problem solving, *Journal of the American Medical Informatics Association*, Vol 3, American Medical Informatics Association, 1996, pp:131-138.
- [8] Balaji R., Bapat R.B. "On Euclidean distance matrices" *Linear Algebra and its Applications* 424 (2007) 108-117.
- [9] Leutbecher M.,Palmer T.N., "Ensemble forecasting" *Journal of Computational Physics* 227 (2008), pp: 3515-3539
- [10] Fadare D.A., "Modelling of solar energy potential in Nigeria using an artificial neural network model", *Applied Energy*, Volume 86, Issue 9, September 2009, Pages 1410-1422.
- [11] Velden M.V.D., Groenen P.J.F., Poblome J., "Seriation by constrained correspondence analysis: A simulation study", *Computational Statistics & Data Analysis*, Volume 53, Issue 8, 15 June 2009, Pages 3129-3138.

- [12] Burez J., Van den Poel D., "Handling class imbalance in customer churn prediction" *Expert Systems with Applications*, Volume 36, Issue 3, Part 1, April 2009, Pages 4626-4636.
- [13] Subasi M., Subasi E., Anthony M., Hammer P.L., "Using a similarity measure for credible classification", *Discrete Applied Mathematics*, Volume 157, Issue 5, 6 March 2009, Pages 1104-1112.
- [14] Demirci R., "Similarity relation matrix-based color edge detection", *Int. J. Electron. Commun. (AEÜ)* 61 (2007), pp:469 – 477.
- [15] Akdemir B., Günes S., Yosunkaya Ş., New Data Pre-Processing on Assessing of Obstructive Sleep Apnea Syndrome: Line Based Normalization Method (LBNM), *International Conference on Intelligent Computing, CCIS (15), ICIC-2008, China*, pp: 185-191, 2008.
- [16] Rumelhart D. E., McClelland J. L., *Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1986.
- [17] Rumelhart, D., Hinton, D., Williams, G., Learning internal representations by error 547 propagation, in Rumelhart D., McClelland F., eds., *Parallel Distributed Processing 7*. 548 M.I.T. Press, Cambridge, MA, USA., 1986.
- [18] http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/index.html?access/helpdesk/help/toolbox/nnet/backpro6.html&http://www.google.com.tr/search?hl=tr&rlz=1R2AMSA_en&ei=H-QTSq2VFw6jAeqz7HABA&sa=X&oi=spell&resnum=0&ct=result&cd=1&q=gradient+descent+momentum+backpropagation&spell=1 Last Access: 21 May 2008
- [19] Lera G., Pinzolas M., Neighborhood Based Levenberg–Marquardt Algorithm for Neural Network Training, *IEEE Transactions on Neural Networks*, Vol. 13 (5), 1200-1203, Sep. 2002.
- [20] Weidhaas J.B., Winter S. Li, K., Ryu J., Jhingra A., Miller B., Dicker A., Gaffney D., "Changes in Gene Expression Predicting Local Control in Cervical Cancer: Results from RTOG 0128", *International Journal of Radiation Oncology Biology Physics*, Volume 72, Issue 1, Supplement 1, 1 September 2008, pp:16.
- [21] http://en.wikipedia.org/wiki/Mean_squared_error, Last Access: 21 May 2009.
- [22] Kaveh, N. S., Ashrafizadeh S. N., Mohammadi F., Development of an artificial neural network model for prediction of cell voltage and current efficiency in a chlor-alkali membrane cell, *chemical engineering research and design*, Vol. 8 (6), 461–472, 2008.

Concept Clustering on Interval Concept Lattice for Web Exploring

ZHOU Wen, ZHAO Yan, LIU Wei, XU Dong, LIU Zongtian, ZHANG Bofeng, XIN Mingjun

Abstract—A concept clustering method on interval concept lattice is introduced in this paper. Based on the distance between two formal concepts defined, the clustering method is presented. Clustering can reduce the size of interval concept lattice to get better interval lattice structure which can be easier to understand. Experimental results show that the reduction algorithm has reasonable performance on the complexity.

Index Terms—interval concept lattice, interval clustering, interval FCA, reduction of interval lattice

I. INTRODUCTION

Formal Concept Analysis (FCA) is a data analysis technique based on the ordered lattice theory firstly introduced by Wille [1]. It defines formal contexts to represent relations between objects and attributes and interprets the corresponding concept lattice. Actually classical formal concept analysis just can deal with the context with Boolean value. Burusco introduced the interval data into FCA at the first time [2]. Combining with interval analysis method [4-7], An interval concept lattice model is introduced in [8] which is inspired by Scaling [3] to process interval attributes by interval attribute scaling. Interval concept lattice (abbr. interval lattice) extends the data process ability of FCA to interval data.

In interval lattice, the objects which have very small different attribute will be separated into different formal concept. In fact, people think they are similar, and normally not divide them into different classes. Base on this observation, a clustering method is proposed for the reduction of interval concept lattice. There are researches on interval clustering [9-12]. The most important thing is to define the distance between intervals and the dissimilarity (or similarity) between interval vectors. On interval lattice, the concept is represented by the attributes of the concept. The attributes can be transformed into interval vector, the dissimilarity between interval concepts are measured by the dissimilarity between interval vectors. Then the clustering algorithm on the interval concept lattice is

Manuscript received Mar. 14, 2009. This work was supported in part by the Shanghai Scientific Special Funds for Cultivation and Selection of Excellent Young Teachers (shu-07027), Innovation project of Shanghai University, and Shanghai Leading Academic Discipline Project (J50103).

ZHOU Wen, LIU Wei, XU Dong, LIU Zongtian, ZHANG Bofeng are with School of Computer Engineering and Science of Shanghai University, Shanghai, China (corresponding author: +86-21-66127707; e-mail: zhouwen, wliu, dxu, ztliu, bfzhang@shu.edu.cn). ZHAO Yan is with School of management of Shanghai University, Shanghai, China. (e-mail: zhaoyan87@shu.edu.cn).

introduced for the reduction of concepts based on the dissimilarity. Actually, the concept lattice is used for web exploring because it's different from the tree structure. A concept node in the lattice can have more than one parent node. When exploring the web, the user can get the target through several routes. And the reduced lattice can help the user get the target faster.

The rest of the paper is organized as following. Section 2 introduces Interval Formal Concept Analysis. Section 3 presents the definition of dissimilarity between interval concepts. Section 4 gives the clustering algorithm for the reduction of interval lattice. The experiment is in Section 5 and Section 6 concludes.

II. INTERVAL FORMAL CONCEPT ANALYSIS

This section concisely introduces the Interval Formal Concept Analysis (IFCA) which was proposed in [8]. Its core data structure is interval concept lattice, which incorporates attribute decomposition based interval scaling to make FCA has the capacity of representing interval information.

An interval $[a, b]$ is the set of all real numbers $\{x: a \leq x \leq b\}$. A natural definition of arithmetic for intervals, represented as pairs of real numbers [13]. The interval $[a, b]$ with $a=b$ is called degenerate interval which degenerate to a real number a . interval decomposition denoted as \odot . There are two rules in decomposition process:

- i. The degenerate intervals do not participate in the decomposition process.
- ii. The decomposition process does not generate degenerate intervals.

Firstly the two intervals decomposition process is analyzed. Both $[a_1, b_1]$ and $[a_2, b_2]$ are not degenerate intervals. Without missing the generality, we assume $a_1 \leq a_2$.

$$[a_1, b_1] \odot [a_2, b_2] = I_1, I_2, I_3 = [a_1, a_2 \wedge b_1], [a_2, (a_2 \vee b_1) \wedge b_2], [(a_2 \vee b_1) \wedge b_2, b_1 \vee b_2] \quad (1)$$

If the interval $I_i = [a^-, a^+]$ ($i=1,2,3$) with $a^- = a^+$ (a^- and a^+ are the lower bound and the upper bound of I_i respectively), it does not satisfy rule 2, and will be deleted from the result. There exists eight conditions of the two interval $[a_1, b_1]$ and $[a_2, b_2]$ shown in Table 1.

Type	Condition	$[a_1, b_1] \odot [a_2, b_2]$	Example
1	B1 -	$[a_1, b_1]$	$[2,5] \odot [2,5] = [2,5]$
2	A1 B2 C3	$[a_1, b_1], [b_1, b_2]$	$[2,8] \odot [2,12] = [2,8], [8,12]$
3	B3 -	$[a_1, b_2], [b_2, b_1]$	$[2,8] \odot [2,5] = [2,5], [5,8]$

4	B1	C3	$[a_1, a_2], [a_2, b_2]$	$[2,6] \odot [3,6]$ $= [2,3], [3,6]$
5		C1	$[a_1, b_1], [a_2, b_2]$	$[2,4] \odot [4,6]$ $= [2,4], [4,6]$
6	A2	B2	C2 $[a_1, b_1], [a_2, b_2]$	$[2,3] \odot [4,6]$ $= [2,3], [4,6]$
7		C3	$[a_1, a_2], [a_2, b_1], [b_1, b_2]$	$[2,5] \odot [4,7]$ $= [2,4], [4,5], [5,7]$
8	B3	-	$[a_1, a_2], [a_2, b_2], [b_2, b_1]$	$[2,11] \odot [5,9]$ $= [2,5], [5,9], [9,11]$

(The condition $A1$ and $A2$ denote $a_1 = a_2, a_1 < a_2$ respectively, $B1, B2$ and $B3$ denote $b_1 = b_2, b_1 < b_2$, and $b_1 > b_2$, and $C1, C2, C3$ denote $b_1 = a_2, b_1 < a_2$ and $b_1 > a_2$ respectively.)

The information table (G, M, R) (the relation of $R(g, m)$ ($g \in G, m \in M$) has the value of $u(g, m)$) can be represented as a cross-table shown in Table 2. The information table has three objects representing three doctors, namely D_1, D_2 and D_3 . In addition, it also has four attributes, "working day" (A), "working time" (B), "registration fee" (C) for out-patient, "physician level" (D) of the doctors. The relation between an object and an attribute is represented by interval data as showing in Table 2.

Table 2. Information table of interval.

	A	B	C	D
D_1	[1,3]	[8,12]	[20,100]	[9,10]
D_2	[2,5]	[14,17]	[10,50]	[1,6]
D_3	[6,7]	[9,11]	[20,50]	[7,8]

Interval attribute decomposition is for $m (m \in M)$, the intervals $u(g, m)$ of $R(g, m)$ in (G, M, R) is decomposed as W_m . The information table can be transferred to interval context by interval attribute scaling.

Definition 1 interval attribute scaling. After interval attribute decomposition the information table (G, M, R) is extended to $(G, M, (W_m)_{m \in M}, R, I)$, each W_m is a set of decomposed intervals of $\mu(g, m)$ for the attribute $m (m \in M, g \in G)$, $(g, m, w) \in I$ if and only if there exists $w \subseteq \mu(g, m)$ ($w \in W_m, \mu(g, m)$ is the value of $R(g, m)$ of information table (G, M, R)).

Definition 2 interval formal context. It is a tuple $K := (G, M, (W_m)_{m \in M}, I)$ where G is a set of objects, M a set of attributes, and $I \subseteq G \times \{(m, w) \mid m \in M, w \in W_m\}$, a relation with $(g, m, w_1) \in I, (g, m, w_2) \in I \Rightarrow w_1 = w_2$. $(g, m, w) \in I$ is read "object g has value w for attribute m ".

The set of W shown in Table 3 is generated from the attribute M from Table 2.

Table 3. W generated by interval attribute decomposition

a_i	$[1,2]$	$a_2 [2,3]$	$a_3 [3,5]$	$a_4 [6,7]$
b_1	[8,9]	$b_1 [9,11]$	$b_3 [11,12]$	$b_4 [14,17]$
c_1	[10,20]	$c_1 [20,50]$	$c_1 [50,100]$	
d_1	[1,6]	$d_1 [7,8]$	$d_1 [9,10]$	

The interval formal context is as shown in Table 4 which is obtained from Table 2 by interval attribute scaling after attribute decomposition.

Table 4. Interval formal context

M	A				B				C				D			
W	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	c_1	c_2	c_3	c_4	d_1	d_2	d_3	
D_1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	
D_2	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0	
D_3	0	0	0	1	0	1	0	0	0	1	0	0	1	0	1	

Definition 3. for $A \subseteq G$, there are two mapping:

$$A' := \{w \in W_m \mid \forall g \in A, m \in M, (g, m, w) \in I\}$$

$$B' := \{g \in G \mid \forall w \in W_m, m \in B, (g, m, w) \in I\}$$

Definition 4 interval formal concept (interval concept). $C(A, B)$ is called an interval formal concept of $K := (G, M, W, I)$ if and only if $A \subseteq G, B(M_B, W_B), M_B \subseteq M, W_m(B) = \{\cup w_m \mid w_m \in B\}, A' = B$ and $B' = A$, A and B are called the **extent** and **intent** of C respectively.

Definition 5 interval concept lattice (interval lattice). The concepts of a given context are naturally ordered by the partial relation defined by $C_1(A_1, B_1) \leq C_2(A_2, B_2) :\Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow W_m(B_2) \subseteq W_m(B_1)$ where $W_m \in B_2$. The ordered set of all interval formal concepts of $K := (G, M, W, I)$ is denoted by $\underline{\mathcal{K}}(G, M, W, I)$, and is called interval concept lattice of (G, M, W, I) . Without ambiguousness, the interval concept lattice is called concept lattice or lattice.

The attribute decomposition is the base of attribute scaling to form the interval context after the attribute decomposition.

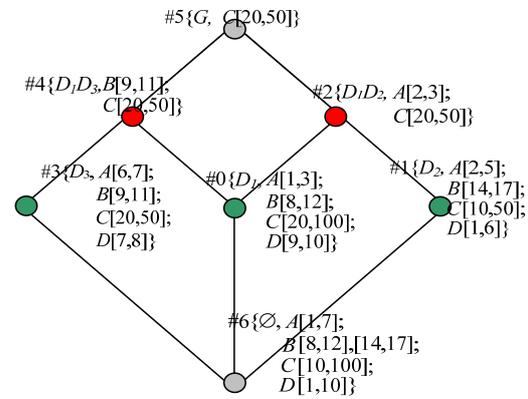


Fig. 1. The interval concept lattice

The interval concept lattice is generated as Figure 1 from context shown as Table 4.

III. DISSIMILARITY BETWEEN TWO INTERVAL FORMAL CONCEPTS

In interval concept lattice, each concept can be represented by a concept vector which is used for calculating the similarities between concepts. The distance function of the two vectors of intervals is the dissimilarity function of two interval formal concepts.

The information in the information table of intervals (like Table 2) and the intents of the concepts are used for the generation of the vectors. Each concept is related to one vector. $|M|$ which is the cardinal number of attributes in information table of interval is the dimensionality of the vector $|M|$. The concept vector is defined as Definition 6.

Definition 6 (Concept Vector). Given interval concept $C=(A, B)$ of $\{G, M, I\}$ ($|G|=k, |M|=q$), the vector of C is defined as $V_C(v_1, v_2, \dots, v_i, \dots, v_q)$, $o \in O, m \in M, g \in G$:

$$v_i = \begin{cases} w_{m_i} & m_i \in B \\ \overline{u(g, m_i)} & m_i \notin B, g \in A \end{cases} \quad (2)$$

where $\overline{u(g, m_i)} = [\overline{u(g, m_i)^-}, \overline{u(g, m_i)^+}]$ and

$$\overline{u(g, m_i)^-} = \frac{1}{|A|} \sum_{j=1}^{|A|} u(g_j, m_i)^-, g_j \in A$$

$$\overline{u(g, m_i)^+} = \frac{1}{|A|} \sum_{j=1}^{|A|} u(g_j, m_i)^+, g_j \in A$$

$\overline{u(g, m_i)^-}$ is an average of $u(g, m_i)^-$ where $g \in A$. $\overline{u(g, m_i)^+}$ is an average of $u(g, m_i)^+$ where $g \in A$.

The element v_i of V_C that is corresponding to attribute m_i is different according to whether m_i is included in the intent B of concept C . Concept vector incorporates the fuzziness of concept into concept lattice. It endows the concept with comparable attribute so any two concepts in the concept lattice can be compared with each other. It's the base for the computing of concept dissimilarity. The interval concept vector is derived from information table of intervals according to the extent and intent of the concept, keeping all the information in context and avoiding information lost.

For the comparison between vectors is meaningful, the normalization of vectors of intervals is necessary. Here the normalization process of vectors of intervals is as following: Corresponding to an attribute m_i , the element v_i of V_C is $[a_i, b_i]$. The normalization of v_i is I_i , and $I_i=[c, d]$.

$$c=(a-\min[i])/(\max[i]-\min[i]);$$

$$d=(b-\min[i])/(\max[i]-\min[i]);$$

$$\text{where } \min[i] = \min\{u(g_j, m_i)^-\}, g_j \in G \text{ and}$$

$$\max[i] = \max\{u(g_j, m_i)^+\}, g_j \in G.$$

Here the concept dissimilarity is given as following.

Definition 7 (Concept Dissimilarity). The dissimilarity between concept C_1 and C_2 of interval concept lattice is defined as the distance of interval vectors of concept C_1 and C_2 :

$$d(C_i, C_j) = \sum_{k=1}^{|M|} \phi(I_k^i, I_k^j) \quad (3)$$

$$\phi(I_k^i, I_k^j) = \sqrt{\frac{1}{3} \left[(I_k^{i-} - I_k^{j-})^2 + (I_k^{i+} - I_k^{j+})^2 + \left(\frac{I_k^{i-} + I_k^{i+} - (I_k^{j-} + I_k^{j+})}{2} \right)^2 \right]}$$

The definition of $\phi(I_k^i, I_k^j)$ is introduced from [14]. When

C_1 is the parent concept of C_2 , or C_2 is the parent concept of C_1 , the distance between C_1 and C_2 is called parent-child distance.

IV. CLUSTERING ALGORITHM FOR THE REDUCTION OF INTERVAL FORMAL CONCEPT LATTICE

The distance between two concepts shows the dissimilarity of the two concepts. Dissimilarity threshold T_S for Clustering determines by the specific application or specified by the user. If the distance between two concepts is less than T_S , these concepts are similar with each other and are included in the same concept cluster (CK). If the distance of concepts greater than T_S , the concept be categorized into different clusters.

Definition 8 (the maximal concept in the Concept Cluster). For $\forall CK \subseteq B(K)$, if there is $C_i(A_i, B_i) \in CK$ and there is no $C_j(A_j, B_j) \in CK$, which makes $A_i \subset A_j$, C_i is the maximal concept of CK.

By definition, there is at least one maximal concept in a concept cluster. The concept clustering process just is carried on between the parent and child concepts.

Definition 9 (Conceptual Hierarchy). Concept clusters have hierarchical relationships that can be derived from interval formal concepts on the interval concept lattice. That is, a concept hierarchy.

The interval concept lattice is also very intricate and complicated as a result of producing many interval formal concepts. Objects that have small differences in terms of attribute values are still classified into different formal concepts. However, such objects should belong to the same concept when they are interpreted.

The clustering algorithm on interval concept lattices is shown in Fig. 2.

Algorithm: Interval Concept hierarchy generalization

Input: $T_S, A // A=B(K)$ -{top concept, bottom concept}

Output: interval concept hierarchy

Process:

```

CK=null; //concept cluster
while(CH1!=null)
{ CH2=null;
  for(i=0; i<n; i++)
  { sameCluster=null;
    judgeConcept(C1,sameCluster,C1,CH2);
  }
  CK.add(sameCluster);
  for(each concept 'sc' in sameCluster)
  { sc.cluster(C1);
    //put cluster sign on "C1" on sc
    CH1=CH2;
  }
  judgeConcept(C1,sameCluster,C1,CH2)
{ sameCluster.add(C1);
  if(C1 has no child concept) {return;}
  else{ for(each child concept cij of C1)
    { if(d(C1,cij)<= TS)
      { judgeConcept(C1,sameCluster, cij,CH2); }
      else
      { If(all the parent concepts of cij are
        already have cluster sign)
        {CH2.add(cij); }
      }
    }
  }
}
}

```

Then, form the concept set
 Finally, add edges to concepts, and generate interval concept hierarchy

Fig. 2. Concept hierarchy generalization algorithm

The first cycle process of the 2nd step in Fig.2, CH₁ is the set of all the child concepts of the top concept in interval concept lattice, CH₂ is null. After substituting the clusters with their supremums and adding edges to concepts, the concept hierarchy is generated.

The conceptual clusters generated have the following properties: each concept cluster is a sub-lattice extracted from the interval concept lattice; a formal concept must belong to at least one concept cluster, but it can also belong to more than one conceptual cluster.

After the clustering based on parent-child concept pairs on interval concept lattice of formal context shown in Table 4, the distance between parent-child concept pairs (not including top and bottom concept of lattice) detected as shown in Fig. 3. When the threshold is set as 0.55, concept pairs (#0, #2) and (#3, #4) are clustered respectively as concept cluster CK₁ and CK₂, concept #1, top concept, and bottom concept are formed as three different clusters. The concepts in the lattice are totally clustered into five categories. The concept set is formed as {#1, #2, #4, #5, #6} by substituting the cluster with the supremum. Then the generalization and specialization relations are found out, and the concept hierarchy is generated, shown as Fig. 4.

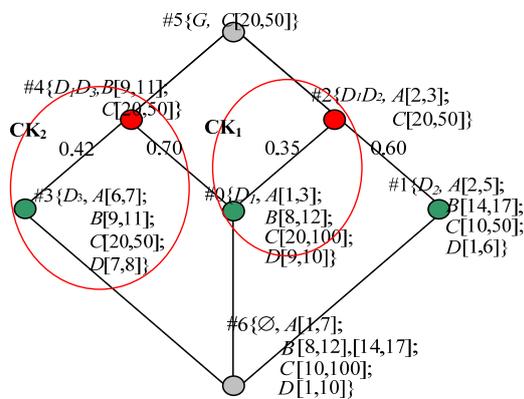


Fig. 3. Conceptual clustering on interval FCA

Then the generalization and specialization relations are found out, and the interval concept hierarchy is generated, shown as Fig. 4.

The interval concept hierarchy reflects the taxonomy relations between the interval concepts. It has smaller size than interval concept lattice.

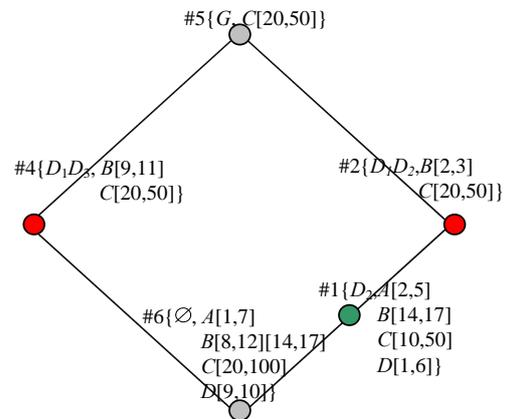


Fig. 4. Interval concept hierarchy

V. EXPERIMENTS

Adopting the data as the test data which is randomly created by JAVA random function, all the algorithms realized with Java running on PC (CPU PIV 2.6G, memory 1.5G) within Windows XP environment. To improve the reality of the experiments, each of experiments with different parameters carries on five randomly generated non-homogeneous data and take the average values of results of five times of experiments as the final result.

The spatial consuming of lattice construction algorithm is tested. Figure 5 expresses the spatial complexity. Curve 1 and Curve 2 show the numbers of lattice nodes changing according to the different number of attributes and the different numbers of objects respectively. The horizontal ordinate indicates both the numbers of objects and attributes are respectively {5,10,20,30,40,50}. The vertical ordinate indicates the size of lattice. With the ascending of object numbers, the size of the lattice has exponential ascending tendency. The influence of attribute number has very slight impact on the size of lattice.

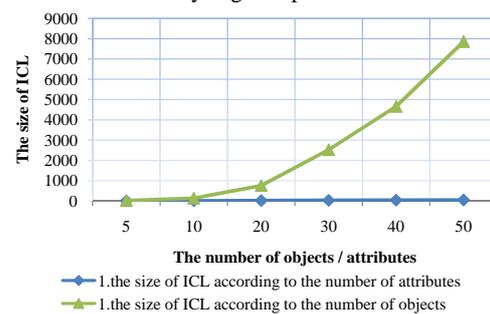


Fig. 5. Relations between the size of lattice and the numbers of both objects and attributes

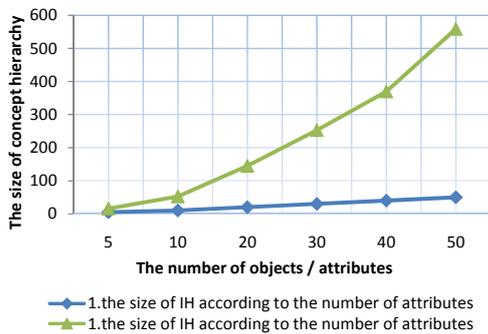


Fig. 6. Relations between the size of interval concept hierarchy and the numbers of both objects and attributes

The curves confirm the results that the numbers of objects have big influence on the lattice in the size and the numbers of attributes haven't so strong impact on the result lattice. Fuzzy lattice spatial consuming character is different. According to the experiments result on fuzzy FCA [15], the numbers of objects has slight influence on lattice size, but the attribute numbers has big impact on it.

The size of reduced interval concept lattice by clustering algorithm is shown in Figure 6 expresses the spatial complexity. Curve 1 and Curve 2 show the size of the reduced interval concept lattice (the interval concept hierarchy) changing according to the different number of attributes and the different numbers of objects respectively. The T_S is set to 0.2. The horizontal ordinate indicates both the numbers of objects and attribute are respectively {5,10,20,30,40,50}. The vertical ordinate indicates the size of interval hierarchy. With the ascending of object numbers, the size of the lattice is sharply reduced.

VI. CONCLUSION

Based on the interval concept lattice model produced extend FCA to IFCA which extends the expression capacity of the classical FCA, the clustering based reduction method is

produced in this paper. With the number of objects increasing the size of interval concept lattice is sharply reduced. The future research includes associate rule extracted from the reduced interval concept lattice and the application of this interval concept lattice as the Grid resource management tools.

REFERENCES

- [1] R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts," in *Ordered Sets*, I.Rival, Ed. Reidel, Dordrecht, 1982, pp. 445-470.
- [2] A. Burusco and R. Fuentes-González, "The study of the interval-valued contexts," *Fuzzy Sets and Systems*, vol. 121, pp. 439-452, 2001.
- [3] S. Prediger and G. Stummé, "Theory-driven logical scaling," in *Proc. Intl. Workshop on Description Logics (DL'99)*, Sweden, 1999.
- [4] R. C. Young, "The algebra of many-valued quantities," *Annals of Mathematics*, vol. 104, pp. 260-290, 1931.
- [5] S. Markov and K. Okumura, *The Contribution of T. Sunaga to Interval Analysis and Reliable Computing*: Kluwer Academic Publishers, 1999.
- [6] R. E. Moore, *Interval analysis*: Prentice-Hall Englewood Cliffs, NJ, 1966.
- [7] A. J. Fernandez and P. M. Hill, "An interval lattice-based constraint solving framework for lattices," in *4th International Symposium on Functional and Logic* vol. 1722, 1999, pp. 194-208.
- [8] W. Zhou, Z. T. Liu, and Y. Zhao, "An Interval Lattice Model for Grid Resource Searching," in *Proceedings of the 7th International Conference on Computational Science (ICCS 2007), Part II, LNCS 4488*, pp. 689-695, Beijing, 2007.
- [9] Antonio Irpino and R. Verde, "Dynamic clustering of interval data using a Wasserstein-based distance," *Pattern Recognition Letters*, vol. to be appeared, p. 30, 2008.
- [10] S. Asharaf, M. Narasimha Murty, and S. K. Shevade, "Rough set based incremental clustering of interval data," *Pattern Recognition Letters*, vol. 27, pp. 515-519, 2006.
- [11] M. Chavent, F. A. T. de Carvalho, Y. Lechevallier, and R. Verde, "New clustering methods for interval data," *Computational Statistics*, vol. 21, pp. 211-229, 2006.
- [12] R. de Souza and F. A. T. de Carvalho, "Clustering of interval data based on city-block distances," *Pattern Recognition Letters*, vol. 25, pp. 353-365, 2004.
- [13] R. E. Moore and F. Bierbaum, *Methods and Applications of Interval Analysis*: Soc for Industrial & Applied Math, 1979.
- [14] Y. F. Yin, "Construction of Interval Databases and Its Applications in Knowledge Discovery." vol. Master Guilin: Guangxi Normal University, 2005.
- [15] Y. Qiang, Z. T. Liu, and e. al., "Research on fuzzy concept lattice in knowledge discovery and a construction algorithm," *Acta Electronica Sinica*, vol. 33, p. 350-353, 2005.

Dagda, a load-balanced middleware to distribute Complex Systems simulations

Antoine Dutot, Damien Olivier, Guilhelm Savin

Abstract—Complex systems which are modeled and are simulated in computer science become increasingly sophisticated. The computing power of a single machine becomes insufficient to execute these simulations. Therefore, it needs to exploit computing power of a set of machines.

DAGDA, the architecture and the platform which are presented in this paper, offers a layer between simulation of a complex system and the available resources. This layer manages spreading of entities on machines to reduce work-load and network-load of each machine.

Index Terms—middleware, dynamic load-balancing, complex systems simulations

I. INTRODUCTION

PROGRAMS asking increasingly computing resources, developers create distributed programs that aim to be executed on several computers. This kind of programs raises some problems. There are problems about communication between remote parts of the program: how to realize a layer to process remote calls, what impact this layer will have on the program's performances. There are also problems about computers architecture : is same architecture needed for computers, etc... Finally, there is problem of how distribute tasks or components of the distributed program.

In this paper, we focus on simulations of complex systems and we propose a dedicated platform, DAGDA, for their distribution. This kind of simulations is often composed of a massive set of entities with many interactions between them. Entity is a generic concept which covers for example agent and object concept. Execution of an entity is not deterministic because of interactions existing between the entity and other entities or environment, what explains a distributed approach rather than others methods used in deterministic programs. Execution of such simulations can be modeled by a dynamic graph which allows to describe interactions (edges) existing in a set of elements (nodes).

DAGDA merges a middleware, to allow communication between remote entities, and a load-balancer, to spread these entities on available machines.

Middlewares are a category of programs that creates a layer between a distributed application and computing resources. They help developers by creating an abstraction of the resources, so developers do not have to deal with resource problems and can focus on the application. The subsection I-B describes middlewares.

DAGDA uses the load-balancing algorithm ANTCO² which is described in II. It has been chosen because it spreads entities considering not only the workload of machines but also interactions existing between entities. The load-balancing concept is described in subsection I-C.

A. Active Object

An important pattern needing to be presented for this paper is the active object pattern[1]. The difference between a *basic* object and an *active* object takes place between method call and its execution. With basic object, calling and execution of a method are synchronous (ie Fig. 1).

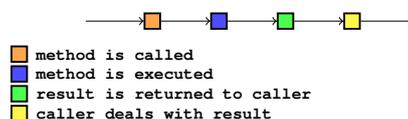


Fig. 1. call of an object method

With active objects, calling and execution are asynchronous. Method calls are request which are sent to an active object. This last one stores requests in a queue and executes them according to a scheduler (ie Fig. 2).

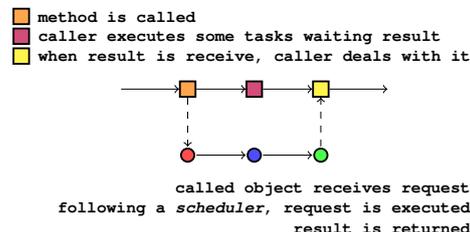


Fig. 2. call of active object method

Figure 3 shows this process : caller send a request, with the help of a stub object, to an active object. This active object returns a *future* with an empty value. When request has been executed, value of the future is set and caller can handle this value.

Interest of active objects is that caller can execute other tasks while it is waiting for a request's result : call is not blocking anymore. Each active object has its own thread to execute received requests. A *stub object* is used to contact an active object on a method call. This call is done through a proxy which is a bridge between stub objects and the active object.

B. Middleware

A middleware provides a connection between softwares or between components of a software. This connection allows a communication between process. These process can be located

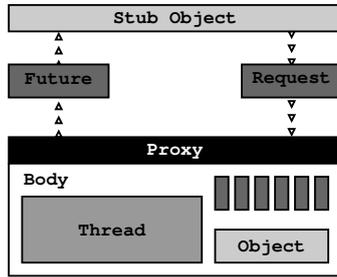


Fig. 3. Active Object pattern

on different machines, with different operating systems. So, a middleware allows a program to exploit power of several participating computers. It can also manage connection and disconnection of participants. So, it creates a dynamic grid of machines used to distribute tasks.

There is different types of middleware. Some of them provide to developers a way to send a job to a machine and then, when job is done, retrieve the result. In this case, there is no interaction between participants, and it is usually a centralized approach : a server (or a restricted set of servers) distributing job to some slaves. For example BOINC[2], which is used by @home projects like SETI@home¹.

Others have an object-approach of the distribution using for some of them the active object pattern. It is the case of ProActive[3] developed by INRIA Sophia Antipolis. This kind of middleware allows interactions between different distributed tasks.

With program running on a single machine, there is an addressing space which allows to attribute a unique id to each object. This id is often the memory address of the object. When a program is distributed (it is running on a set of machines), uniqueness of an id is more difficult to assure. Middlewares provide a global addressing space which allow an unique id for each object.

C. Load-balancing

The concept of load-balancing is that a set of machines \mathbb{S} are grouped as an unique virtual machine \mathbb{M} . If \mathbb{T} is the set of all executable tasks, then load-balancer is a function $l : \mathbb{T} \rightarrow \mathbb{S}$ which attributes to a task t a machine $l(t) = m$ that will execute the task t . So the load-balancer l defines a policy to distribute tasks on \mathbb{M} .

This technique is used by web services, for example, to spread users's requests between servers : several servers look like one for users. When a user sends a request to a web service, load-balancer redirects this request to one of the available servers. This allows to equilibrate work-load of servers and provide a best quality of service.

In distributed programming, it allows to optimize work-load of each machine by establishing a policy to spread tasks on machines. Some load-balancing algorithms can depend on the type of distributed network: this network can be synchronous or asynchronous and its topology can be dynamic.

¹<http://setiathome.ssl.berkeley.edu/>

D. Dynamic Graph

Execution of simulations distributed by DAGDA can be described as a dynamic graph, so the concept of graph and then of a dynamic graph need to be defined.

A graph G is a pair (N, E) where N is a set of elements called nodes and E is a set of nodes-pair (u, v) called edges such that $u, v \in N$. A dynamic graph is a sequence $G_i = (N_i, E_i)$ such that $\forall (u, v) \in E_i, u, v \in N_i$. So, it is a graph that can change over time, by adding/removing nodes and/or edges, but such that if an edge exists at a time i then its ends also exist at the same time i .

Nodes of the graph are the entities of the simulations and edges model interactions between entities. Edges of the graph modeling the execution are weighted : more intensive is an interaction between two entities, more important is the weight of the corresponding edge.

II. ANTCo²

ANTCo² is a distributed algorithm dedicated to load balancing and communication minimisation.

AntCo² considers only the dynamic graph of the application to compute the distribution.

As communications and agents appear and disappear, as the importance of communication evolve, the graph changes. Therefore the load balancer should also handle this dynamic process and be able to provide a distribution as the graph evolve.

Each computing resource is associated with a color, then by assigning a color to a node, the algorithm specify the distribution.

One can see the distribution as a weighted partitioning of the graph. In this partitioning we try to distribute evenly the load (number of entities weighted by their computing demand) and to minimize communications between computing resources to avoid saturating the network. These two criteria are conflicting, therefore a trade-off must be found.

We see the partitioning as a dynamic community detection algorithm. We call such dynamic communities "organizations". Communities are often seen as group of vertices that are more densely connected one with another than with the rest of the graph. An algorithm able to detect organizations is able to follow communities as they evolve when nodes and edges appear, evolve and disappear in the communities.

There exists several graph partitioning algorithms ([4], [5], [6]) and community detection algorithms ([7]), but few handle evolving graphs. It is always possible to restart such algorithms each time the graph changes, but this would be computationally intensive. AntCo² is an incremental algorithm that start from the previous partitioning to compute a new partitioning when the graph changes.

Having a load balancer running on a single machine, to distribute applications that are often very large could be inefficient. Another goal of AntCo² is to be able to be distributed with the application.

AntCo² uses an approach based on swarm intelligence, namely colonies of ants. This algorithm provides several advantages: ants can act with only local knowledge of the

graph representing the application to distribute. AntCo² tries to avoid any global computation, therefore allowing it to be distributed with few communications and no global control.

In AntCo², each colony represents a computing resource and has its own color. Inside colonies, ants collaborate to colonize organizations inside the graph and assign their color to nodes. Inversely, colonies compete to keep and conquer organizations.

Ants color nodes using numerical colored pheromones corresponding to their colony color. Such pheromones "evaporate" and therefore must be maintained constantly by ants. This allows to handle graph dynamics by forgetting old partitioning solutions and discovering new solutions by the constant exploration of ants inside the graph. The details of the algorithm are given in ([8]).

The change of a color for a node indicates a "migration advice", meaning that the corresponding agent should migrate on the computing resource associated to the new color. An inertia mechanism allows to avoid oscillatory advices.

III. DAGDA

DAGDA is a middleware dedicated to the distribution of Complex Systems simulations. It uses an existing middleware as a base which is extended with new features. The final aim is to provide a simple way to create distributed complex system simulation.

The main words of Dagda are decentralized, portable, load-balanced. Decentralized means that there is no restricted set of machines on which depend all machines. Dagda aims to be as portable as possible, is that any machines (computer,pda,phone,super-calculator,...) can participate to the distribution.

The used middleware is ProActive[3]. This choice is motivated by the active object approach which is used in ProActive.

A. Entities

Dagda is based on the concept that the distributed application is composed of a massive set of objects. These objects are called *entities* and are hosted on a machine by an *agency*. Entities are active objects.

Entities can interact with each other and can migrate from one agency to an other. This rises a problem: how to identify each entity through the network and how to get a remote entity ? The second part of this problem, how to get entity, is treated on section III-B. Entities are identified by an id which is unique through time and network. Uniqueness is assumed by the fact that id depends on the agency's address (agency who creates the entity) and on a timestamp.

B. Communication between agencies

Dagda aims to have a decentralized architecture so there is no master server to reference informations as for example entities location. Therefore, some mechanisms are needed to provide functionalities like entity-search.

Each agency has a dedicated active object whose role is to detect other agencies and to provide to user functionalities through connected agencies.

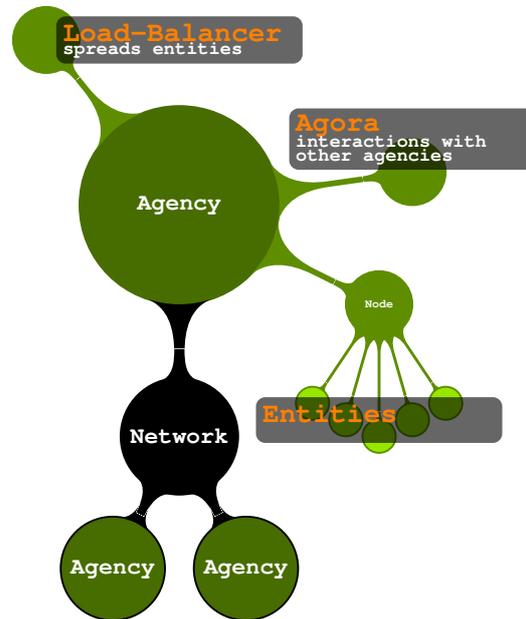


Fig. 4. Dagda overview

C. Context

A program may have some parameters which create a context that is used by components of this program. With programs running on a simple process, this is easy to do by declaring global variables. But with distributed programs, each machine has its own memory and one other can not see changes.

Dagda creates a context divided in two parts. It contains a *local* part which contains parameters that do not have to be shared. Second part is *global* and changes on this part will be spread on all machines. Context users can access to parameters without local or global distinctions.

D. Interactions Graph

Dagda profiles method calls between entities. For example, if an entity A calls a method $m()$ of an entity B, this call will be detected and registered. Then this detection of interactions between entities is used as provider to a dynamic graph which models these interactions through the time. Nodes of this graph are the entities host on the machine and remote entities such that there is an interaction between these remote entities and one of the hosted entities. Edges of the graph represent interactions between entities. Greater is the number of interactions between two entities, greater is the weight of the corresponding edge. There is a mechanism which decreases edges's weight through the time. The GRAPHSTREAM[9]² API is used to create the graph.

This graph can be used by tools, for example to monitor entities activity and have a look on this activity.

²<http://www.graphstream-project.org>

E. Load-balancing

Entities are spread on the available machines with the ANTCO² load-balancing algorithm. This choice allows :

- equilibrate the work-load of machines;
- reduce the network-load;
- distribute the load-balancer.

Distribution of the load-balancer is an important thing to have a decentralized platform. In [10], three ways are presented to run the ANTCO² algorithm. The first and the second one run with a restricted set of servers (size of this set is one in the first case). In these two cases, computing-load of server is fully used and ANTCO² has a global view of the distributed application. The last one uses each machines to run the algorithm. In this case, only a few computing-load is used on each machine and ANTCO² has just a local view of the distributed application. This case allows to decentralize ANTCO², so it has been chosen in DAGDA.

Work-load dedicated to ANTCO² is function of number of entities, so by balancing entities-load, it balances itself : it is auto-distributed.

IV. RESULTS

AT this time, DAGDA is still in development. The platform is able to create entities and profile interactions between them. So it is possible to view the graph of the simulation's execution in real time. It is also able to connect agencies and migrate entities from one agency to another.

A. Test application

To realize some tests, a simple application has been written which aimed to generate interactions between entities and migrate them between agencies. Entities used for this application can be described as follows :

```
TESTENTITY:
attributes :
  List<TestEntity> neigh
methods :
  call( TestEntity te ) {
    while( neigh.size() > MAX )
      neigh.poll();
    neigh.add(te);
  }
execute() {
  int i,j;
  i = random() % neigh.size();
  j = random() % neigh.size();
  neigh.get(i).call(neigh.get(j));
  if( random() < P_MIGRATION )
    migrateSomewhere();
}
```

The application creates a set of `TestEntity` and inits randomly the `neigh` attribute of entities. Then each agency run the `execute()` method of each hosted entity.

Figure 5 shows the graph of the execution of this application with 64 entities.

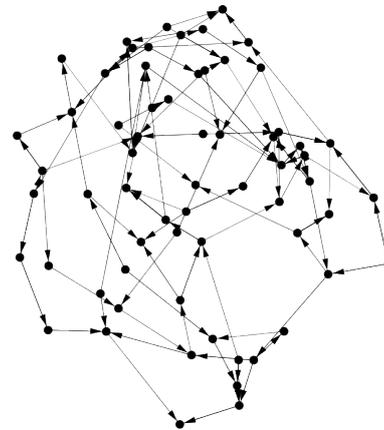


Fig. 5. Execution of the test program

V. CONCLUSION

In this paper, concepts of middleware and load-balancing have been described. Then the DAGDA platform which merges a middleware and the load-balancer ANTCO² has been presented.

DAGDA is still in development but it ables to launch a program and profile execution of this program. Next step is to finalize implementation of the load-balancer and validate the platform making battery of tests.

Then we need to provide an application running on DAGDA and realize tests on performances to show the gain brought by DAGDA.

REFERENCES

- [1] C. Hewitt, P. Bishop, and R. Steiger, "A universal modular actor formalism for artificial intelligence," in *IJCAI*, 1973, pp. 235–245.
- [2] D. P. Anderson, "Public computing: Reconnecting people to science," in *Conference on Shared Knowledge and the Web*, Residencia de Estudiantes, Madrid, Spain, Nov. 2003.
- [3] L. Baduel, F. Baude, D. Caromel, A. Contes, F. Huet, M. Morel, and R. Quilici, *Grid Computing: Software Environments and Tools*. Springer-Verlag, January 2006, ch. Programming, Deploying, Composing, for the Grid.
- [4] B. Kernighan and S. Lin, "An efficient heuristic procedure for partitioning graph," *The Bell System Technical Journal*, vol. 49, no. 2, pp. 192–307, 1970.
- [5] B. Hendrickson and R. Leland, "An improved spectral graph partitioning algorithm for mapping parallel computations," *SIAM J. Scien. Comput.*, vol. 16, no. 2, pp. 452–469, 1995.
- [6] C. M. Fiduccia and R. M. Mattheyses, "A linear time heuristic for improving network partitions," in *ACM IEEE Design Automation Conference*, 1982, pp. 175–181.
- [7] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev.*, vol. 69, 2004.
- [8] C. Bertelle, A. Dutot, F. Guinand, and D. Olivier, "Organization detection for dynamic load balancing in individual-based simulations," *Multi-Agent and Grid Systems*, vol. 3, no. 1, p. 42, 2007. [Online]. Available: <http://litis.univ-lehavre.fr/~dutot/biblio/MAGS2007.pdf>
- [9] A. Dutot, F. Guinand, D. Olivier, and Y. Pigné, "Graphstream: A tool for bridging the gap between complex systems and dynamic graphs," in *EPNACS: Emergent Properties in Natural and Artificial Complex Systems*, 2007.
- [10] A. Dutot, "Distribution dynamique adaptative à l'aide de mécanismes d'intelligence collective," Ph.D. dissertation, Université du Havre - LIH, 2005.

GeoCells model : European Structural Funds and regional interactions, which convergences for the European regions?

Bernard Elissalde

UMR CNRS IDEES (6228). University of Rouen (France),
bernard.elissalde@univ-rouen.fr

Patrice Langlois

UMR CNRS IDEES (6228), University of Rouen (France)
patrice.langlois@univ-rouen.fr

Dominique Goyat

UMR CNRS IDEES , University of Rouen (France)

RÉSUMÉ.:L'objectif de cette communication est de présenter le projet de développement d'une plate-forme de simulation de processus géographiques (Géocells), permettant de modéliser le comportement des régions européennes en fonction des aides accordées par l'Union Européenne. Le schéma d'ensemble de la plate-forme repose sur l'utilisation des techniques de simulation par automates cellulaires. La prise en compte des disparités régionales par les politiques structurelles de la Commission européenne, l'analyse des positions relatives des régions européennes sous l'angle des indicateurs macroéconomiques et budgétaires, va permettre d'évaluer l'efficacité globale des Fonds structurels européens, de mesurer l'influence des modifications des règles d'attribution sur les trajectoires des différentes régions aidées, évaluées par des indicateurs de convergence.

ABSTRACT : The aim of this paper is about the trends of regional disparities in the European Union who can be considered as a complex system.. For modelling the uncertain efficiency of the regional policy we attempted to use a cellular automata (Geocells) developed by P.Langlois). Methodologically, this cellular automata Geocells is based on interrelated processes between variables (like time periods, growth rates in the GDP per head, flows of public investments) and three geographical levels (european level, national level and regional level). This three levels are used to lay the emphasis on the fact that the EU structural expenditure, and its spatial impact, work as a rules-based system. Simulations were made, in order to evaluate, on the one hand, the specific role of each level and each variable, and on the other hand how some change in one part affects the whole. In this perspective we could underline the role of increasing or decreasing budgets, the weight of national policies and national economic trends and the contiguity effect according to the geographical location of each eligible region.

MOTS-CLÉS :automates cellulaires, politique régionale européenne, disparités régionales, régions

KEYWORDS: cellular automata, regional policy, regions, regional inequalities

1. Introduction.

The aim of this paper is to present the functioning of GeoCells : the geographic simulation platform, as well as a first application enabling to model the European regions' behaviour according to the variation in aid granted by the European Union and to neighbourhood effects. The platform's overall diagram relies on the use of cellular automata techniques.

The taking into account of regional disparities by the European Commission's structural policies, the analysis of European regions relative positions from the angle of macroeconomic and budgetary indicators, will allow an evaluation of the overall effectiveness of the aforementioned policies, and to measure the influence in the modification of granting rules. The different settings offered by the simulation platform ensure to simulate both the impact of aid on the overall evolution of the Europe of the Fifteen regions (measured by indicators such as beta or sigma convergence), as well as on each region's specific future.

The introduction of simulation and forecasting methods in the debates on the European Union's regional policy does not aim to enable to find the one and only response to the problem of European regions' unequal development, but to suggest a range of credible options as a decision support tool for territorial solidarity and economic and social cohesion, in a European space in perpetual evolution. Even though they belong to an interdependent

group such as the European Union, these spatial units each own their specific trajectories, in which the reaction delays, the transformation rhythms strongly vary from one region to another. All of these differentiated laps of time will build a European regional mosaic, making it unlikely to happen a mechanical adjustment between the impulsions of the European Union's regional policy (Structural Funds, Cohesion Funds, etc.) and the regional readjustment initially planned.

2. Community solidarity policies and the regions' future

The issue of the solidarity effort between State members and the regions, as well as their adequacy to the cohesion principles displayed in the European texts and treaties is at the centre of the debates on European regional policy. The main questioning is about the European public policies' ability to adjust disparities produced by the single market. How can we improve redistribution and territorial equity in an Union of low economic growth ? In such an economical context, should we limit the solidarity effort of rich countries or on the contrary emphasize it in order to accelerate the catching-up of backward regions ?

2. 1 *The issue of allocated Funds' effectiveness*

In order to evaluate the effectiveness of passed programmes and to discuss the necessity of possible reorientations in the granting

of structural Funds within the European Union, it is essential to recall budgetary stakes and a few functioning principles of European intervention as far as regional policy is concerned. Since the creation of the ERDF in 1975, redistribution rules and amounts granted, within the framework of what we put behind regional policy have changed during the last twenty years and have not affected the same regions during the same periods of time. The general conditions of eligibility have changed over the reforms, and individually for each region, because economic results made them eligible or not from one period to another. Various reforms in 1984, then in 1988, carried along by the imperative of economic and social cohesion, have changed the amount of granted budgets. A Cohesion Fund concerning four countries (Spain, Portugal, Greece, and Ireland), was born subsequently from 1994 onwards, in order to offset these countries' budgetary effort, with the aim of a convergence towards criteria of participation to the Economic and Monetary Union. These so called cohesion countries have received ever since aid of a substantial weight in their GDP.

For many years, surveys have tried to evaluate the consequences of the European economic integration consolidation over regions, as well as the effectiveness of the various European Structural Funds' reforms¹. They all agree to admit, at country scale, the alternation of catching-up stages, notably for the Cohesion countries (Spain, Portugal, Greece,

¹ July 1997 and March 1999 CEPII journal numbers

and Ireland), more particularly for Ireland which has reached the European average from 1997, with more uncertain stages. Regional convergence would have known a slowing down in the first half of the 1980's, then an improvement at the end of this same 1980's, followed by an overall divergence period in the 1990's. This changing diagnosis, at European scale, is, in most cases, tempered by a maintaining, or even an increase in disparities between regions of a same country. Since regional growth dynamics are not exactly the same as the ones at a national level, we have been wondering after the works of Martin (2000), Maurel (1999), and Riou (2002) about the spatial consequences of the European economic integration and about the reasons of an unequally distributed and polarizing growth. The conditions for the reduction of this gap between the scopes of interregional solidarity and territorial cohesion and reality, outlined in many official texts such as the European Spatial Development Perspective or the Second report on economic and social cohesion (Eurostat 2000), as well as in the results of our previous works (Elissalde 2005) were the purpose of different simulations achieved in this paper.

In addition to the issue of European Structural Funds' effectiveness, this shift between scale levels makes it wonder in the future about the solutions to mobilize in order to bring down the development gaps significantly increased by the progressive transition from fifteen to twenty seven members in the European Union. The use of a

simulation platform through cellular automaton aims at answering the question of knowing on which conditions (of settings in terms of budgetary redistribution), according to which duration of financial aid programs, and according to which objective levels of reduction, convergence, or adjustment, European solidarity policies could be effective.

Two hypothesis can be considered: either the results of financial aid programs at European level are due to other factors such as the differentiated effects of each country's particular national situations (lets not forget that regional aid work according to the principle of additionality), either the selectivity of growth dynamics would be due to the region's relative position inside a wider spatial unit, affected itself by a

situation of growth or stagnation. With regard to the analysis presented above, we have therefore added the hypothesis that neighbourhood effects between regions played a role in the spreading of growth or in the reduction of disparities.

For all these reasons, we join R. Geyer [2003], in order to consider that the European Union shows certain specific characteristics of a complex system. The European Union's spatial dynamic would thus begin to look like a partly auto-organized system where the reactivity of spatial units would behave in a more or less autonomous manner in relation to the redistributive impulsions of the European Commission, which on its side assumes and tries to demonstrate in its reports on cohesion, the fundamental importance of regional aid's immediate positive effects

2. 2. The issue of regions' convergence

In order to evaluate the consequences of different parameters settings noted above on simulation results, we have used the concept of convergence and two indicators in order to refine diagnosis on the reality of disparities reducing between European regions. Following other economic works, Beine and Docquier (2000) suggest three declensions of this idea of convergence :

- absolute convergence which supposes that per capita income would converge towards each other

independently from initial conditions and led policies

- conditional convergence which puts forward the hypothesis that identical territories in terms of demographic growth, public policies, but with different initial conditions, are supposed to converge towards the same stationary state the ones according to the others.

- « club » convergence which puts forward the hypothesis that convergence is not achieved in a global manner but by group of regions. Per capita incomes would

not converge at an identical pace for all regions but through groups of regions, which would distance themselves by an original growth dynamic.

As for evaluation, we adopted the distinction of X.Sala-i-Martin (1996) which opposes:

- sigma convergence C_σ measuring the evolution of the GDP per capita over n years, calculated as the annual variation gap of the coefficient of variation between two periods t_0 et t_n , where m_i is the GDP mean of the year i and σ_i its standard deviation. Thus we have:

$$C_\sigma = \frac{\frac{\sigma_n}{m_n} - \frac{\sigma_0}{m_0}}{n}$$

- beta convergence C_β which measures the relation between the GDP per capita (logarithmical) variation over a given period compared to the initial level.

If we call:

$$x_k = \log PIB_{k0}$$

the GDP neperian log of the region k for the initial year t_0 and

$$y_k = \log(PIB_{ki}) - \log(PIB_{k0}) = \log\left(\frac{PIB_{ki}}{PIB_{k0}}\right)$$

the GDP logarithmical variation of k between t_0 and t_i , beta convergence between two periods t_0 et t_n , is then defined by the slope a of the straight line $Y = aX + b$ of the variable $Y = \{y_k\}_{k=1,\dots,K}$ in relation to the variable $X = \{x_k\}_{k=1,\dots,K}$. Therefore, we can write:

$$C_\beta = \frac{Cov(X, Y)}{Var(X)}$$

Lets note that the more these indicators are negative, the more they indicate a better convergence.

The diffusion of wealth between regions (if we suppose it exists) tends to make the regions' standards of living converge. This mechanism is opposed to the regions' internal growth, which increases exponentially from these wide range of wealth levels between regions. This growth then tends to make them diverge. The combination of these two effects is generally to the benefit of growth. Even though the growth of the wealthier improves the wealthiness of the neighbours, the gap deepens nevertheless. The financial aid policy towards the most underprivileged regions seeks therefore to counterbalance the predominant effects of this divergence. Lets note also that the way Europe measures the "convergence" (through beta-convergence) allows in theory to mask the exponential widening of absolute gaps between regions using relative gap logarithms... However, simulations show that the behaviour

of these indicators are not so simple to interpret.

These indicators were already used in the research works of Charleux (2003) and Le Gallo (2006) and initially in France by F.Maurel's team (1999) from the General Commissariat for National Planning in order to test the hypothesis of an « absolute » catching-up of the most underprivileged regions, then by introducing complementary

explanatory variables (ex: facilities endowment) in order to evaluate their weight in growth regional gaps. These calculations carried out by comparing separately convergence tests at States' level and at regions' level lead simultaneously to the empirical observation of a convergence between countries and of a regional divergence internal to countries.

3. *The GeoCells project*

4.

4.1. From the strict notion of cell to the complex spatial agent's one

The GeoCells project is part of the continuity of two experiences developed by our laboratory (MTG Rouen, UMR IDEES) [Langlois et al 02], [Guermond et al 03], [Guermond et al 04].

First of all the SpaCelle software, which is a cellular automaton based on rules dealing only with qualitative spatial variables, followed by another very different cellular model, allowing to simulate continuous flows such as surface flow in hydrology called the RuiCells model.

The strict formal concept of cellular automata defines a cell as purely reactive. A cell in the broad sense does not know how to do anything else but to change its internal state according to the perception of its environment and of its own state. Nevertheless, in a wider sense, the notion of cell can be

understood in its analogy with the living world. Moreover, its based on this analogy that von Neumann defined the term of cellular automaton, in order to try to model the phenomenon of living agents' reproduction. In this sense, a cell is not only reactive, but becomes an agent which can own a complex behaviour. A cell must be delimited in space by its membrane, which is both a frontier separating the inside from the outside, and a filter controlling solid, energy, and information flows with its environment.

In addition, we often need to define spatial units of any forms, because this is the way that geographical information is available : agricultural plots, land registry, urban units, etc. Thus we have tried to define directly a behaviour (a dynamic) to spatial units of any forms. The polygonal form must thus be directly usable as a geometric support of a cell in a geographic simulation system.

(Batty 2005) (Longley1996)(White 2004)

4.2. The GeoCells project, an automaton to simulate the regional policy

The purpose of the GeocCells project is to create a simulation platform based on geographic information layers. Its main driving force is a geographic model generator based on topologic cellular agents. –

From a model description considered as representative of an issue in association with a real or virtual geographic environment, the user builds, through this platform, an automated structure allowing this model to live, in order to see its evolution and to check its pertinence level.

The overall functioning principle of GeoCells is given by the following diagram :

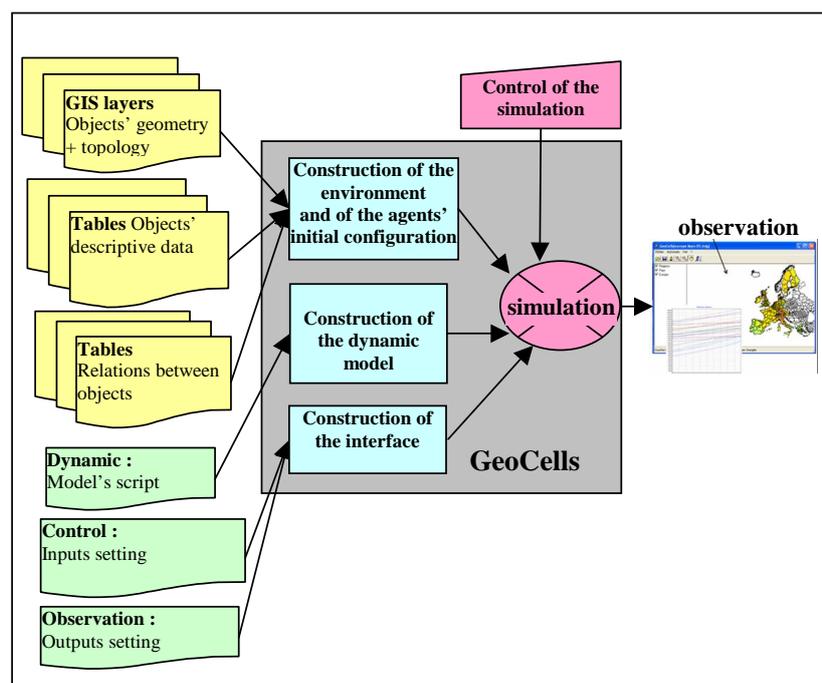


Figure 1. GeoCells functioning principle

The different distinct parts visible on the diagram match the following stages :

1 - The user calls for a set of static information used as a support to the model and thus describing its initial physical environment

2 - The user, relying on the above data, provides the desired model dynamic, that is to say the group of action features peculiar to the simulation and relating to the simulator-user interaction,

3 - GeoCells translates this information and generates an automated model and a command interface,

4 - the user communicates by means of the interface in order to change certain parameters relating either to the simulation or else to the ones linked to the results display.

4.3. Basic principles

a) Cellular structure

The system is based on a group of geographic information layers. Each

layer is made of features from a same class. To each layer of information, and to each feature of this layer, matches a cell, which main asset is to own, in addition to the feature's physical components (location, shape, size...), the knowledge of its neighbourhood and above all its behaviour dynamic.

Each cellular class owns :

- Behaviour rules giving to the cells of its class the same function in the system (district, plot, department),
- properties and attributes (perimeter, surface area, budget...),
- relations with cells from other layers of the system.

b) Hierarchized structure

The system takes into account the hierarchical relations existing between layers, such as the ones explained on the diagram below (a district [layer 1] belongs to a department [layer 2] and the department is made up of districts...), but nothing prohibits to implement other relations between cells of different layers, such as for example transport connections, etc.

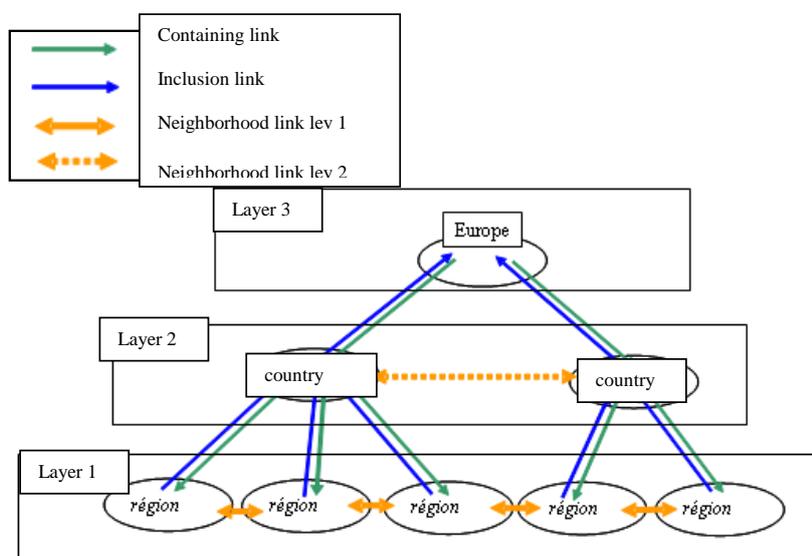


Figure 2. Hierarchy of cellular layers

4.4. Generated automata characteristics

4.5.

Time definition - In order to scan all cells of all layers, therefore to give them the means to carry out their actions, the system takes a certain (physical) machine time that depends on the performances of the host computer. This proceeding time matches the *lapse of time* which is the system's temporal reference (logical time). We can therefore build from this reference a time unit adapted to each simulation.

Synchronization - One of the difficulties of this type of mechanism is to maintain the temporal coherence between every cellular layers. In order to do so, the automaton synchronizes the system by ordering itself the starting of each cell's life

stage. This synchronization applies to :

- all of the cell's input reading (incoming flows),
- the execution of all of the cell's peculiar processes, with according to the circumstances, change of its internal states
- all of the cell's output writing (outcoming flows).
- saving of all current contexts.

Communication canals - An unidirectional communication canal feature was introduced when it was necessary to implement the system's multilayer nature with the flexibility required by exchanged data's multiform nature, combined, according to the models, with the possible plurality of flows. This

feature is automatically generated and sized. Each cell owns the input and output references relating to the canals that concerns it, and this is true for each communication action (informational and/or energetic). For this reason, it has the knowledge of its environment and enters into dialogue with it.

Command interface - According to the needs, each of the model's influential magnitude is combined with an interface component, for example, in the form of a cursor, giving the opportunity at initialization to change its value. All of these components, created dynamically, associated to a certain number of structural dialog components (menus, etc.), makes the system's command interface, the simulator's control panel.

4.6. GeoCells' field of action

GeoCells can be used in any model that brings into play phenomena where the spatial component is predominant. Among these phenomena, we can cite the diffusion or the propagation of magnitudes (physical or not), whether they are generated by rules of neighbourhood-contact (ex : frontiers between cells of a same layer) and/or by rules of neighbourhood-transport, by inserting between layers one or more transport layers (linear topology) and access relations that are associated with them. We can also take on rules relating to virtual exchanges of a hierarchized type or not. In the following example, we

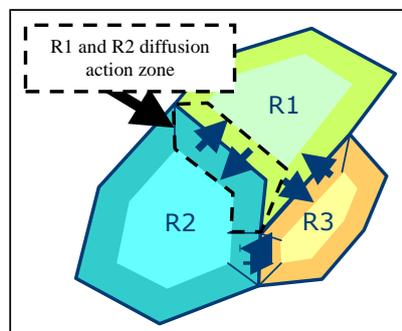
carry out several hierarchized economic exchange flows.

The GeoCells model applies to the evolution of the GDP per capita in the Europe of the 15 and to the influence, in this evolution, of the game of the various aid granted as regional policy (ERDF). The platform is made up of three topological layers:

- Administrative regions level NUTS2 (512 cells : 511 regions + 1 cell representative of the outside)
- Countries (17 cells : 15 country cells + 1 cell rest of Europe + 1 « external » cell)
- EU (3 cells : 1 cell Europe of 15 + 1 cell rest of Europe + 1 « external » cell).

5. Model and simulations

5.1. Description of the growth-diffusion model between European regions



We will now clarify the diffusion model a bit unusual that we have used. Let's note X_i the GDP of the region i , P_i its population and $Y_i = X_i/P_i$ its GDP per capita at a moment t .

We put forward the following hypothesis. Each cell has the **aim** to homogenize, through time, its standard of living Y in relation to its neighbours. But the standard of living is not comparable to a physical magnitude capable of diffusing like a flow. It is through the variation of wealth (X) symbolized by the GDP (by internal growth and by diffusion) or through the variation of population (P) (also by internal growth or by

migrations) that each region can work in order to achieve its goal. In this model, at first approximation, we have considered that population was constant throughout time. It is therefore here only on the variation of X that relies the diffusion mechanism in order to reach the goal.

Another hypothesis is to consider that only a small fringe close to the borderline (area in dotted line on the figure) takes part in the diffusion of wealth, by the levelling-out of standards of living of the two neighbouring border fringes. Since we do not have any information on the spatial distribution of the populations inside a region, we must put forward the hypothesis of a uniform distribution. Consequently, instead of launching forth into geometric calculations of insane zoning, we use a simple proportionality parameter, accessible in the user interface by a cursor, called **diffusion rate**, which sets the rate k (of surface area, population, and wealth all at once, since we consider them as uniformly distributed over the region's surface area), which takes part in the diffusion between regions. This rate defines therefore the part of the region's surface area matching the border fringe, in which the standard of living is going to attempt to level up through time, with the neighbouring regions' homologous fringes.

In order to model the diffusion between two regions i and j , we then introduce the coefficient k_{ij} which is the surface area's proportion i matching the intersection between the border fringe defined by k and the

proportion p_{ij} of its land borderline shared by the region j , defined by

$$p_{ij} = \frac{l_{ij}}{\sum_{k \in N(i)} l_{ik}}, \text{ where } l_{ij} \text{ is the}$$

borderline's length between i and j .

We then have : $k_{ij} = k \cdot p_{ij}$

If the wealth on the two sides of the border fringe between i and j was evenly distributed like connected vessels, we would obtain a levelled-out standard of living (which is not the average of the two previous standards), defined by:

$$Y_{ij} = \frac{k_{ij} \cdot X_i + k_{ji} \cdot X_j}{k_{ij} \cdot P_i + k_{ji} \cdot P_j}$$

We can then define the variation dX_{ij} (positive if it emits or negative if it receives) of the diffusion from the region i towards the region j during a short lapse of time dt as being proportional to the concerned population ($k_{ij}P_i$) and proportional to the difference between the current standard of living (Y_i) and the (local) aim of levelling-out (Y_{ij}) of standards of living i and j . This can be translated into the following equation :

$$\frac{dX_{ij}}{dt} = K \cdot k_{ij} \cdot P_i (Y_i - Y_{ij})$$

The value of K is set internally (since we can already play on k).

By adding the border fringes of the region i , we note down:

$$dX_i = \sum_{j \in N(i)} dX_{ij}$$

One should notice that this diffusion is, by construction, preservative of the mean $\sum_{i=1}^n X_i$. (because one can verify easily that for any couple (i, j) we have : $dX_{ij} + dX_{ji} = 0$)

Moreover, the variable X_i is subjected to an a priori exponential internal growth, $\frac{dX_i}{dt} = C_i X_i$

Internal growth is adjustable, either individually region by region through the attribute table, either on the whole as being the same for all regions with the help of a cursor present in the user interface.

The final growth-diffusion equation is thus given by :

$$X_i(t + dt) = X_i(t) + (C_i \cdot X_i(t) + K \cdot k_{ij} P_i (Y_i - Y_{ij})) dt$$

The lapse of time for the discretization of growth and diffusion processes are small compared to redistributing flows, because they correspond to continuous processes. We have selected the month as lapse of time, that also matches the time unit that we chose, so $dt=1$. (C_i is then the twelfth of the annual growth rate).

The equation with this lapse of time is then written :

$$X_i(t + dt) = (1 + C_i) X_i(t) + K \cdot k_{ij} P_i (Y_i - Y_{ij})$$

The model introduced here attempts to give an account of the crossed recursion of the processes' effects contributing to regional

dynamics : the region's peculiar growth, the redistribution mechanism linked to its membership to a wider territorial group (State, Union), and finally neighbourhood effects. Many economic models try to isolate the various sectoral variables in the final growth. Here, we suppose that the three components presented above create system effects introducing a large part of uncertainty in terms of growth.

We will not describe in detail the part of the model concerning aid and contributions, insofar as they show through clearly enough in the settings of the user interface.

5.2. Simulator's settings

Given the data available for the group of regions NUTS2 of the fifteen member States European Union, the ij model retained, as the main indicator, the variation of the GDP per capita of each European region. The variation of this magnitude linked only to the variation of the GDP (we have made the choice of a constant population), is subjected within the platform to several influences adjustable for each simulation:

- The GDP variation rate is, either specific to the region, either identical to the group of regions of a same country, either, by simple hypothesis, identical for the whole group of regions.
- The terms of public intervention include the mechanisms relating to contributions (Countries and EU), to the aid linked to regional policy, such

as eligibility thresholds (75%) for Structural Funds.

- The European budget weight was taken into account, stabilized around a threshold of 1% of the European total GDP since fifteen years (threshold reached since 1984). From this average budget, simulations were able to make the Community budget weight vary from 0,5% to 3% of the EU total GDP.

- The principle of additionality between the States and the European Union in the Structural Funds financing was also taken into account, as well as the variability of the relative importance of regional policy in the Community expenditures.

- Finally, the rule of 4% maximum weight of European aid in the GDP of a region or of a State was applied.

- To these principles officially ratified by the European Commission, we have added to our model a spatial dynamic parameter: the hypothesis of the role of spatial interactions and of contiguity effects in the regions' trajectories.

The diffusion by contact with neighbouring regions, made possible

5.3. Test of the simulator

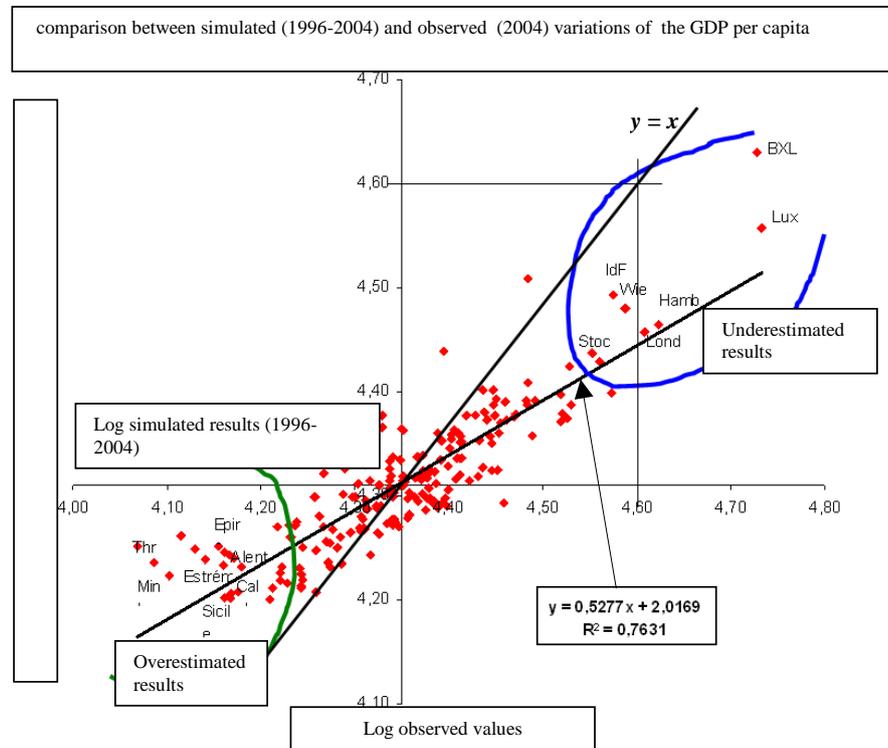
In order to test the validity of the model, one of the first priorities was to attempt to "calibrate" the simulator's results in relation to the regions' real variations. Therefore, we have compared, given the most recent data available, the 2004 GDP

by the functioning of the cellular automaton, is carried out therefore naturally in one way or another. Many regional growth models analyse the region as a stand-alone unit and ignore spatial interaction phenomena linked to proximity, neighbourhood, or contiguity effects. What is happening in the neighbouring regions is ignored, while sensitivity to exchange and migration distance is very large. Now, many works have shown that economic interaction and territorial interaction acted in a multiplicative way (Heylen.C et al, 2001), and some empirical assessments evaluating the spatial auto-correlation degree between European regions, as far as GDP per capita is concerned, confirm the pertinence of the reasoning process (Elissalde, 2005). The existence of territorial cores matching either regional areas or national spaces having similar development characteristics and trajectories corroborates this idea. While a situation of spatial competition between activities and between territorial units exists, the taking into account of contagion, of mimicry phenomena, of power struggles linked to neighbourhood effects proves to be necessary.

per capita actual results with a simulation for the period 1996-2004 (see diagram figure 5). It emerges from the comparison between simulated and observed variations that the platform shows an important degree of credibility, both in its overall results (values of the coefficient of determination) and for regions taken individually. For

certain settings of the simulator, the differences between the simulated and the observed variations, reveals a trend towards significant results. They underestimate the big metropolitan regions' final values (Brussels, London, Hamburg, Paris area, etc.), and on the contrary overestimate the less-developed regions' ones (Alentejo, Epire, Calabria, Estremadura). These results reproduced on the diagram match settings including a 10% GDP per neighbourhood diffusion rate. By repeating simulations over the same period with a diffusion rate reduced to 5%, but with a doubling of the Community budget (2% of the GDP), added to a substantial increase of regional aid in the aforementioned budget, an overall result is reached, which, this time, underestimates, with one or two exceptions, the entire values of the regions' GDP per capita. The more the GDP diffusion rate is decreased from one region to

another, the more the underestimation is important. Several interpretations can henceforth be suggested. It seems that the simulator gets closer to reality, when it includes an important degree of permeability from the neighbouring regions GDP (through the diffusion rate), giving indirectly an account of exchanges and interdependencies between them. Moreover, it seems necessary to wonder about the overall impact (and not only at a particular region's level) of structural policies in relation to the co-variation general dynamics of the Europe of the fifteen entire group of regions. From these results, it emerges that the reducing of disparities are potentially plausible for regions that are eligible for European Funds, and that the options of European regional policies comes within a choice between egalitarianism by readjustment, equity without hierarchical upheaval, and lack of solidarity.



6. Simulations and impacts of regional policy in the Europe of Fifteen

6.1. Mixing of variables and political options

Over four hundred simulations were achieved from 1996's population and GDP values. They attempted to evaluate the relative weight of the simulator's variables, the impact of budgetary variations, and can be grouped together in three main categories of scenarios. This type of reasoning concurs with the

one achieved from "qualitative" scenarios about the demands of a polycentric development in order to attempt to weaken the weight of the "Middle-European ridge".

a) The first scenario (simu 1) is the one of free competition between regions without the intervention of Structural Funds. It is tantamount to abolishing European "interventionism" and to "renationalizing" aid, just as recommended in the Sapir Report. Simulations include a GDP growth

specific to each region and a diffusion rate by neighbourhood, as well as a redistribution rate of 20% or 30% of GDP. The difference of growth rates has few effects, on the other hand the redistribution rate by neighbourhood has consequences on the diffusion of prosperity. The readjustment strongly depends on the level of wealth transfers by neighbourhood, whatever the mechanism. In other words, a backward region will have more chances of developing itself if it is located near an already thriving region.

b) The second scenario (Simu 2) consists in searching out territorial equity.

Territorial equity includes ideas of parity of treatment, equivalence of access, and more generally of solidarity between territorial groups, with what it takes in terms of public action, especially by bringing

corrective as far as resources and facilities are concerned. The project consists in endowing each region or each region with a measure of autonomy with the necessary conditions to its development. Structural Funds are used alone, by magnifying the part devoted to regional policy to 1,5% of GDP, but by limiting the diffusion phenomena by neighbourhood to 5 or 10%.

c) Finally, a last approach offers a selective allocation of aid (Simu 3). It consists in searching out results more than automatic distribution of resources, granting compensations to the most underprivileged units. Simulations are achieved with growth rates specific to each region, bringing Structural Funds aid to 2% or 3% of the EU's GDP, according to the usual eligibility criteria, but by stopping neighbourhood effects. This hypothesis of watertightness between regions attempts to isolate what could be a pure effect of European aid.

Table 1: Results of 400 simulations from the GeoCells platform

	Coeff. of correlation	Beta convergence	Sigma convergence
Simu1 (see details of settings in the text)			
10 years	-0,64	-0,343	-0,014
20 years	from -0,50 to -0, 531	from -0,560 to -0,590	0,000
25 years	-0,476	-0,70	0,004
Simu2 (see details of settings in the text)			
10 years	from -0,631 to -0,695	from -0,456 to -0,500	-0,009
20 years	from -0,519 to -0,592	from -0,661 to -0,731	0,002
25 years	from -0,463 to -0,533	from -0,773 to -0,846	0,005
Simu3 (see details of settings in the text)			
10 years	from -0,54 to -0,55	from - 0,414 to -0,429	from -0,009 to -0,008
20 years	from -0,41 to -0,42	from -0,582 to -0,606	0,003
25 years	from -0,35 to -0,36	from -0,675 to -0,696	0,007

In order to evaluate their weights in regional trajectories, these simulations were based on

associations of variables between, on the one hand settings between endowment parameters cited above and coming from European regional

policy, and on the other hand on the neighbourhood effects and on the duration of redistribution phases. The results obtained in the above chart emphasize:

- Whichever the scenario, a strong sensitivity to the duration of redistribution programs, more than to the percentage of European budget (within a range evolving from 0,5% to 3% of the Europe of the Fifteen GDP) devoted to Structural Funds. While periods of regional aid programs take place more often over periods of five or six years added to repeated adjustments in allocation criteria, beta convergence values give results showing the strongest values of tendency reversal, therefore the most significant as far as catching-up is concerned, only after 20 or 25 years of actions.
- convergence indicators do not obtain the most performing scores for

the same scenarios. The convergence measure based on the variation of the statistical dispersion (sigma convergence) is more important for scenario 3 (simu3), relying on the regional policy's budget increase, while the prospect of a catching-up (beta convergence) is more credible with the scenario of territorial equity (simu2).

- the « liberal » scenario (simu1) of the diffusion of prosperity only through neighbourhood effects and without European aid seems very little discriminatory. Given equivalent temporal sequences, the values are supplanted by the ones from the two other scenarios. It is only over a period of a quarter of a century after possible substantial changes of the regions' activity profiles that results become equivalent.

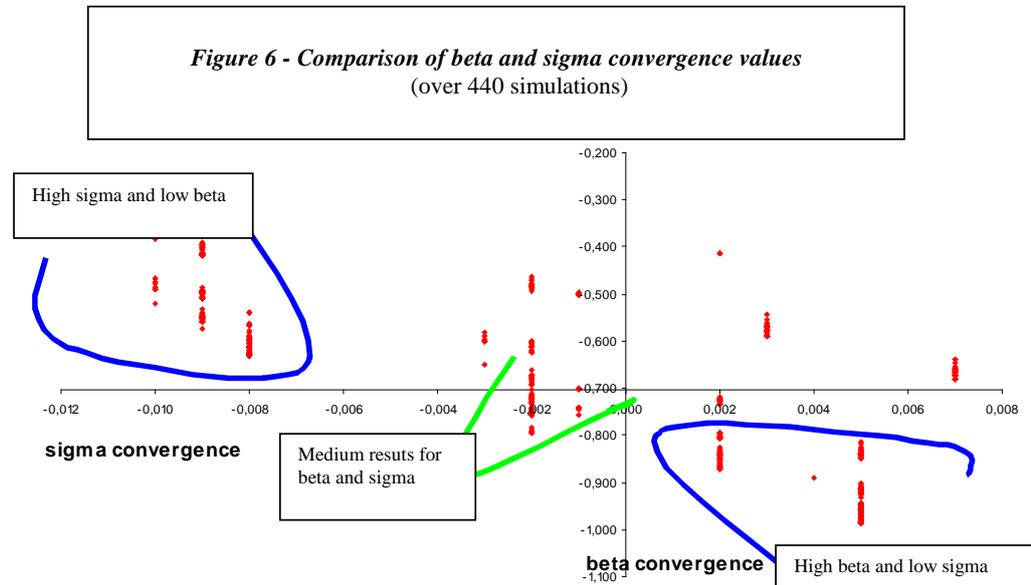
Chart 4. Statistical indicators of simulation results according to the importance of European budget in percentage of the EU's total GDP

		Beta convergence	Sigma convergence
Budget at 1% of GDP (1996)	mean	-0.698	-0,0012
Budget at 1,5% of GDP (1996)	mean	-0,701	-0,0007
Budget at 2% of GDP (1996)	mean	-0.707	-0,0007
Budget at 2.5% of GDP (1996)	mean	-0.717	-0,0006
Budget at 3% of GDP (1996)	mean	-0,720	-0,0005

5.3. Towards an optimization of the two convergence indicators via simulations ?

If multiple regression calculations achieved above brought out the most discriminating variables, the next aim of the work consists in selecting specific settings for each one of the four variables that lead to the most

effective results for the two aspects of the convergence idea. Knowing that the two chosen indicators give an account of various convergence process, reduction of dispersion and disparities for the sigma indicator and improvement of development levels through time for the beta indicator, variations in settings lead to more or less close or more or less far, not to say completely opposite results.



By reproducing on the above diagram the variation in values taken by beta and sigma convergence over a series of more than 400 simulations, it comes out that there are no simulation for which settings lead to high values **both** for sigma and beta convergence. On the contrary, opposite values, divergence for an indicator, convergence for the other, are the most frequent. While high values for both would correspond, on the contrary, to an ideal situation of regional growth and of convergence for the entire group of regions.

This apparent paradox results from each indicator's and each variable's functioning modes. When diffusion mechanism by contiguity are activated between two regions, the weakest receives an additional contribution, but on its side, the strong region carries on to develop itself. In other words, there is not automatically a reducing of disparities (sigma) even if there is progression of every regions compared to the previous situation (beta around -0,9). Conversely, when by assumption, neighbourhood effects are almost nonexistent, weak regions are simulated externally only by the aid of Community Funds, the

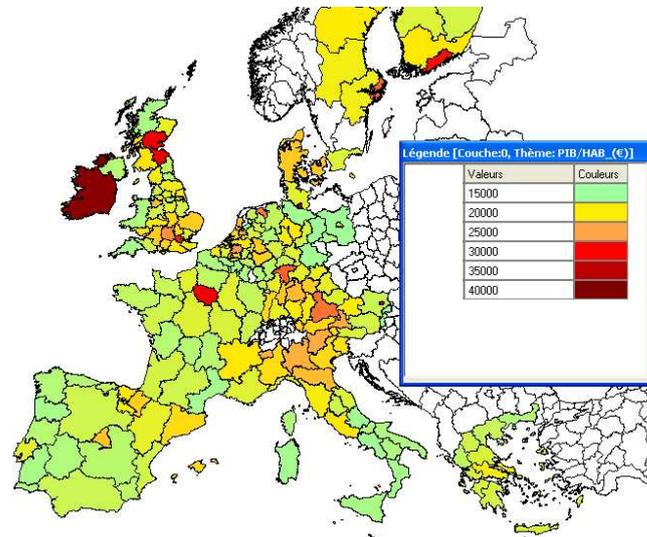
overall gap between regions decreases but with a beta convergence that does not improve much (values around -0,4). We find here again the strongly discriminating impact of neighbourhood effect (already pointed out with multiple regressions) compared to the Community aid factor.

When a certain proximity exists between two indicators, it is significant to notice that it does not match necessarily the maximum settings offered by the four variables. The tendencies towards an (ideal) complete convergence does not result, according to our simulations, from the one and only increase of the Community budget and from the allocated Funds to the regional policy, but seems more likely due to particular mixing that take into account interregional regulations. Hence the sometimes deceiving or “unexpected” impact of structural policies which does not bring the best effectiveness in terms of reducing of regional disparities within the EU.

These efficient mixing are the indication of a multicriteria complexity which relies on specific proportions between Funds levels and the average durations of Community aid (unlike short and standardized durations of current programs), and above all on the taking into account of the interdependency between neighbouring regions. For the reason of the co-evolution of the European regions system’s various magnitudes, seldom are situations favourable to complete convergence processes of wide regional groups.

6. Which configuration for territorial cohesion ?

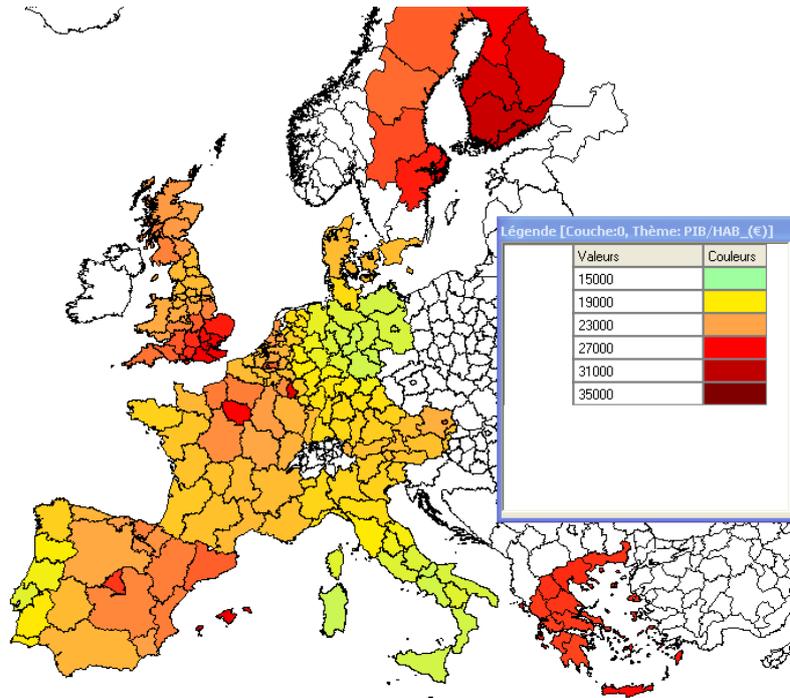
The cartography of the various types of simulation gives concrete expression to the impact of territorial cohesion modes’ geographic distribution chosen by each option :



Map 1 - GDP per capita value after a simulation over 10 years without diffusion

This first map reproduces the result of a situation where values of beta convergence and sigma convergence evolve in a completely opposite way. The simulation concerns a Community budget brought up to 2% of the GDP during a 10-year period with, by hypothesis, a total lack of GDP diffusion by contiguity between neighbouring regions. This hypothesis of regional growth's watertightness gives a good

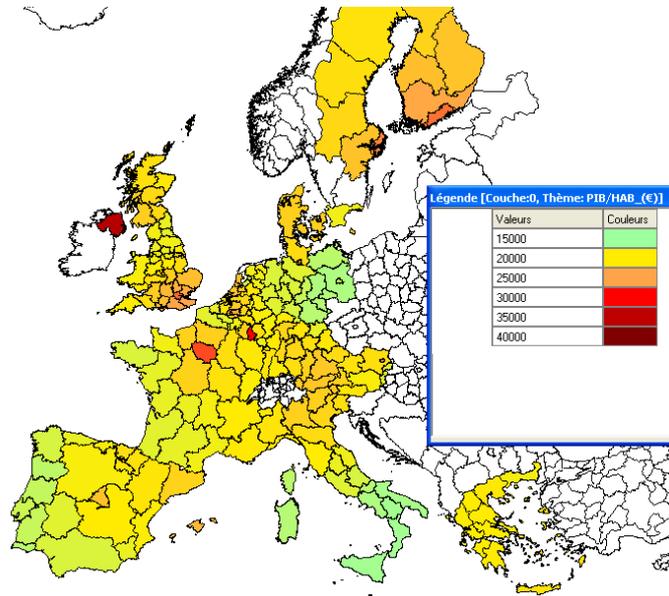
sigma with a low dispersion of incomes between regions, since backward regions saw their GDP per capita rise up, but, on the other hand, a bad beta convergence, since the GDP per capita of the entire group of regions, except for a few metropolitan regions, increased only slightly.



Map 2 - GDP per capita value after a simulation over 25 years with a diffusion by neighbourhood of 40%

This second type of setting concerns a 25-year period with a budget representing 2% of the Fifteen's GDP, and a diffusion rate by neighbourhood of 40%, distributed among every neighbouring regions. The results of this type of setting show a good beta convergence and a bad sigma convergence. Representative of an inegalitarian growth, this setting reveals a certain effectiveness at national level, but establishes itself as less homogeneous at European level. The high GDP diffusion rate between contiguous regions reflects the importance of

European spaces' growing integration. It produces prosperous regions aggregation phenomena by expansion very often from metropolitan regions : Parisian Basin, South of England, North East of Spain (hence the high values of beta convergence). But on the contrary, neighbouring effects also work between peripheral and poor region areas that do not pull out of their backwardness (hence a bad sigma convergence). Growth develops itself by clusters of regions, but the development gaps are not on the whole being shortened



Map 3 - GDP per capita value after a simulation over 15 years with a diffusion by neighbourhood of 30%

This third map represents the result of settings combining values of beta and sigma convergence close to each other ensuring an average progression of underprivileged regions and a reducing of the overall disparities. It is obtained with a 15-year duration of policies application, relying on a doubling of the European budget, by devoting to it half of the regional policy. While diffusion by integration of neighbouring regions concerns 30% of the GDP. Here, it is about an attempt of a compromise trying to reconcile the beneficial effects of each of the two types of convergence. From this setting, emerges a configuration of the European space which is relatively homogeneous, but dominated by a few very big metropolis (London, Paris, Brussels, Luxemburg,

Stockholm), which development level stands out clearly from the rest of the regions.

From these two indicators, in theory complementary, and often found in literature about regions' convergence, the introduction of a growth propagation variable by contiguity, changed the expected scenarios worked out by the instigators of the Community policies. This introduction of spatial interaction by neighbourhood transforms the Community policies determinist projections in a system of regional units reacting according to a multi-level and a multi-localized complexity. To the multiplicity of settings offered by the simulation platform answers a few seldom co-occurrence probabilities of the two forms of convergence. The taking

into account of the neighbourhood effects reveals the existence of an auto-organized process, which is only seldom holder of a global

7. Conclusion

The use of the GeoCells cellular automaton is an attempt in order to simulate the combined effects of the role of neighbourhood interactions and of the variations in the European regional policies on the reducing of disparities and the European group global cohesion. GeoCell's multi-layer organization is adapted to the specificity of this policies' functioning involving a transfer of resources between State members via the European Union's budget and the neighbourhood effects linked to the integration in the European space. Starting from the double assessment that European Union's spatial dynamic revealed an inegalitarian growth and strongly polarized mode in which a third of European regions produce two thirds of the Community GDP, and that structural aid did not have, in this context, the expected consequences at regional level, we attempted to optimize the combination of duration and of public expense levels in order to make probable the reducing of disparities. The GeoCells model contributed to highlight the incidence of spatial interactions and the complexity induced by diagnosis brought up by two type of indicators of economical convergence. The fact that all of the European regional system's components vary together by interactions between scale levels and between neighbouring regions, make it difficult to achieve the coincidence

convergence at the level of a group of regions as well as of the entire European space.

between improvement of the growth of the whole and the reducing of disparities between regions. It is however at these conditions that the cohesion policy, which makes up one of the pillar of the European building will be likely to answer the challenge of territorial integration in an enlarged Union. On the contrary, there is a high probability of a perennization of an inegalitarian growth mode, associating consolidation of the European integration and regional divergence process partially compensated by redistribution funds.

8. Bibliography

-Batty.M , Goodchild.M ,
Maguire.D, 2005, *GIS, Spatial
Analysis and Modelling*, ESRI
Press, Redlands, California

Beine.M, Docquier.F (eds), 2000,
*Croissance et convergence
économique des régions*, De Boek,
université

Bussi M., Langlois P., Daudé E.,
« Modéliser la diffusion spatiale de
l'extrême droite : une expérimentation
sur le Front national en France ». Colloque AFSP, Paris, 2005.

Charleux . L., La politique
régionale de l'Union européenne: des
régions à l'espace?, Thèse de
doctorat, université J.Fourier ,
Grenoble, 2003.

Elissalde.B , « Géographie, temps
et changement spatial », *L'espace
géographique*, n°3, 2000.

Elissalde.B, « Temporalités et
modélisation des dynamiques
régionales : le cas de l'Union
Européenne » in *Modélisations
géographie, Déterminismes et
complexités*, Guermond Y.(dir),
Hermès-Science, 2005.

Elissalde.B et al, *Géopolitique de
l'Europe*, ed. Nathan, 2006.

-Engelen.G, Neis.B, Whalen.J,
White.R , 2004, *An individual based
model of the Lobster Fishery in St
John Bay Newfoundland Canada*,
(Research Institute for Knowledge
Systems, RIKS)

Eurostat ; Deuxième rapport sur la
Cohésion économique et sociale,
Commission des Communautés
européennes, 2000.

Geyer.R, « European integration,
complexity and the revision of
theory », *Journal of Common Market
Studies*, vol41, 1, 2003.

Guermond Y; Delahaye D.
Dubos-Paillard E., Langlois P.,
« From modelling to experiment »,
GeoJournal, vol. 59, iss. 3, pp. 171-
176(6), Kluwer Academic Publishers,
2004.

Guermond Y., Dubos-Paillard E.,
Langlois P., « Analyse de l'évolution
urbaine par automate cellulaire : le
modèle SpaCelle », *L'espace
géographique*, vol 4, p 357-378,
2003.

Gross C., Opérations topologiques
et géométriques sur les multicartes
combinatoires ; application à la
cartographie thématique, thèse
Université de Strasbourg, 1989.

Heylen.C et al, Transports de
marchandises par la route au niveau
régional, Eurostat, Commission
européenne, 2001.

Kiel,L.D et Elliott. E, *Chaos
theory in the social sciences*,
University of Michigan Press, 1997.

Langlois P., « Formalisation des
concepts topologiques en
géomatique », *revue internationale de
géomatique*, Vol 4, n°2, 1994.

Langlois P., Delahaye D.,
« RuiCells, automate cellulaire pour
la simulation du ruissellement de
surface ». *Revue Internationale de*

Géomatique, pages 461-487, vol 12, n°4, 2002.

-Longley.P , Batty.M, 1996, *Spatial Analysis: Modelling in a GIS Environment*, Wiley, New York

Martin. P, A quoi servent les politiques régionales européennes?, *Economie internationale*, n°81, 2000

Maurel. F, *Scénario pour une nouvelle géographie économique de l'Europe*, Commissariat général du Plan, ed. Economica, 1999.

Riou.,S, « Géographie, croissance et politique de cohésion en Europe », *Revue française d'économie*, n°3, vol XVII, 2002.

Salah-i-Martin,X, « Regional Cohesion: Evidence and Theories of Regional Growth and Convergence », *European Economic Review*, n°40, 1996.

-White.R, Straatman .B and Engelen.G, 2004, *Planning Scenario Visualization and Assessment: A Cellular Automata Based Integrated Spatial Decision Support System*, in: *Spatially Integrated Social Science*, edited by: M. F. Goodchild and D. Janelle, p.420-442, Oxford University Press, New York, USA

Young.A.T, Sigma convergence versus Beta convergence: evidence from U.S. County-Level Data, *Emory Economics*, univ. Atlanta, Sept 2003

1 numéro de la revue du CEPII de juillet 1997 et de mars 1999

Hybrid Approaches for Stock Price Prediction

Qinghua Wen, Zehong Yang, Yixu Song

Abstract—There is a wide acceptance of benefit of the synergy effect. The object of this paper is to investigate whether a hybrid approach combining different stock prediction approaches together can dramatically outperform the single approach and compare the performance of different hybrid approaches. The hybrid model includes three well-researched prediction algorithms: back propagation neural network (BPNN), adaptive network-based fuzzy neural inference system (ANFIS) and support vector machine (SVM). First, we utilize them independently to single-step forecast the stock price, and then integrate the three forecasts into a final result by a combining strategy. Two different combining methods are investigated. The first method is a linear combination of the three forecasts. The second method combines them by a neural network. We have all of the algorithms experiment on the S&P500 Index. The experiment verifies that by combining the single algorithm considerably, a better performance can be received.

Index Terms—BP neural network, ANFIS, Support vector machine

I. INTRODUCTION

IT is assumed that the behaviors of stock market in the future could be predicted with previous information given in the history [1]. This is the very idea behind Technical Analysis of Investment trading. Therefore, there exists a function

$$p(t+1) = f(p_{t-k}, \dots, p_t; x_{t-l}, \dots, x_t; y_{t-m}, \dots, y_t; \dots)$$

where p is the stock price, x and y are the other influence factors such as daily highest prices, daily lowest price, exchange volume, consumer confidence index etc. So the work of predicting the price in the future of stocks transforms to the problem of data regression in the computer science sense.

Many effective artificial intelligence methods can do this work [2]. Among them, as claimed by Grudnitski and Osburn [3], artificial neural networks (ANN) are particularly well suited for finding accurate solutions in an environment characterized by complex, noisy, irrelevant or partial information. So many works focus on applying the different neural networks into stock prediction. Weigend et al [4] and Refenes et al [5] applied multilayer forward network (MLFN) models in their forecasts of foreign exchange prices. Tenti [6] applied recurrent neural network (RNN) models to forecast exchange rates. Kuan and Liu [7] provided a comparative evaluation of MLFN's performance and an RNN for the prediction of an array of commonly traded exchange rates. In a more recent study by Leung et al. [8], Chen and Leung [9] used

an error correction neural network (ECNN) model to predict exchange rates and good forecasting results can be obtained with their model. Although the successful application of neural networks, however, neural networks suffer from a "black box" syndrome and involve difficulties to deal with qualitative information.

On the other hand, fuzzy logic as an effective rule-based modeling system in artificial intelligence not only tolerates imprecise information, but also makes a framework of approximate reasoning. But the fuzzy logic lacks self-learning capability. Therefore, the approach of combining a neural network model with fuzzy logic techniques becomes more popular. For example, Li et al. [1] and Cheng et al. [10] employ ANFIS to predict future stock price. Brent et al [11] compared the Mamdani with Takagi Sugeno Fuzzy inference system learned using neural learning and genetic algorithm, which shows the Takagi Sugeno based Fuzzy inference is much better than the Manani based Fuzzy inference system in the considered stock.

At the same time, recently, a novel type of learning machine, called the support vector machine (SVM), has been receiving increasing attention in areas ranging from its original application in pattern recognition to the extended application of regression estimation. This was brought about by the remarkable characteristics of SVM such as good generalization performance, the absence of local minima, and sparse representation of solution. L. J. Cao and Francis E. H. Tay [12] show that SVM forecasts significantly better than the BP network, but has a similar performance with the regularized RBF neural network in time series forecasting. Y.K. Bao et al [13] conclude the support vector machines for regression is a robust technique for function approximation.

So it is a natural question about what will happen if we combine these approaches together. In this paper, we focus on studying the hybrid of different single stock price prediction algorithm to investigate whether a hybrid approach combining different stock prediction approaches together can dramatically outperform the single approach and compare the performance of different hybrid approaches. The three widely used algorithms are employed, they are BP neural network, adaptive network-based fuzzy neural inference system and support vector machine. First, we utilize them independently to single-step forecast the stock price, and then integrate the three forecasts into a final result by a combining strategy. The experiment results indicate that creating a considerate combining strategy, the performance of the hybrid way is better than the performances of using the three algorithms independently. Two different combining methods are investigated.

The rest of the paper is organized as follows: Section 2 simply describes the three main algorithms used in this paper. Section

3 explains the combining strategies for the hybrid approach. The detailed experiment setup and the corresponding analysis are shown in section 4, and finally some concluding remarks are described in section 5.

II. OVERVIEW OF THE ALGORITHMS

A. BPNN

The BP neural network implies the feed-forward neural networks of which the weight matrix is adjusted by back-propagation (BP) algorithm. BP neural network had emerged as one of the most powerful tool for regression and classification problems. Many papers have investigated the application of BP network in stock price prediction in which most of them employ other optimization algorithms to improve the structure and the parameter of the network to achieve a better forecast. A detail of the application of the BP network in stock can refer to [14], which make a completely discussion of the financial application of BP network. The standard three-layer BP neural network is used in this paper.

B. ANFIS

Adaptive Network-based Fuzzy Inference System (Jang 1993, Jang 1995) is a class of adaptive networks that are functionally equivalent to fuzzy inference system, which combines the advantage of artificial network and fuzzy inference system, and avoids the “black box” of common neural network.

ANFIS is based on Tagaki-Sugeno fuzzy model. The fuzzy rule is:

$$\text{Rule: if } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = p_1x + q_1y + r_1$$

Generally speaking, ANFIS incorporates the following three important features: meaningful and concise representation of structured knowledge, efficient learning capability to identify parameters and clear mapping between parameters and structured knowledge. ANFIS is a class of adaptive networks that are functionally equivalent to fuzzy inference system.

C. SVM

Support vector machine (SVM) is a relatively new approach of data mining. It was developed by Vapnik and his co-workers [15][16] which initially used in pattern classification, with the introduction of Vapnik’s insensitive loss function, SVMs have been extended to solve non-linear regression estimation problems and they have been shown to exhibit excellent performance in financial time series forecasting [17].

Compared to other neural network method, SVM has three distinct characteristics [17]. First, SVM estimates the regression using a set of linear functions that are defined in a high-dimensional feature space. Secondly, SVM carries out the regression estimation by risk minimization, where the risk is measured using Vapnik’s insensitive loss function. Third, SVM implements the structured risk minimizing principle which minimizes the risk consisting of the empirical error and a regularized risk. The SVM employed in this paper is the LibSVM developed by Lin [18]; the details about the LibSVM

refer to [18].

III. COMBINING STRATEGY

The combining strategy is employed to merge the three forecast result of the single forecasts algorithm to form the final prediction of the stock price. Unlike the voting strategy common used in classification problem, in this paper, two special combining strategy is used which called Hybrid-1 and Hybrid-2 in the following experiment respectively. Hybrid-1 is a simply linear combination of the three predictions. Considering the different performance of the three algorithms, the weight assigned for different algorithms is proportional to their accuracies. The weight determined by:

$$W_i = \frac{1}{\sum_{i=1}^3 \frac{1}{MSE_i}}, i = 1, 2, 3$$

Hybrid-2 is a more complex approach which combines the forecasts by a three-layer feed-forward neural network in which there are 3 nodes in the input layer, 3 nodes in the hidden layer, 1 node in the output layer.

IV. EXPERIMENT AND ANALYSIS

A. Data preprocessing

The stock price is quite different in value may affect the performance of the prediction algorithm. So the original data is mapped to [0,1] by *min-max* normalization:

$$x_{scaled} = \frac{x - m}{M - m}$$

for the time series data x , $m = \min(x)$, $M = \max(x)$. In the experiment, three kinds of time series, e.g. the close price, highest price and lowest price are normalized independently. For convenience, the prediction result don’t map back to the true stock price, just compare the scaled value.

B. Feature selection

Feature selection is a key issue for the prediction accuracy. Many papers have dealt with input selection. In the study, the genetic algorithm (GA) is employed to select the most appropriate features for the three algorithms from 15 candidate features respectively. The original 15 inputs considered are:

$$C_t, C_{t-1}, C_{t-2}, C_{t-3}, C_{t-4} \\ L_t, L_{t-1}, L_{t-2}, L_{t-3}, L_{t-4} \\ H_t, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}$$

where C , L , H represent the close price, the lowest price and highest price respectively. The target is to select some most important features as the inputs of the models to forecast the stock price in the $t+1$ transaction day. The number of the selected feature varies with the forecast models.

The genetic algorithm is built with a population size of 40 and trained for 100 generations. The mean squared error (MSE) of regression is used as the measurement.

C. Experiment setup

For the BPNN model, the 8 features selected are $C_t, C_{t-1}, C_{t-2}, L_t, L_{t-4}, H_t, H_{t-1}, H_{t-4}$. The standard three-layer BP neural network is used. There are 8 nodes in the input layer which is equal to the number of features, 1 output node and 10 hidden nodes. The number of hidden node is determined based on the experimental result. The sigmoid transfer function is used for the hidden nodes and the linear transfer function for the output node.

For the ANFIS model, the final selection of inputs is: $C_t, C_{t-1}, C_{t-2}, L_t, H_t, H_{t-1}$. For every input variable, two membership functions are created, and all of the membership functions are global bell functions.

The input of SVM model is same to the BP network model. When applying SVM to financial forecasting, the most important thing that needs to be considered is what kernel function is to be used. As the dynamics of financial time series are strongly nonlinear, it is intuitively believed that using nonlinear kernel functions could achieve better performance than the linear kernel. In this study, the radial basis function (RBF) is used as the kernel function. The two parameters of the RBF are optimized by cross-validation method.

D. Result and analysis

For every model, we test them on the S&P 500 index under different size of the training set vs. test set. The result shows in the Table.1.

The algorithm is implemented in MALAB tool box.

TABLE I
PERFORMANCE OF DIFFERENT MODEL WITH DIFFERENT SIZE OF TRAINING SET VS. TEST SET

Training set : test set	80:20	800:20	800:100	800:200
ANFIS	0.000351	0.000142	0.000121	0.000133
BPNN	0.000295	0.000148	0.000114	0.000130
SVM	0.000486	0.000154	0.000121	0.000144
Hybrid-1	0.000335	0.000147	0.000117	0.000132
Hybrid-2	0.000263	0.000138	0.000113	0.000129

In the first row, the number before the colon is the size of training set, the number after is the size of testing set. The testing set follows the training set in time. E.g. 80:20 means that the price of the first 80 days is the training set, whereas the price in the next 20 days (from the 80th to the 100th) as the test set. Hybrid-1 and Hybrid-2 mean the two combining strategies to combine the three prediction result. The performance is measured by average MSE.

The prediction of Hybrid-1 and Hybrid-2 of 80:20 shows in Fig 1.

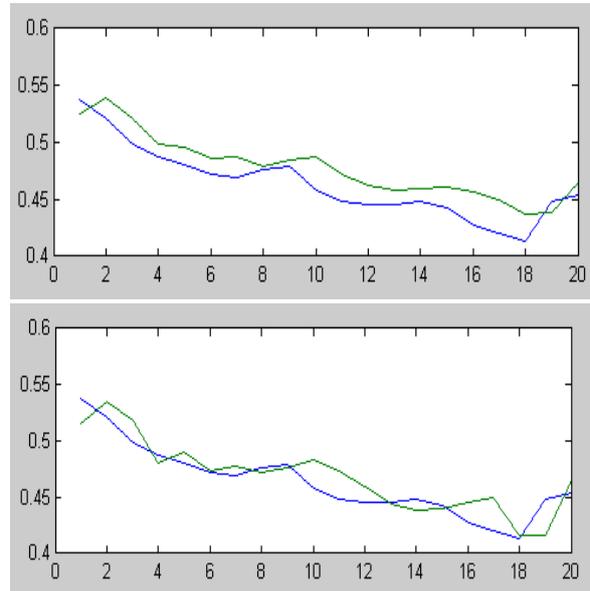
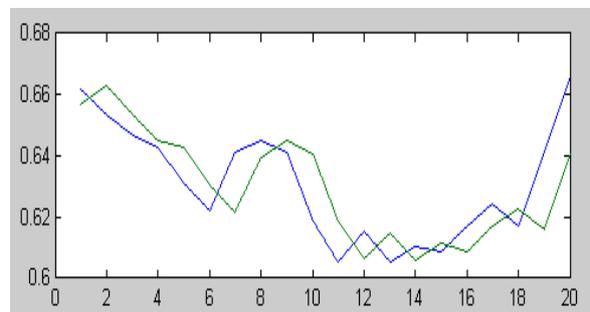


Fig. 1. The prediction of the Hybrid-1 model (top) and the Hybrid-2 model (bottom) for the price of 20 days using the 80 historical daily prices. Blue lines show the true prices, green lines show the forecasts.

As shown in Table 1, the three models share a similar performance. All of them are fit to the short-term prediction, as the test size increases, the performance decrease. The training set also influences the performance. This characteristic implies the stock price reflects the long-term historical prices, not just determined by the short-time historical prices. The performance of hybrid approach is generally better than the single algorithm.

On the other hand, the performance of Hybrid-2 model is much better than the Hybrid-1 in all and outperforms any single algorithm. Hybrid-1 is just slightly better than the average performance of the single algorithms. The reason is that the prediction of the three models is approximately similar. This characteristic shows in Figure 2. The performance of the three models is approximately similar, so the Hybrid-1 is equivalent to using mean prediction of the single algorithm as its final prediction.



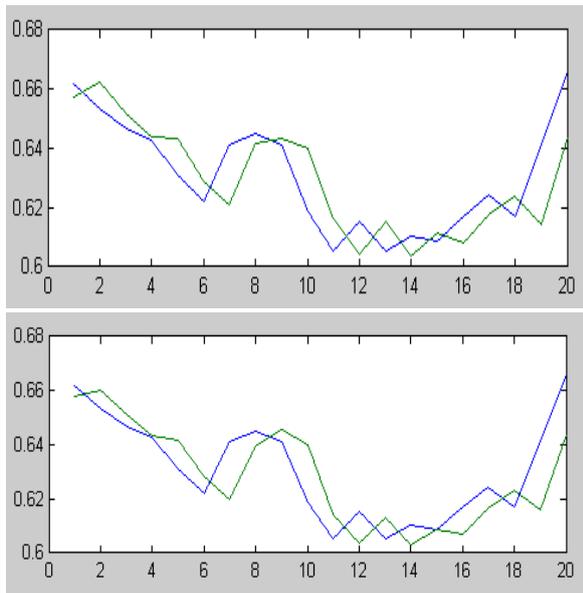


Fig. 2. From top to bottom: the predictions of the SVM model, the BPNN model and the ANFIS model for the price of 20 days using the 800 historical daily prices. Blue lines show the true prices, green lines show the forecasts.

V. CONCLUSION

Many classical soft computing approaches have successfully applied in the prediction of stock price and showed good performance. In this study, the object of this paper is to investigate whether a hybrid approach combining different stock prediction approaches together can dramatically outperform the single approach and compare the performance of different hybrid approaches. The soft computing approaches employed in this paper are the ANFIS model, the BP network model and the SVM model. The experiment is implemented on S&P 500 index. As shown in the experiment, by combining them considerably, a better performance receives

REFERENCES

- [1] R.J. Li, Forecasting stock market with fuzzy neural networks. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005
- [2] Lean Yu, Adaptive smoothing neural networks in foreign exchange rate forecasting. ICCS 2005, LNCS 3516, 2005, pp.523-530.
- [3] Grudnitski, G., Osburn, L., Forecasting S&P and gold futures prices: an application of neural networks. *Journal of Futures Market*, 13, 1993, pp.631-643.
- [4] Weigend, A.S., Huberman, B.A., Rumelhart, D.E., Generalization by weight-elimination with application to forecasting. In: Lippman, R.P., Moody, J.E. and Touretzky, D.S.(Eds), *Advances in Neural Information Processing Systems*, 3, Morgan Kaufman, San Mateo, CA, 1991, pp.875-882.
- [5] Refenes A.N., Azema-Barac M., Chen L., Karoussos, S.A., Currency exchange rate prediction and neural network design strategies. *Neural Computing and Applications*, 1, 1993, pp. 46-58.
- [6] Tenti P., Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, 10, 1996, pp. 567-581.
- [7] Kuan C.M., Liu T., Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, 10 (4), 1995, pp.347-364.
- [8] Leung M.T., Chen A.S., Daouk, H., Forecasting exchange rates using general regression neural networks. *Computers and Operations Research*, 27, 2000, pp.1093-1110.
- [9] Chen, A.S., Leung, M.T., Regression neural network for error correction in foreign exchange forecasting and trading. *Computers and Operations Research*, 31, 2004, pp.1049-1068.
- [10] P. Cheng, Predicting the impact of anticipatory action on US stock market: an event study using ANFIS (A neural fuzzy model).
- [11] Brent Doeksen, Real Stock Trading Using Soft Computing Models, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05)
- [12] Yu-Kun Bao, Forecasting stock composite index by fuzzy support vector machines regression, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005.
- [13] Refenes A.N., Zaprani A.D., Bentz Y, Modeling Stock Returns with Neural Networks. Neural Network Applications and Tools. Workshop on (1993) September 13-14, 1993, pp.39 – 50
- [14] Vapnik VN, Golowich SE, Smola AJ, Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems* 1996, 9:281–7.
- [15] Vapnik VN. *The nature of statistical learning theory*. New York: Springer, 1995.
- [16] L. J. Cao, Francis E. H. Tay, Support vector machine with adaptive parameters in financial time series forecasting. *IEEE transactions on neural networks*, VOL. 14, NO. 6, november, 2003
- [17] <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

A New Method to Improve Network Exchanges of Multimedia Stream

Bechir Alaya, Claude Duvallet, Bruno Sadeg
 LITIS, UFR des Sciences et Techniques
 25 rue Philippe Lebon, BP 540
 F-76058 Le Havre Cedex
 Firstname.Lastname@univ-lehavre.fr

Abstract—Multimedia applications usually manage large quantities of data in the form of frames of certain types. To ensure a good traffic of these frames through the network, temporal constraints must be respected when sending and receiving these frames. If the temporal constraints are not met, then the quality of service (QoS) provided to users decreases. In this paper, we exploit some results obtained in QoS management in Real-Time Databases Systems (RTDBSs), and we apply them to multimedia applications, because of similarities existing between these two fields. We propose a new method allowing to control the QoS provided to clients according to the network congestion, by discarding some frames when needed.

Index Terms—Distributed multimedia systems, Feedback control loop, Quality of service, MPEG format, (m,k)-frame constraints.

I. INTRODUCTION

THE recent improvements in networks area allow to consider the large exploitation of new services in many applications, particularly in multimedia applications. These applications deal with large volumes of data and require real-time processing, i.e., they must be completed before fixed dates, to guarantee an acceptable quality of service (QoS) in the streams presented to the users. Systems adapted to the management of these kinds of data with QoS guarantees are real-time database systems (RTDBSs) [13] [14].

Many distributed multimedia applications must face to the unpredictable loads that cause the system overload. For example, user-demands may arrive in a bursty manner during a short period. Currently, all applications need to provide a good QoS to the users (a good flow of video frames). To this purpose, it will be interesting to adapt the existing techniques to multimedia applications in order to obtain more reliable and efficient transfer of the video packets, without modifying the initial infrastructure. The main problems are related to the adaptation of available resources (bandwidth, buffer size, video servers, etc.) and to the proposition of new techniques which deal with system instability periods (overload or under-utilization). The proposition must allow to ensure an acceptable QoS while respecting the multiple requirements of the video streams.

Many works on QoS management in RTDBSs have been done [1] [10]. Almost all these works are based on a feedback control scheduling architecture (FCSA) that controls the system behavior thanks to a feedback loop.

The feedback loop begins to measure the performances of the system in order to detect overload periods. Then, according to the results observed, the values of the parameters are modified to adjust the system load to the real conditions. As these conditions always vary, this process is repeated indefinitely. Because of the similarities existing between RTDBSs and multimedia applications[6], in this paper, we propose to apply the results obtained on the QoS management in RTDBSs to multimedia applications. The main objective is to allow to design multimedia applications that will be able to provide the QoS guarantees and a certain robustness when user's demands quickly grow up or when the network becomes congested. These works are especially applied to video on demand (VoD) applications. We adapt a method called FCA-DMS (Feedback Control Architecture for Distributed Multimedia Systems). The architecture proposed contains three main components, such as, the master server, the video server and the clients. We will apply a control of the network congestion by discarding or not some multimedia frames of certain types according to the network state, notably to the shared bandwidth. This will increase the QoS provided to the users.

In this article, we present a method allowing to take into account the network congestion in order to increase the QoS provided to the users, especially, how to achieve an optimal value of frames in a MPEG stream [6][12]. In Section 2, we present the multimedia system architecture that we use. In Section 3, we develop our approach which allows to increase the applications QoS during the overload periods (network congestion). Then, in Section 4, we present the simulations results that we have done to test the validity of our approach. Finally, we conclude this article and we give some perspectives.

II. A QUALITY OF SERVICE APPROACH

A. Quality of service in distributed multimedia systems

The usually admitted definition of the QoS in a multimedia application is the whole of requirements in terms of bandwidth, quality of visualization, delay and rate of video packets loss. Our approach consists in taking into account researches already done on the management of QoS in RTDBSs [11] [2] and their adaptation to multimedia systems. To this purpose, we propose an adaptation of a method based on feedback control architecture to distributed multimedia systems [6]. We

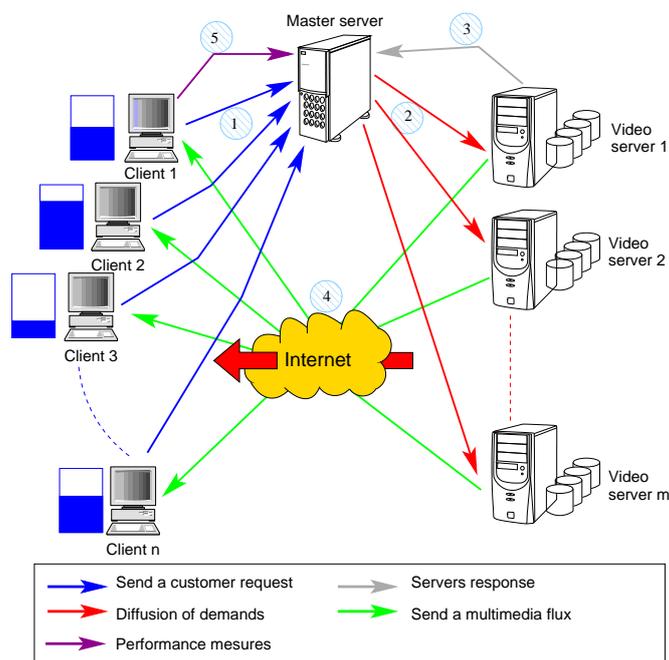


Fig. 1. Functional model of the FCA-DMS architecture.

exploit, to this purpose, the notion of (m,k) -firm constraints used in real-time systems [5] and in RTDBSs [8][3].

B. Feedback control architecture

In a previous work, Natalia Dulgheru has proposed an architecture, named QMPEGv2 [6] which deals with distributed multimedia systems (cf. figure 1). The architecture proposed contains three main components:

- **Master server:** it accepts requests from clients, chooses the video servers able to serve the demand, supervises the system state and adjusts the video streams in order to maintain the QoS initially fixed.
- **Video servers:** They run under the master server control and send the video packets to the clients.
- **Clients:** they send requests to the master server and receive the video frames from the video server. When a state change occurs, they send a feedback report to the master server.

In the following, we describe briefly a typical procedure which is executed when a video on demand is requested, based on FCA-DMS architecture:

- 1) A client sends a request to the master server to get a video, with a certain level of QoS.
- 2) The master server broadcasts the request to the video servers available in the system.
- 3) The video servers send back their response to the master server, which chooses one among them.
- 4) A stream is opened between the chosen video server and the concerned client.

- 5) The master server asks the video servers to adapt their QoS, when necessary.

The feedback loop consists on adapting the QoS according to the load system conditions (servers and network congestion). The system observes the QoS obtained by the client and, if necessary, asks the concerned video server to improve it.

In order to improve the QoS, the system increases or decreases the number of transmitted frames of certain types. To this purpose, we based our action on the characteristics of the standard MPEG format [9], that defines a mechanism to code frames at the time of the video compression. A video sequence enters the system. It's then compressed and coded according to three types of frames: *Intra frames* (I), *Predicted frames* (P) and *Bidirectional frames* (B). I frames are references frames. P frames allow to rebuild a frame using an I frame. B frames use I frames and P frames to rebuild a sequence. Therefore, I frames are the most critical. To decrease the eventual network congestion, it is necessary to remove some frames from a video sequence, but these suppressions must be done in a controlled manner. We propose in the following section a method based on the controlled frames suppression in order to control the QoS provided to users.

C. Feedback control loop

Using the feedback loop allows to stabilize the system during the instability periods [4]. It is based on the two principles: observation and auto-adaptation. The principle consists of observing the results obtained by the system and checking if the current QoS observed is consistent with the QoS initially

required, e.g. in VoD application, the system checks if the videos sequences are presented to users without interruptions. The auto-adaptation consists for the system to adapt the results according to the QoS needed par the clients, by adjusting some network and video parameters, e.g. the system increases or decreases the number of accepted frames¹. In this way, the feedback loop ensures the stability of the system.

III. A NEW METHOD TO CONTROL THE NETWORK CONGESTION

A. The (m,k)-frame method

According to certain conditions, the system load varies from overload state to under-utilization state and vice-versa. Indeed, since the number of video servers sending the video packets is unknown, sometimes this causes severe damages on the service level provided to clients. Consequently, the number of transmitted packets is also unknown and can be important. Moreover, when a high number of video packets access to network resources, it is necessary to keep a high priority level for more critical packets (I frames, then B frames, then P frames)[6][12].

We propose an approach based on (m,k)-firm method [7]. The (m,k)-firm model is characterized by two parameters m and k . An application is said under (m,k)-firm real-time constraint if at least m operations among k consecutive operations meet their deadlines. We adapt this method to the context of multimedia applications. An multimedia operation consists of sending/receiving a video frame. To adapt this method, we consider that m video packets among k must be correctly sent. To this purpose, we propose a new technique of video packets management crossing the network, called (m,k)-frame.

A video stream is decomposed into several classes according to their tolerance to the loss of frames characteristics, i.e. each class contains the video packets of similar (m,k)-frames constraints. The three classes, we consider, refer to the three types of frames: I, B and P. With this technique we realize a trade off between the shared resources and the QoS granularity in the same class of a video stream.

B. The quality of service adaptation

In this work, we focus on the adaptation of the video stream to the network state. We assume that measures of the network capacity are available, in one hand, and that we have an important number of frames to send, on the other hand.

The three classes of frames (I, B and P) are used to adapt the quality of stream sent to the network capacity. We consider the following constraints: (m_I, k_I) -frame, (m_P, k_P) -frame and (m_B, k_B) -frame, i.e. m_i frames of a certain type must be received among the k_i frames sent. Then the network capacity is measured by the formula: $m_I + m_P + m_B$. Recall that I frames are the most critical. The parameters are ordered in the following manner: $m_I > m_P > m_B$. We usually have $m_I = k_I$, i.e., I frames are critical and it is forbidden to remove them.

¹note that I frames (critical) are not removed

We assume the situation where the network, whose current capacity is N , is congested. We also assume that QoS_{max} is the quality of the stream to send including M frames. To be consistent with the network capacity, it is necessary to remove $(M - N)$ frames. Therefore, we have to degrade the quality of the MPEG stream. When we apply no method of congestion control, frames will be randomly removed, i.e. they are lost by the network, causing the degradation of the video presentation, notably if some I frames are removed. Here, we apply our (m,k)-frame method, which consists of removing frames in an intelligent manner. We have: (1) $M = k_I + k_P + k_B$, and (2) $N = m_I + m_P + m_B$. The number of frames to remove is then: $M - N = (k_I - m_I) + (k_P - m_P) + (k_B - m_B)$, where $k_I = m_I$ (I frame are the most critical, and are not to remove).

C. Bandwidth fair sharing

With the previous assumptions, we deal with the problem of sharing bandwidth between servers in case of network congestion phases. In the previous section, we have seen how to reduce the QoS at the stream level, according to the available capacity of the network. Here, we need to share fairly the bandwidth between all sources that wish to send a stream. We compute the total capacity needed by all servers. Then, we compute R , the ratio between the needed capacity and the available network capacity (N).

$$R = \frac{N}{\sum_{i=1}^m RC_i}$$

such as :

- m : the number of video servers.
- RC_i : The Required Capacity of the video server i .

Example: let 3 video servers wishing to send flows of 40, 30 and 20 frames per second respectively. The total capacity of the network needed to answer to this demand must be $40+30+20=90$. If, however, the network only arranges a capacity of 75 frames per seconds, it is not able to sent all the frames. The ratio R is computed as follows: $(75/90)*100 = 83.33\%$. Then, we apply this rate to each of the three required capacities $40*83.33\%=33$, $30*83.33\%=25$ and $20*83.33\%=17$. If we sum the three obtained numbers, we find 75 frames per second. It corresponds to the current capacity of the network.

In the following, are listed some advantages of the bandwidth fair sharing:

- To control of the congestion of the network by fair sharing resources between all streams. A bad stream doesn't affect the service provided to the other streams. Only this service will be concerned if a stream wants to consume more resources than available.
- To guarantee an acceptable bandwidth and delay.
- To guarantee a link sharing between the different classes of service.

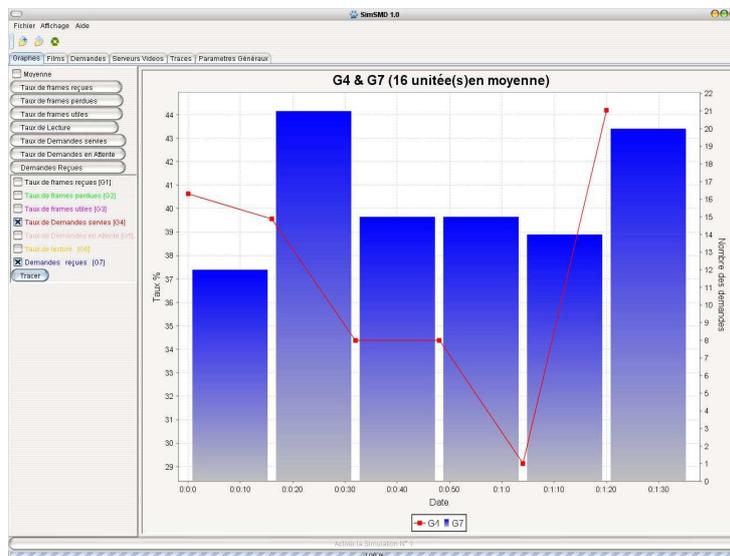


Fig. 3. The simulator

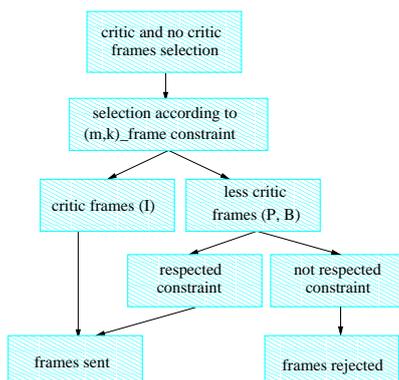


Fig. 2. The (m,k)-frame algorithm.

IV. SIMULATIONS AND RESULTS

A. The simulator

To test the validity of our proposition, we have developed and implemented a simulator (cf. Figure III-C) in order to verify the behavior of system and its adaptation to different load variations.

This simulator is based on the architecture components presented in Section II. A master server serves all video servers participating to the video diffusion. Moreover, the master server allows to add new video servers, when necessary. A new server is assigned a number and a list of accessible objects. After starting the master server, the video servers who want to participate in the video distribution, refer themselves to the master server, and then get a number.

In order to make a request, a client dialogs with the master server, that distributes the request to the available video

servers. When the master references a video server able to satisfy the client request, i.e. the server is able to provide the required QoS, it sends the video server reference to the client. Then, the broadcast from the video server can begin. After some time, the client sends to the server the QoS level he obtained.

The three parts of the architecture of this simulator are modeled thanks to an object modeling language and realized in JAVA language.

The objective of the simulations is to demonstrate how our method, called (m,k)-frame, is able to adapt the QoS to the real conditions of a multimedia application, according to the current system load. Notably, the system must adjust the QoS when the client number that does requests varies (dynamic arrived of clients). The network congestion can have different sources:

- *internal*: when there is a large number of clients doing requests in the system. We can limit this client number by using an admission controller located at the master server level.
- *external*: when the network is used by other applications that can cause the congestion.

Our architecture must adjust the QoS by reducing the number of frames broadcasts in the network.

B. Simulation results

In these Figures, *sim0* indicates the rates before (m,k)-frame application and *sim1* indicates the rates after (m,k)-frame application. We noticed in these Figures, a difference in the rates of lost frames, the rate of received frames and the rate of useful frames before and after using the (m,k)-frame application. In Figure 4, we observed more gain to rate of received frames after the application of the (m,k)-frame technique in case of system overload.

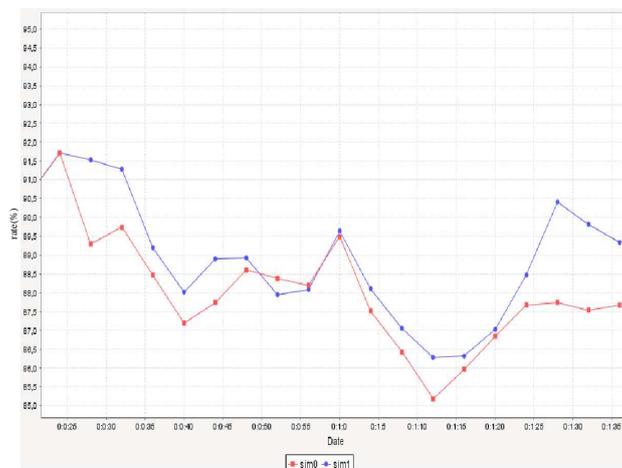


Fig. 4. Received frames rates with and without application of (m,k)-frame technique.

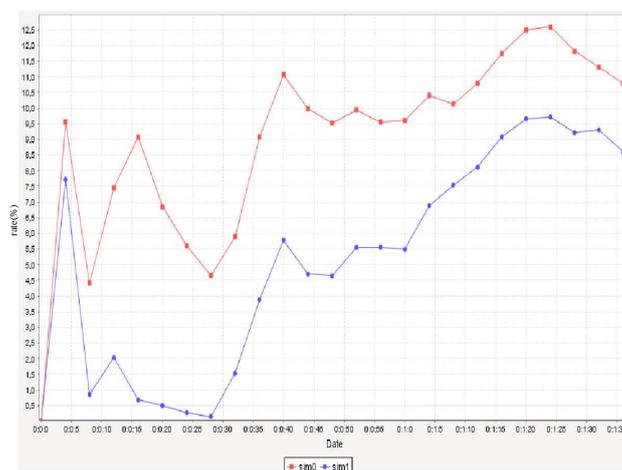


Fig. 5. Lost frames rates with and without application of (m,k)-frame technique.

In Figure 5, we found a difference in loss rate after the implementation of the (m,k)-frame classification. With the increasing of the number of clients over time, we notice in the Figure 5 that reduces of loss frames, we also note an increase in the rate of received and useful frames in Figure 4 and Figure 6 respectively. We have not found a significant gain in the rate of received frames and the rate of useful frames, because our technique of (m,k)-frame classification is not sufficient to effectively stabilize the QoS desired by the clients.

Our simulation shows nearly 10 % gain in the rate of loss frames. But by combining our (m,k)-frame application with other techniques, we can have a significant gain in terms of received frames.

V. CONCLUSION AND FUTURES WORKS

In this work, we proposed an improvement of the feedback control architecture for distributed multimedia systems (FCA-

DMS). Our objective is to provide a deterministic temporal guarantee according to the temporal constraints to real-time video streams. Our main contribution concerned the adaptation of (m,k)-firm constraints for the video packets and the establishment of a video distribution strategy.

A possible extension of this work consists on enhancing the architecture presented. Notably, in order to bring some breakdowns tolerance because of the presence of only one master server. We also have presented the importance of the bandwidth sharing and given a certain priority to the video packets in the network resources level, in order to increase its reliability and robustness and to converge towards the QoS specified by the client.

The simulator design allowed to validate the feasibility of our approach and should permit to provide results demonstrating the real contribution of this new approach.

A possible future work would consist also of building a

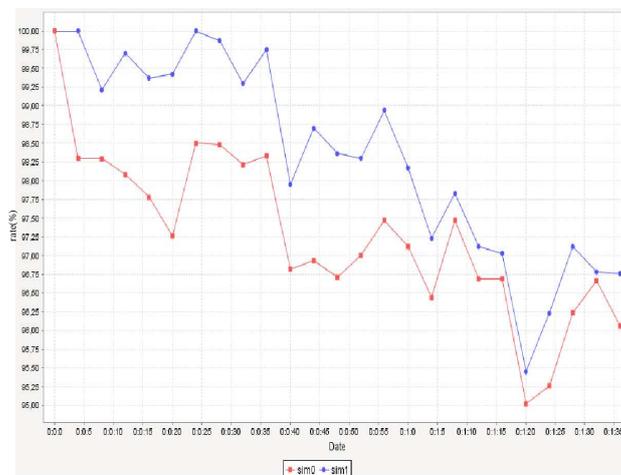


Fig. 6. Useful frames rates with and without application of (m,k)-frame technique.

real video on demand server based on the architecture that we proposed.

REFERENCES

- [1] M. Amirijoo, J. Hansson, and S. H. Son. Specification and management of QoS in real-time databases supporting imprecise computations. *IEEE Transactions on Computers*, 55(3):304–319, 2006.
- [2] M. Amirijoo, J. Hansson, and S. Song. Error-driven QoS imprecise management in imprecise real-time databases. In *Proc. of Euromicro Conf. on Real-Time Systems (ECRTS'03)*, pages 63–72, Portugal, 2003.
- [3] S. Andler and J. Hansson, editors. *Active, real-Time and Temporal Database Systems (ARTDB'97)*, Como, Italy, 1997. Proceedings of the Second International Workshop on Active, Real-Time and Temporal Database Systems, Springer.
- [4] E. Bouazizi, C. Duvallet, and B. Sadeg. Management of QoS and data freshness in RTDBSs using feedback control scheduling and data versions. In *Proceedings of 8th IEEE International Symposium on Object-oriented Real-time distributed Computing (ISORC'2005)*, pages 337–340, Seattle, United State, May 18-20 2005.
- [5] G. Bucci, M. Campanai, and P. Nesi. Tools for specifying real-time systems. *Journal of Real-Time Systems*, 8:117–172, 1995.
- [6] N. Dulgheru. Management of QoS in Distributed MPEG Video. Master's thesis, University of Linköping, 2004.
- [7] M. Hamdaoui and P. Ramanathan. A Dynamic Priority Assignment Technique for Streams with (m, k) -Firm Deadlines. *IEEE Transactions on Computers*, 44(4):1325–1337, 1995.
- [8] J. Hansson and S. Son. Overload Management in RTDBs. In *Real-Time Database Systems: Architecture and Techniques*, chapter 10, pages 125–140. Kluwer Academic Publishers, 2001.
- [9] ISO/IEC 13818-2. Information Technology-Generic Coding of Moving Pictures and Associated Audio, Part 2: Video. Recommendation ITU H.262, International Standardization Organization, November 1994.
- [10] K. Kang, S. Son, J. Stankovic, and T. Abdelzaher. A QoS-Sensitive Approach For Timeliness and Freshness Guarantees in Real-Time Databases. In *Proceedings of the Euromicro Conference on Real-Time System*, pages 203–212, 2002.
- [11] K.-D. Kang, S. Son, and J. Stankovic. Service Differentiation in Real-Time Main Memory Databases. In *5th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (IEEE ISORC'02)*, pages 119–128, Washington D.C, 2002.
- [12] J. Ng, K. Leung, and W. Wong. Quality of Service for MPEG Video in Human Perspective. Technical report, Hong Kong Baptist University., 2000.
- [13] K. Ramamritham. Real-time databases. *Journal of Distributed and Parallel Databases*, 1(2):199–226, 1993.
- [14] K. Ramamritham, S. Son, and L. DiPippo. Real-Time Databases and Data Services. *Real-Time Systems*, 28:179–215, 2004.

Index

- Andrew ADAMATZKY, 91
Sergey AGAFONOV, 132
Bechir ALAYA, 206
Ramon ALONSO-SANZ, 91
Mohammad Riyad AMEERUDDEN, 154
Pierre AUGER, 5
- Akdemir BAYRAM, 160
Igor BELYKH, 110
Maamar BETTAYEB, 138, 144
Paul BOURGINE, 7
Mohamed BOUTAYEB, 70, 74
Samuel BOWONG, 79
- Liliana CARDOZA-AVENDAÑO, 56
Enrique CASTRO, 118
Di CHANGCHUN, 124
Hugues CHATÉ, 12
Estelle CHERRIER, 65
Leon O. CHUA, 2
Nathalie CORSON, 100
Michel COTSAFTIS, 8
Cesar CRUZ-HERNANDEZ, 56
- Alexey DANEEV, 132
Jacques DEMONGEOT, 4
Jian DING, 29
J. DONGES, 11
Marco DORIGO, 10
Mehmet DURSUN, 112
Antoine DUTOT, 171
Claude DUVALLET, 206
Bernard ELISSALDE, 175
- Yves EMVUDU, 79
Nurettin ÇETINKAYA, 150
- Guizhen FENG, 29
Jean-Pierre FRANÇOISE, 3
- Antolino GALLEGO, 118
A GENÇ, 160
Jean-Marc GINOUX, 51, 61
Serter GÜMÜŞ, 125
Salih GÜNEŞ, 112, 125, 160
Juan José GONZÁLEZ DE LA ROSA, 118
Dominique GOYAT, 175
- Sébastien HÉNAFF, 47
- Everardo INZUNZA-GONZÁLEZ, 56
- Sajiya JALIL, 110
Zheng JIAN, 124
Zhi JIANZHUANG, 124
- Sadik KARA, 125
Jürgen KURTHS, 11
- Patrice LANGLOIS, 175
Laurent LAVAL, 41
Feng LIU, 15
Wei LIU, 166
Zongtian LIU, 166
René LOZI, 20, 47
Rosa Martha LÓPEZ-GUTIÉRREZ, 56
- Mohammed M'SAAD, 65
Genaro J. MARTÍNEZ, 91

N. MARWAN, 11
Qiong MENG, 25
Antonio MORENO-MUNOZ, 118
Dany Pascal MOUALEU, 79
Djamila MOULAY, 105

Ayse ÜNAL, 125

Damien OLIVIER, 171

Rosa PIOTRKOWSKI, 118
Kemal POLAT, 112

Wen QINGHUA, 202
Céline QUÉMARD, 40

Bruno ROSSETTO, 51, 61
Otto ROSSLER, 6
Harry C. S. RUGHOOPUTH, 154
Vyacheslav RUSANOV, 132

Bruno SADEG, 206
Mehmet Emin SAKARYA, 125
Reem SALIM, 138, 144
Ali SALLI, 125
Guilhelm SAVIN, 171
Dmitry SHARPINSKIY, 132
Guodong SHI, 15
Andrey SHILNIKOV, 110
Jean-Jacques SLOTINE, 70

Ina TARALOVA, 47
Roger TAULEIGNE, 34
Jean Jules TEWA, 79

Timofey UDILOV, 132

Zhiqing WENG, 15
Weili WU, 29

Mingjun XIN , 166
Dong XU, 166

Song YIXU, 202
James A. YORKE, 9
Şebnem YUSUNKAYA, 112

Yang ZEHONG, 202
Ali ZEMOUCHE, 70, 74
Bofeng ZHANG, 166
Yan ZHAO, 166
Wen ZHOU, 166
Ali Osman ÖZKAN, 125
Y. ZOU, 11

ICCSA 2009

June 29 - July 02

3rd International Conference on
**Complex Systems
and Applications**

**Conference Proceedings
Special Sessions**

Cyrille Bertelle, Xinzhi Liu, M.A.Aziz-Alaoui (eds)



Université du Havre

ICCSA 2009

**Third International Conference on Complex
Systems and Applications**

Special Sessions

June 29 – July 02 2009

Le Havre, Normandy, France

Cyrille Bertelle, Xinzhi Liu and M.A. Aziz-Alaoui (eds)

Advisory Committee:

- Auger P., Académie des Sciences, France
- Bourguine P., ISC, Paris, France
- Chen G., City University of Hong-Kong
- Chua O.L., University of California at Berkeley, USA
- Demongeot J., IUF, TIMC Grenoble, France
- Françoise J.-P., UPMC, Paris 6, France
- Grebogi C., University of Aberdeen, UK
- Pumain D., ISC, Paris, France
- Rossler O., IPTC, Tübingen, Germany
- Siljak D.D., Santa Clara University, USA

General Chairs:

- Cyrille Bertelle, University of Le Havre, France
- Xinzhi Liu, University of Waterloo, Canada
- M.A. Aziz-Alaoui, University of Le Havre, France

Local Organization Committee:

- M. Aziz-Alaoui, LMAH, Le Havre, France
- Julien Baudry, LITIS, Le Havre, France
- Stefan Balev, LITIS, Le Havre, France
- Cyrille Bertelle, LITIS, Le Havre, France
- Nathalie Corson, LMAH, Le Havre, France
- Antoine Dutot, LITIS, Le Havre, France
- Rawan Ghnemat, LITIS, Le Havre, France
- Frédéric Guinand, LITIS, Le Havre, France
- Rabah Labbas, LMAH, Le Havre, France
- Djamila Moulay, LMAH, Le Havre, France
- Damien Olivier, LITIS, Le Havre, France
- Yoann Pigné, LITIS, Le Havre, France
- Claire Roussin, LITIS-LMAH, Le Havre, France
- Nathalie Verdière, LMAH, Le Havre, France
- Adnan Yassine, LMAH, Le Havre, France

Regional Organization Committee:

- Abdulrab H., INSA Rouen, France
- Bouzid M., University of Caen, France
- Caputo J.-G., INSA Rouen, France
- Cherrier E., ENSI of Caen, France
- Daude E., University of Rouen, France
- Elissalde B., University of Rouen, France
- Letellier C., University of Rouen, France
- Mutabazi I., University of Le Havre, France

International Scientific Committee:

- Afraimovich V., IICU, University of San Luis Potisi, Mexico
- Ahmed N.U., University of Ottawa, Canada
- Akhmet M.U., Middle East Technical University, Turkey
- Ayesha A., De Montfort University, UK
- Banos A., LIV, Strasbourg, France
- Barbot J.P., ENSEA, Cergy, France
- Basin M., Autonomous University of Nuevo Leon, Mexico

- Belykh I., Department of Mathematics and statistics, Georgia State University, USA
- Biswas S.K., Temple University, USA
- Bollt E.M., Clarkson University, USA
- Boukas E. K., Polytechnique de Montreal, Canada
- Boutayeb M., Strasbourg University, France
- Chavalarias D., CREA, Ecole polytechnique, France
- Collet P., Centre de Physique théorique, Ecole polytechnique, Paris
- Cotsaftis M., LTME/ECE, Paris, France
- Courbage M., Université Paris Diderot Paris 7, France
- Danca M.F., Avram Iancu University, Romania
- Datta B.N., Northern Illinois University, USA
- Diehl M., K.U.Leuven, Belgium
- Duan, Z., Peking University, China
- Duchamp G.H.E., Paris XIII University, France
- Elwakil A.S., University of Sharjah, Emirates
- Gerdtts M., University of Birmingham at Edgbaston, UK
- Ginoux J.-M., PROTEE, Toulon, France
- Gulyas L., Lorand Eotvos University, Hungary
- Han Q., Central Queensland University, Australia
- Ho D., City University of Hong Kong
- Hong, L., Wright State University, USA
- Ishitobi, M., Kumamoto University, Japan
- Jost C., University of Paul Sabatier, Toulouse, France
- Kolumban G., Budapest University of Technology and Economics, Hungary
- Kurths J., Postdam University, Germany
- Lam L. University of Hong Kong
- Letellier C., University of Rouen, France
- Li B.-L., California University, USA
- Li T., Tsinghua University, China
- Lozi R., Nice University, France
- Lu Jinhu, Chinese Academy of Sciences, China
- Mao X., University of Strathclyde, UK
- Migorski S., Jagiellonian University, Poland
- Momani S., Mutah University, Jordan
- Mordukhovich, B.S., Wayne State University, USA
- Müller J.-P., CIRAD, Montpellier, France
- Odibat Z., Al Balqa Applied University, Jordan
- Paolini E., Universidad Nacional del Sur, Argentina
- Respondek W., INSA Rouen, France
- Rossetto B., PROTEE, Toulon, France
- Sanjuan M., Department of physics, Universidad Rey Juan Carlos, Madrid, Espana
- Shen, X., University of Waterloo, Canada
- Smith R., North Carolina State University, USA
- Tang W.K.S., City University of Hong Kong, China
- Teo K.L., Curtin University of Technology, Australia
- van Wyk M., University of the Witwatersrand, South Africa
- Volny D., University of Rouen, France
- Wang P.K.C., University of California at Los Angeles, USA
- Ying H., Wayne State University, USA
- Yong J., University of Central Florida, USA
- Yu P., University of Western Ontario, Canada
- Yue W., Konan University, Kobe, Japan

Contents

Contents	iii
1 From Nonlinear Dynamics to Biology & Biomedicine	1
Representations of dynamical systems	
Daniel J. CROSS , Robert GILMORE	3
Stochastic modeling of the mammalian circadian oscillator	
Didier GONZE	8
Close return plots for a analysing patient-ventilator interactions during nocturn ventilation	
Christophe LETELIER , D. BOUNOIARE , R. NAECK , Ubiratan. S. FREITAS , A. PORTMANN , A. CUVELIER , J.-F. MUIR	13
Controlling the interaction between wild and transgenic mosquitoes	
Marat RAFIKOV , Ana Paula P. WYSE , Luiz BEVILACQUA	20
TBA	
Marc LEFRANC	24
Analysis of the sequential organisation of mice trajectories	
Philippe FAURE , Annick LESNE	24
2 Complexity in Soft and Condensed matter	25
Statistical analysis of nanoscale real space images for surface, interface and thin film studies and applications	
T. NGUYEN , O. BEZENCENET , A. BATAILLE , Daniel BONAMY , A. BARBIER , J. COUSTY , Luc BARBIER	29
Crackling noise in cracks	
Daniel BONAMY , S. SANTUCCI , L. PONSON	30
Aging and effective temperatures near a critical point	
Sylvain JOUBAUD , Artyom PETROSSYAN , Sergio CILIBERTO	31
Effective rheology and transport properties of an active suspension of E-Coli	
Eric CLEMENT , J. DAUCHET , G. MINO , T. DARNIGE , M. HOYOS , A. ROUSSELET	32
Dynamics and thermodynamics of the thermohaline circulation in a two-dimensional ocean model: theory	
Gilles COLLETTE , B. DUBRULLE , D. PAILLARD	33

Crack propagation and pinning in heterogeneous materials: effects of microscopic disorder and toughness pattern	
Davy DALMAS , E. BARTHEL , D. VANDEMBROUCQ	34
From cage jumps to dynamical heterogeneities in granular media	
Olivier DAUCHOT , R. CANDELIER	35
New approach for treating of spatio-temporal diagram: conditional averaging of two-dimensional fields	
Alexander EZERSKI , N. ABCHA , Innocent MUTABAZI	36
How herds move: a study of polar self-propelled particles	
É. BERTIN , M. DROZ , Guillaume GRÉGOIRE	37
Modulated spiral in a Couette-Taylor system submitted to a high radial temperature gradient	
Raphaël GUILLERM , Arnaud PRIGENT , Innocent MUTABAZI	38
Experimental investigation of the dynamical correlations between magnetic nanoparticles in isotropic and textured superspin-glasses	
K. KATSUYOSHI , Denis L'HÔTE , S. NAKAMAE , Y. TAHRI , Caroline THIBIERGE , E. VINCENT , V. MOSSER , A. KERLAIN , M. KONCZYKOWSKI , E. DUBOIS , V. DUPUIS , R. PERZYNSKI	39
Finite-size effects on sub-critical bifurcations in high-dimensional lattices of coupled logistic maps	
Paul MANNEVILLE	40
Oscillations in bidisperse fluidized suspensions	
A. DEBOEUF , G. GAUTHIER , Jérôme MARTIN , Dominique SALIN	41
Intermittent regime in flow oscillations investigated by means of symbolic dynamics	
Luc PASTUR , François LUSSEYRAN , Thierry FAURE , Christophe LETELLIER	42
Intermittency in granular flow	
Marc RABAUD , P. GONDERT , R. FISCHER	43
Evolution of the fabric tensor in amorphous silica: via molecular dynamics simulations	
Cindy ROUNTREE , M. TALAMALI , D. VANDEMBROUCQ , S. ROUX , E. BOUCHAUD	44
Lock exchange with autocatalytic reaction front	
I. BOU MALHAM , N. JARRIGE , Jérôme MARTIN , N. RAKOTOMALALA , L. TALON , Dominique SALIN	45
Dynamic fracture: how brittle are brittle amorphous solids?	
Julien SCHEIBERT , C. GUERRA , F. CÉLARIÉ , Davy DALMAS , Daniel BONAMY	46
Dynamical length scale in glass transition: experimental study	
Caroline THIBIERGE , Denis L'HÔTE , F. LADIEU , C. BRUN	47
Trajectories of sand grains in a wave boundary layer	
Tien DAT CHU , A. JARNO-DRUAUX , Alexander EZERSKI , F. MARIN	48
Elastic instability of polymer solutions in the Couette-Taylor system	
Fayçal KELAI , Olivier CRUMEYROLLE , Innocent MUTABAZI	49
The renormalized fragmentation equation and its exact solutions	
M.A. GOROKHOVSKI , V.L. SAVELIEV	50
3 Combinatorics, Physics and Complexity	51
Asymptotic analysis and convergence of some leader election algorithms	
Christian LAVAULT	53
Learning self organizing maps as a Markov mixture	
Mustapha LEBBAH , Younès BENNANI , Nicoleta ROGOVSKI	53
Statistics on graphs, exponential formula and combinatorial physics	
Gérard H.E. DUCHAMP , Laurent POINSOT , Silvia GOODENOUGH , Karol A. PENSON	60

Exponential random graphs as models of overlay networks	
Moez DRAIEF	64
Coding rhombus tilings by multidimensional words: a first attempt	
Damien JAMET	64
Synchronization of countable cellular systems, localization of quasi-periodic solutions of autonomous differential systems	
Laurent GAUBERT	64
Drawing solution curve of a differential equation	
Farida BENMAKROUHA , Christiane HESPEL , Edouard MONNIER	75
Generating series: a combinatorial computation	
Farida BENMAKROUHA , Christiane HESPEL	79
On approximation of nonlinear generating series by rational series	
Mikhail V. FOURSOV , Christiane HESPEL	84
Counting rooted and unrooted triangular maps	
Samuel VIDAL , Michel PETITOT	91
About a group of Drinfel'd associators	
Hoang Ngoc MINH	97
Combinatorics & Schelling's model	
Cyril BANDERIER , Hanane TAFAT	97
Systemic modelling	
Daniel KROB	97
4 Risk and Complexity	99
Exploring crowd dynamics based on individual behaviors and local interactions	
Haïfa ABDELHAK , Éric DAUDÉ , Damien OLIVIER	102
Cholera in the 19th century: constructing epidemiological risk with complexity methodologies	
Éric DAUDÉ , Emmanuel ELIOT , Emmanuel BONNET	110
A multiagent urban traffic simulation Part II: dealing with the extraordinary	
Éric DAUDÉ , Pierrick TRANOUEZ , Patrice LANGLOIS	116
Use of geosimulations and the complex system theory to better assess flash floods risks in the Paris Basin watersheds (France)	
Johnny DOUVINET , Daniel DELAHAYE , Patrice LANGLOIS	121
Compiled risks of spatial complexity: the map algebra contribution	
Emmanuel BONNET , Thierry SAINT-GÉRARD , Eliane PROPECK , David GAILLARD	125
Modeling and simulation of pedestrian behaviors in transport areas: the specific case of platform/train exchanges	
Jérémy FIEGEL , Arnaud BANOS , Cyrille BERTELLE	129
Agent-oriented approach for detecting and managing risks in emergency situations	
Fahem KEBAIR , Frédéric SERIN	134
Validation of an ontology of risk and disaster through a case study of the 1923 Great Kanto Earthquake	
Damienne PROVITOLLO , Jean Pierre MÜLLER , Edwige DUBOS-PAILLARD	140
Exploitation of a displacement survey to detect road network use vulnerability	
Michel NABAA , Cyrille BERTELLE , Antoine DUTOT , Damien OLIVIER , Patrick LIONS	151
5 Evolution Problems and Complexity	157

Oscillatory behavior of solutions for a class of second order nonlinear differential equation with perturbation	
Quanxin ZHANG , Lei WANG	159
On quasistatic models of contact phenomena	
Anna OCHAL , Stanislaw MIGORSKI	164
On an isotropic differential inclusion	
Ana Cristina BARROSO , Gisella CROCE , Ana Margarida RIBEIRO	169
On some transmission problems in Hölder spaces	
Ahmed MEDEGHRI	174
Coefficients de singularités géométriques pour des problèmes d'évolution	
Mohand MOUSSAOUI	174
Sturm-Liouville problems for a complete abstract second order differential equation of elliptic type in UMD spaces	
Stéphane MAINGOT	174
A general decay result in a viscoelastic Timoshenko system	
Salim MESSAOUDI	174
On some transmission problems with boundary Dirichlet conditions	
Fatimetou MINT AGHRABATT	175
Complete abstract differential equations of elliptic type on the half-line : application of Dore-Venni and Da Prato-Grisvard sum theory in Lp-spaces	
Amine ELTAIEF	175
On some abstract fourth differential equation with transmission and boundary conditions	
Hassan DIARAMOUNA SIDIBE	175
6 Chaotic Dynamics, Control and Applications of Complex Systems	177
Finding a realistic guaranteed-chaotic physical model is not easy	
James A. YORKE	181
New developments on partial control of chaotic systems	
Samuel ZAMBRANO , Miguel A. F. SANJUÁN	182
Riddled basins in complex physical and biological systems	
Ricardo Luiz VIANA	183
Synchronization of time varying networks	
R.E. AMRITKAR	184
A simple way of calculating the topological entropy for interval maps	
José AMIGÓ , Rui DILÃO	185
Chimera states in non-locally coupled phase oscillators with propagation delays	
Abhijit SEN	186
Phase control and synchronization in excitable systems	
Jesús M. SEOANE	187
Avoiding escapes in open dynamical systems using phase control	
Miguel A. F. SANJUÁN	188
Targeting synchronized response in chaotic oscillators	
Syamal Kumar DANA	189
Synchronization in chains of oscillators	
Ioan GROSU , M. HASLER , A. BIRZU	190
Simple discrete 3-dimensional stirring models	
Judy KENNEDY	191

How to compute safe sets with escape time sets	
Juan SABUCO , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	192
Point-vortex interaction in an oscillatory deformation field: Hamiltonian dynamics, harmonic resonance and transition to chaos	
Xavier PERROT , Xavier CARTON	193
Applying ordinal patterns to spatially extended systems	
José AMIGÓ , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	194
Control and synchronization of linearly coupled inertia ratchets with co-existing attractors	
U. E. VINCENT , I. O. OLUSOLA , D. MAYER	195
Control of noisy transient chaos in a discrete time continuous analog circuit	
Alexandre WAGEMAKERS , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	196
A Hindmarsh-Rose type silicon neuron	
Takashi KOHNO , Kazuyuki AIHARA	197
Estimation of the control parameter of a map through the analysis of its order patterns	
David ARROYO , Gonzalo ALVAREZ , José AMIGÓ	198
A pulsed-therapy of heterogeneous tumor model	
Abdelkader LAKMECHE	199
Partial control of the Duffing oscillator with fractal basins	
Mattia COCCOLO , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	200
Anti-phase, in-phase synchronization, bifurcation and robustness of nonlinear interacting oscillators	
Rui DILÃO	201
A new type of orbits in the five body problem	
Arsen DZHANOEV , Alexander LOSKUTOV , Miguel A. F. SANJUÁN	202
Dynamical consequences of mesoscopic organization in complex systems: dynamical stability & synchronization in modular and hierarchical networks	
Sitabhra SINHA	203
State space of a rotating rotor	
V.M. SOKOL	204
7 Chaos and Synchronization of Fractional-Order Systems	205
A note on chaos control and synchronization of fractional order systems	
Zaid M. ODIBAT	208
A special kind of synchronization of different chaotic discrete-time systems	
Yong CHEN , Yiliang JIN , Xin LI	213
Non-standard discretization of fractional differential equations	
G. Hussian ERJAEI , Modi ALNASR , Shaher MOMANI	218
Complexified dynamical systems from real fractional actionlike with time-dependent fractional exponent on multi-fractals sets	
Ahmad Rami EL-NABULSI	223
Designing modified projective synchronization for fractional order chaotic systems	
Luo RUNZI , Deng SHUCHENG , Wei ZHENGMIN	228
On some stability conditions and hyperchaos synchronization in the new fractional order hyperchaotic Chen system	
Ahmed E. MATOUK	232
Backstepping control of fractional-order chaotic systems	
E. NASERI , A. RANJBAR , S.H. HOSSEINIA , Shaher MOMANI	237

Control of Genesio-Tesi and Chen chaotic systems using a fractional-order controller	
M. MAHMOUDIAN , A. RANJBAR , E. NASERI , S.H. HOSSEINIA , Shaher MOMANI	242
Synchronization of chaotic fractional-order Coulet system via ASMC	
M. SHAHIRI T. , A. RANJBAR , R. GHADERI , S.H. HOSSEINIA , Shaher MOMANI	247
Dynamical behaviors, linear feedback control and synchronization of the fractional order Liu system	
Ahmed E. MATOUK	253
Chaotic synchronization of fractional-order Chua's system with time-varying delays	
Shangbo ZHOU , Xiaoran LIN , Hua LI	258
Solution of the fractional diffusion equation with absorbent term and external force	
Subir DAS	264
Solution of vibration equation by homotopy analysis method	
Ms. S. CHAKRABORTY	268
8 Transportation and Logistics	271
Conception and methodologies to delimitate hinterlands and an application in a container terminal	
Nélio D. PIZZOLATO , Luiz F. SCAVARDA , Rodrigo PAIVA	273
Normative location of neighborhood's shopping centers: a case study in the district of Barra da Tijuca, Rio de Janeiro, Brazil	
Márcio ROZENTAL , Nélio D. PIZZOLATO	281
Dynamical handling of straddle carriers activities on a container terminal in uncertain environment - a swarm intelligence approach -	
Stefan BALEV , Frédéric GUINAND , Gaëtan LESAUVAGE , Damien OLIVIER	290
A multiagent urban traffic simulation Part I: dealing with the ordinary	
Pierrick TRANOUEZ , Patrice LANGLOIS , Éric DAUDÉ	297
Space-time self-organization for the dynamic VRPTW	
Besma ZEDDINI , Mahdi ZARGAYOUNA , Adnan YASSINE , Moncef TEMANI	302
French ports reform: reasons and perspectives	
Sidi Mohamed OULD MOHAMED MOCTAR	308
Industrial dynamics as applied to raw material supply process modeling: the case of a small company from the building industry	
Charles Henri FREDOUET	321
9 Nonlinear Analysis of Time Series and Applications	329
Localizing epileptic foci using surrogate-baseline corrected nonlinear synchronization measures	
Ralph Gregor ANDRZEJAK , Daniel CHICHARRO , Florian MORMANN , Klaus LEHNERTZ	331
Using a nonlinearity detection as a prior step for global modeling	
Ubiratan S. FREITAS , Christophe LETELLIER	333
Improvement of symbolic transfer entropy	
Dimitris KUGIUMTZIS	338
Assessing the degree of synchronization in time series using symbolic representations	
Roberto MONETTI , Wolfram BUNK , Thomas ASCHENBRENNER , Stephan SPRINGER	343

Characterization of pre-seizure states by a multi-signal analysis of scalp EEG and ECG signals during slow wave sleep	
Stavros NIKOLOPOULOS , Mario VALDERRAMA , P. MILANI , V. NAVARRO , Michel LE VAN QUYEN	347
Complex networks as a tool for nonlinear time series analysis	
Michael SMALL , Xiaoke XU	347
Non-Gaussian signal processing of biological signals	
Max LITTLE	353
10 Dynamic Graphs and Complex Systems	355
A survey of connectivity in mobile ad hoc networks	
Cédric ABOUE NZE , Frédéric GUINAND	357
On the use of social agents for image segmentation	
Richard MOUSSA , Marie BEURTON-AIMAR , Pascal DESBARATS	369
Dynamical systems based on dynamic graphs	
Stéphane DESPRÉAUX , Aude MAIGNAN	376
Dynamic graph as a new type of graphs	
Mounira NEKRI , Abdelkader KHELLADI	379
Index	381

From Nonlinear Dynamics to Biology & Biomedicine

Organizing Committee

- Christophe LETELLIER
CORIA, University of Rouen, France
Christophe.Letellier@coria.fr

Description

Biology and BioMedicine is a field where the systems are complex by nature. There are two opposite ways to investigate the resulting dynamics. One approach consists that a complex system is nothing but the sum of its parts, and that an account of it can be reduced to accounts of individual constituents. The second approach corresponds to the idea that all the properties of a given system (physical, biological, chemical, etc.) cannot be determined or explained by its component parts alone. Instead, the system as a whole determines in an important way how the parts behave. In this latter case, it is important to start with a time series produced by the whole system. The next step is then to use one of the most important results from the nonlinear dynamical systems theory that can be summed up by the Takens theorem: it is possible to reconstruct from a single time series a phase portrait equivalent to the original phase space that cannot be measured. It is therefore relevant to possess powerful technique of analysis to deeply investigate the dynamics underlying biological and biomedical systems. This session is devoted to recent techniques developed in the context of nonlinear dynamical systems theory as well as some applications in biology and/or biomedicine. The aim is to develop a common language and a unified methodology between physicians, biologists and physicists.

Contents

Representations of dynamical systems	
Daniel J. CROSS , Robert GILMORE	3
Stochastic modeling of the mammalian circadian oscillator	
Didier GONZE	8
Close return plots for a analysing patient-ventilator interactions during nocturn ventilation	
Christophe LETELLIER , D. BOUNOIARE , R. NAECK , Ubiratan. S. FREITAS , A. PORTMANN , A. CUVELIER , J.-F. MUIR	13
Controlling the interaction between wild and transgenic mosquitoes	
Marat RAFIKOV , Ana Paula P. WYSE , Luiz BEVILACQUA	20
TBA	
Marc LEFRANC	24
Analysis of the sequential organisation of mice trajectories	
Philippe FAURE , Annick LESNE	24

Representations of Dynamical Systems

Daniel J. Cross and Robert Gilmore, *Fellow, APS*

Abstract—Data from an experiment can be embedded in R^N in many different ways. Different embeddings may provide different information. A set of labels is needed to distinguish inequivalent embeddings, or representations, of the data. We describe how the number of labels decreases as the dimension, N , of the embedding increases. These ideas are illustrated in terms of the representation theory for three dimensional ($d = 3$) dynamical systems, where all appropriate results are available.

Index Terms—Chaos, nonlinear dynamics, embedding, representation, inequivalent, isotopy

I. INTRODUCTION

WHEN you analyze embedded experimental data: What do you learn about the embedding and what do you learn about the underlying dynamics that create the data? This is the first fundamental question confronting the analysis of chaotic experimental data.

The same analysis techniques may provide different information when applied to different embeddings of the same data. It is therefore essential to separate the invariants of the analysis, that point to the fundamental processes responsible for the dynamics, from those (the “variants”) that may differ from one embedding to another.

We show below that embeddings are representations of the dynamics; that different representations may or may not be equivalent; that distinct classes of equivalent representations are labeled by an appropriate set of representation labels; that fewer and fewer labels are necessary to distinguish among representations as the embedding dimension N increases; and that in a sufficiently high dimension all representations become equivalent. We illustrate these ideas for three dimensional dynamical systems ($d = 3$) where three types of labels are needed to distinguish representations in three dimensions ($N = 3$), one type is necessary in four dimensions ($N = 4$), and all representations become equivalent in five and higher dimensions ($N \geq 5$).

II. EMBEDDINGS

The first challenge for the analysis of experimental data is the construction of a suitable representation of the data. This amounts to a mapping of the data into a space that can be used as a phase space to describe the dynamics. This space is typically R^N , with N sufficiently large.

Such a mapping is called an embedding. An embedding (in the sense of Whitney [1]) is a diffeomorphism between the phase space that governs the experimental dynamics and a submanifold in R^N . We are usually concerned with mappings

of data generated by strange attractors into strange attractors in R^N . These are not manifold embeddings, rather they are embeddings in the sense of Takens [2]. In the discussion that follows embeddings should be understood in the sense of Whitney.

When data consist of a single experimental time series taken at equally spaced time intervals, the data must be used to create N -vectors [2], [3]. Standard recipes include time delay embeddings and differential embeddings. When data consist of fields, for example fluid height data $z(x, y; t)$ on a planar grid at equally spaced time intervals, other methods must be used. These include Fourier, Galerkin, and SVD projections [4].

III. REPRESENTATIONS

An embedding is a diffeomorphism (1-1, onto, differentiable with differentiable inverse) between the original phase space and a recreated phase space in R^N . As such, it provides a representation of the dynamics. What we learn about the original dynamics is what we learn by analyzing its representation, the embedding.

Different embeddings (e.g., time delay with different delays, differential-integral) provide different representations of the dynamics. As in the theory of Group Representations we must ask: Are they equivalent or not?

A. Equivalence:

For dynamical systems the natural notion of equivalence is under topological deformation (isotopy). Two embeddings of a dynamical system are defined to be equivalent if they are isotopic [5]. The isotopy smoothly deforms the phase space given by one representation into the phase space provided by the other; at the same time it smoothly deforms the attractor (strange or otherwise) contained in one phase space into that contained in the other.

B. Geometric and Dynamical Measures:

Geometric measures, such as fractal dimensions, are diffeomorphism invariants. As a result they are representation-independent, and in principle it doesn't matter which representation they are determined in. However, implementation of this theorem on experimental data can be problematic. In [6] significantly different estimates of the correlation dimension of experimental data were made on sets that differed by a simple logarithmic transformation.

In principle, dynamical measures such as the spectrum of Lyapunov exponents are also diffeomorphism invariants. However, the number of exponents depends on the embedding dimension, so grows with N . Further, if N grows past the

D. J. Cross and R. Gilmore are in the Physics Department, Drexel University, Philadelphia, PA 19104 USA e-mail: d.j.cross@drexel.edu, robert.gilmore@drexel.edu.

Manuscript received April 15, 2009.

“proper” dimension, “spurious exponents” are determined that are sometimes even larger than the largest Lyapunov exponent in the phase space of appropriately small dimension [7], [8].

To summarize, in principle geometric and dynamical measures are representation-independent quantities.

C. Topological Measures:

A useful theorem by Wu Wen-Tsün [9] guarantees that all embeddings are equivalent in sufficiently high dimension. Specifically, mappings of a d dimensional manifold \mathcal{M}^d into R^N are isotopic when $N \geq 2d + 1$ for $d > 1$. For $d = 1$, $N \geq 2d + 2 = 4$: $N = 3$ is knot theory. For N in the range $d \leq N \leq 2d$ different representations may be inequivalent.

When there are obstructions to isotopy [10], representations fall into equivalence classes. It should be possible to find representation labels that are sufficient to distinguish among these classes. As the dimension increases from $N = d$ to $N = 2d$ these representation labels “fall away”. In higher dimensional spaces there is more room to maneuver so there are fewer obstructions to isotopy. Finally, for $N = 2d + 1$ and larger, no representation labels are necessary to distinguish among representations, since all are equivalent.

The only topological imprint that remains, like the smile of the Cheshire cat, is the mechanism that is responsible for generating chaotic dynamics [11].

IV. REPRESENTATIONS, $d = 3$, $N = 3$

A set of representation labels is known for strange attractors that are generated by three dimensional dynamical systems $d = 3$ and embedded in R^3 [11], [12]. The representation labels are of three types: parity, global torsion, and knot type. The spectrum of these representation labels depends on the genus g of the attractor [12], [13], [14]. The spectrum is simplest in the simplest case, $g = 1$. This case is discussed first. The more general case, $g \geq 3$, is discussed in the following subsection.

Wu’s theorem guarantees that in both cases, all representations are equivalent for $N \geq 7$. In fact, in both cases all representations become equivalent for $N \geq 5$. In dimension five or greater the only remaining topological signal in any representation is the mechanism that creates chaotic motion.

A. Genus $g = 1$

Strange attractors in three dimensions that are created by repetition of the stretching-twisting-folding processes are contained in bounding tori of genus $g = 1$. Such attractors include two-dimensional nonlinear oscillators that are periodically driven as well as autonomous dynamical systems such as the Rössler attractor. For such systems the phase space is $D^2 \times S^1$. All embeddings have a “hole in the middle” when viewed from the proper perspective. Two of the representation labels for this class of attractors are derived from the mapping class group of the torus [10], [11]. The mapping class group of the torus surface $\partial(D^2 \times S^1)$ that preserves direction of flow (longitudes) has discrete representatives with matrix form

$$\begin{bmatrix} 1 & n \\ 0 & \pm 1 \end{bmatrix} \quad (1)$$

The integer index n describes how often a longitude is twisted around the centerline (“core”) of the torus. The index ± 1 is the parity index: it determines the handedness of the diffeomorphism.

The third index required to distinguish among representations of strange attractors with $g = 1$ is directed knot type. The torus $D^2 \times S^1$, and the attractor it contains, can be mapped into R^3 in many different diffeomorphic ways. Specifically, a directed knot $\mathbf{K}(\theta) = (\xi(\theta), \eta(\theta), \zeta(\theta))$ is adopted that satisfies periodic boundary conditions: $\mathbf{K}(\theta) = \mathbf{K}(\theta + 2\pi)$. Harmonic knots [15] provide useful representations for such knots. An attractor in $D^2 \times S^1$ is mapped into R^3 by the mapping $(x(t), y(t), t) \in D^2 \times S^1 \rightarrow \mathbf{K}(\theta) + x(t)\mathbf{n}(\theta) + y(t)\mathbf{b}(\theta) \in R^3$. Here $\mathbf{n}(\theta)$ and $\mathbf{b}(\theta)$ are the unit normal and binormal to $\mathbf{K}(\theta)$ and the angle θ and the time are synchronized according to

$$\frac{t}{T_d} = \frac{\theta}{2\pi} \quad (2)$$

where T_d is the period of the driving term in S^1 [11]. In order to prevent self-intersections from occurring in this mapping of $D^2 \times S^1$ into R^3 we require the attractor be scaled so that the nonlocal minima of the knot distance function $|\mathbf{K}(\theta) - \mathbf{K}(\theta')| > R$, where $\max x(t)^2 + y(t)^2 < R^2$.

The parity index distinguishes among embeddings into $D^2 \times S^1 \subset R^3$ but is no longer an obstruction to isotopy in R^4 [5]. To see this, we lift coordinates $x(t), y(t), z(t)$ from R^3 to R^4 by means of the injection

$$\begin{bmatrix} x(t) \\ y(t) \\ z(t) \\ 0 \end{bmatrix} \xrightarrow{R_{34}(\phi)} \begin{bmatrix} x(t) \\ y(t) \\ z(t) \cos \phi \\ z(t) \sin \phi \end{bmatrix} \xrightarrow{\phi=\pi} \begin{bmatrix} x(t) \\ y(t) \\ -z(t) \\ 0 \end{bmatrix} \quad (3)$$

The projection back down into R^3 reverses the parity index of the original representation. The rotation $R_{34}(\phi)$ is not only an isotopy, it is an isometry as well.

Knot type no longer is an obstruction to isotopy in R^4 . The idea is similar to that presented in Eq. (3). Not only do knots in R^3 fall apart in R^4 , but tori surrounding knots in R^3 can also be untangled in R^4 [5].

Representations with different global torsion in R^3 become equivalent in R^5 [5]. Coordinates in a torus $D^2 \times S^1$ with a global torsion n can be taken in the form $[\theta, re^{i(\phi+n\theta)}]^t$. Here $\theta \in S^1$ and $re^{i\phi}$ is a complex number identifying a point in the disk D^2 . The curve with $r = 0$ is the core of the torus. A homotopy between a torus with $n = 0$ and $n = 1$ is given by

$$\begin{bmatrix} \theta \\ re^{i\phi} \end{bmatrix} \xrightarrow{\text{Inject}} \begin{bmatrix} \theta \\ re^{i\phi} \\ re^{i(\phi+\theta)} \end{bmatrix} \xrightarrow{C_{23}(\psi)} \begin{bmatrix} \theta \\ re^{i(\phi+\theta)} \\ -re^{i\phi} \end{bmatrix} \quad (4)$$

The rotation by $\psi = \pi/2$ radians in the complex 2-3 plane is an isometry as well as an isotopy.

Global torsion continues to distinguish inequivalent representations even when they are embedded in R^4 . The proof is somewhat involved [5]. In summary, it involves linking a longitude (S^1) on the torus surface with a sphere $S^2 \subset R^4$ in such a way that the linking number of these two manifolds is the same as the linking number between the longitude and the core of the torus in R^3 . Since isotopy cannot change linking numbers of the two- and one-dimensional manifolds in R^4 , global torsion remains an obstruction to isotopy in R^4 .

B. Genus $g \geq 3$

Strange attractors in three dimensions that are created by repetition of the stretching-tearing-squeezing processes are contained in bounding tori of genus $g \geq 3$ [13]. Such attractors include many autonomous three dimensional systems with a symmetry, such as the Lorenz attractor.

Each bounding torus of genus $g \geq 3$ can be built up systematically by combining $g - 1$ pairs of stretching and squeezing units [12]. These units are shown in Fig 1. The rules for combining these units are the usual: outputs are connected to inputs and there are no free ends (as is the case for branched manifolds [4]). There is one additional constraint: all connections are colorless. Fig. 2 illustrates the use of two pairs of stretch-squeeze units to build up a genus-three bounding torus of the type that contains the Lorenz attractor. Disks at the entrances of the stretch units (blue) serve to define components of the global Poincaré surface of section. Alternatively, disks at the exits of the squeeze units (yellow) can be used. The global Poincaré surface of section of an attractor contained in a bounding torus of genus g is the union of $g - 1$ disks [13]. The topological period of a closed trajectory in the attractor is the number of its intersections with the disks comprising the Poincaré surface of section.

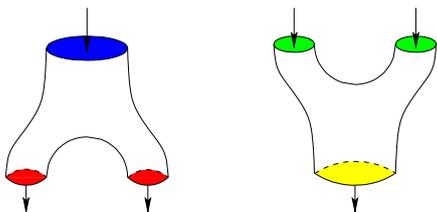


Fig. 1. Stretching (left) and squeezing (right) units are needed to build up a bounding torus. The flow direction is indicated by arrows. Input and output ends are color-coded.

Diffeomorphisms map strange attractors in a genus- g bounding torus to other representations in bounding tori of the same genus and index [13], [14]. The representation labels in this case are similar to the genus-one case, but more extensive [12]. Parity is one of the indices in three-dimensional embeddings. This is no longer an obstruction to isotopy in four-dimensional embeddings by arguments that follow closely those surrounding Eq.(3) [17].

There are a number of integer indices that are related to torsion. A “flow tube” can be inserted into each bounding torus of genus- g as shown in Fig. 3. One such flow tube can

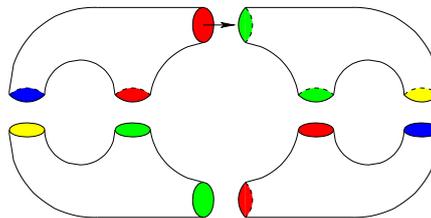


Fig. 2. The bounding torus that contains the Lorenz attractor is a genus-three torus [13]. This is built up using two pairs of stretching and squeezing units as shown. Outflows from one type of unit are inflows to the other. All connections are colorless. The global Poincaré surface of section is a union of blue (or yellow) disks.

be inserted following each outflow. There are three outflow regions for each pair of stretch and squeeze units, and $g - 1$ pairs make up each genus- g torus. As a result, $3(g - 1)$ integer indices distinguish inequivalent representations of strange attractors interior to a genus- g bounding torus. These remain obstructions to isotopy in R^4 by arguments similar to those given above in the genus-one case, involving the links of spheres S^2 and longitudes. In R^5 all representations become equivalent. The argument is somewhat more involved than that surrounding Eq.(4), and is given elsewhere [17].

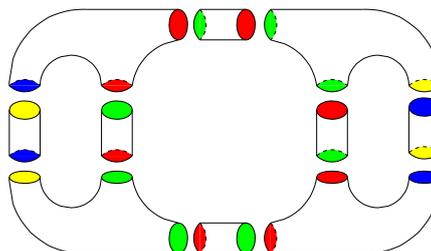


Fig. 3. Flow tubes are inserted at the outflow ends of all stretching and squeezing units. Each can be twisted independently, giving rise to $3(g - 1)$ integer indices describing the “local torsion” in various parts of the attractor’s representation.

Finally, there is an analog to knot type as a representation label. It is related to the link type of the g fundamental longitudes in the genus- g torus. This link index disappears as a representation label for embeddings into R^4 [17].

V. APPLICATION

Two-dimensional nonlinear oscillators $\dot{x} = F_1(x, y)$, $\dot{y} = F_2(x, y)$ typically have a natural time scale T_n . When such systems are periodically driven at an angular frequency $\omega = 2\pi/T_d$, they exhibit behavior that becomes increasingly complicated as the ratio of the drive to the natural time scale T_d/T_n increases.

A rotational transformation with global torsion n is very useful for simplifying analysis of the data. This takes the form

$$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \quad (5)$$

The rotation angle is synchronized to the drive time and global torsion through $\frac{\theta}{2\pi} = n\frac{t}{T_d}$. A useful value of n is typically near $\pm T_d/T_n$. This value can be estimated by computing two simple quantities that measure how the attractor is “wound up”. These are the torsion along the direction of the flow and the (kinetic) energy of the flow [18]:

$$\mathcal{T} = \langle u\dot{v} - v\dot{u} \rangle \quad (6)$$

$$\mathcal{KE} = \frac{1}{2} \langle \dot{u}^2 + \dot{v}^2 \rangle \quad (7)$$

These are time averages of classical quantities. These averages are plotted in Fig. 4 for a number of representations of a strange attractor generated by the van der Pol equations [18].

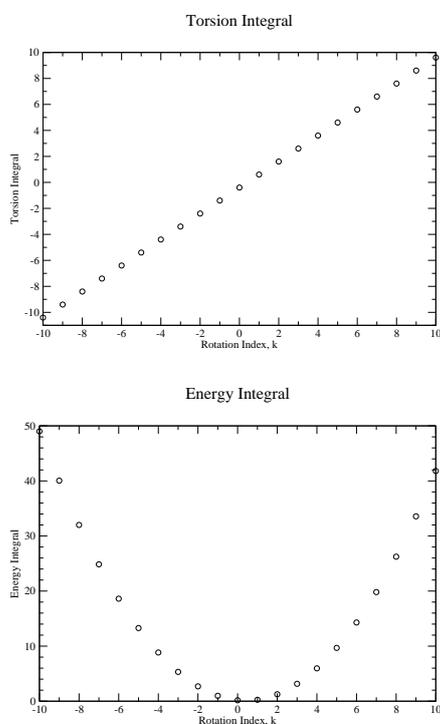


Fig. 4. Plots of the classical integrals \mathcal{T} and \mathcal{KE} for a number of representations of the attractor generated by the van der Pol equations $\dot{x} = by + (c - dx^2)x$, $\dot{y} = -x + a\sin(\omega_d t)$ with $(a, b, c, d, \omega_d) = (0.25, 0.7, 1.0, 10.0, \pi/2)$. The representations differ by the global torsion k .

In general, topological analyses of data become increasingly difficult as the representation of the attractor becomes increasingly wound up. For this reason it is useful to choose the representation of the attractor that minimizes \mathcal{KE} . The topological indices (linking numbers, relative rotation rates [4], [16], [19]) resulting from an analysis of this representation can easily be related to the corresponding indices in representations with other values of global torsion, parity, and knot type. The mechanism is representation-independent.

VI. IMPLICATIONS

Topological analyses of three dimensional embeddings of strange attractors provide information about both the embedding and the underlying physics. Such analyses identify a branched manifold that is a caricature of the strange attractor [4], [16]. This caricature reveals the mechanism responsible for chaotic behavior; that is, how the different branches are stretched and recombined repetitively. As the embedding dimension increases the topological indices fall away as obstructions to isotopy, leaving behind only information about mechanism. For $N \geq 5$, mechanism is the only remaining information available from analysis of an embedding [5]. It should be possible to construct an index in dimensions $N \geq 5$ that identifies the underlying mechanism. Such an index is not yet known.

VII. CONCLUSION

Many representations of experimental data are possible. From an experimental point of view, representations of lower dimension are preferable to those of higher dimension, and those of the lowest possible dimension are the most preferable of all. As the dimension decreases, the number of representation labels required to distinguish inequivalent representations increases. In reverse, as the embedding dimension increases fewer labels are necessary to distinguish inequivalent representations, and in sufficiently high dimension, not greater than $N = 2d + 1$ for d -dimensional dynamical systems ($d > 1$), and 5 for 3-dimensional dynamical systems, all embeddings are equivalent. The useful set of labels needed to distinguish representations of three-dimensional dynamical systems is known. Very little is known about representations of dynamical systems with $d > 3$.

ACKNOWLEDGMENT

The authors would like to thank Prof. Christophe Letellier, Prof. Gregory L. Naber, and Mr. Timothy Jones for useful discussions.

REFERENCES

- [1] H. Whitney, *Differentiable manifolds*, Ann. Math. **37**, 645-680 (1936).
- [2] F. Takens, in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, edited by D. A. Rand and L. S. Young (New York: Springer-Verlag, 1981), vol. 898, pp. 366-381.
- [3] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Geometry from a times series*, Phys. Rev. Lett. **45**, 712-715 (1980).
- [4] R. Gilmore and M. Lefranc, *The Topology of Chaos*, NY: Wiley, 2002.
- [5] Daniel J. Cross and Robert Gilmore, *Representation theory for strange attractors*, unpublished.
- [6] M. Lefranc, D. Hennequin, and P. Glorieux, *Improved correlation dimension estimates through change of variable*, Phys. Lett. **A163**, 269-274 (1992).
- [7] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, *Lyapunov exponents from time series*, Phys. Rev. **A34**, 4971-4979 (1986).
- [8] U. Parlitz, *Identification of true and spurious Lyapunov exponents from time series*, Int. J. Bifurcation Chaos Appl. Sci. Eng. **2**, 155-165 (1992).
- [9] Wu Wen-Tsun, *On the isotopy of C^r -manifolds of dimension n in Euclidean $(2n+1)$ -space*, Science Record, New Ser. Vol **II**, 271-275 (1958).
- [10] D. Rolfsen, *Knots and Links*, Providence, RI: AMS Chelsea Publishing, 1990.
- [11] N. Romanazzi, M. Lefranc, and R. Gilmore, *Embeddings of low-dimensional strange attractors: Topological invariants and degrees of freedom*, Phys. Rev. **E75**, 066214 (2007).

- [12] R. Gilmore, C. Letellier, and N. Romanazzi, *Global topology from an embedding*, J. Phys. A: Math. Theor. **40**, 13291-13297 (2007).
- [13] T. D. Tsankov and R. Gilmore, *Strange attractors are classified by bounding tori*, Phys. Rev. Lett. **91**, 134104 (2003).
- [14] R. Gilmore and C. Letellier, *The Symmetry of Chaos*, Oxford: University Press, 2007.
- [15] A. K. Trautwein, *Harmonic knots*, Ph. D. Thesis, July 1995, Department of Mathematics, University of Iowa, Iowa City, Iowa.
- [16] R. Gilmore, *Topological analysis of chaotic dynamical systems*, Rev. Mod. Phys. **70**, 1455-1530 (1998).
- [17] Daniel J. Cross and Robert Gilmore, *Complete set of representations for three-dimensional dynamical systems*, unpublished.
- [18] R. Gilmore, *Two-parameter families of strange attractors*, Chaos **17**, 013104 (2007).
- [19] H. G. Solari and R. Gilmore, *Relative rotation rates for driven dynamical systems*, Phys. Rev. A **37**, 3096-3109 (1988).



Daniel J. Cross received a B.A. degree from the physics department and a B.S. degree from the mathematics department of Cedarville University in 2002. He received his M.S. in physics from Drexel University in 2005 and is current working toward his Ph.D. in physics at Drexel. His research interests include topological aspects of nonlinear dynamics and group theory.



Robert Gilmore received B.S. degrees from the Physics Department (1962) and from the Mathematics Department (1962) at M.I.T. and his Ph.D. from the Physics Department at M.I.T. in 1967. He has been professor of physics at the University of South Florida (1971-1979) and at Drexel University since 1981. He spent 1979-1981 at I.D.A. in Arlington, VA, and has been visiting professor at Alabama A. & M. University and a number of european universities. He has spent sabbatical years at Louvain la Neuve, University of Lille, and INSA Rouen.

Stochastic modeling of the mammalian circadian oscillator

Didier Gonze

Abstract—In mammals circadian rhythms are generated autonomously at the molecular level in the neurons of the suprachiasmatic nucleus of the hypothalamus. In these cells, oscillations in the expression of the clock genes and in the concentration of the clock proteins are generated through a complex gene regulatory mechanism involving interlocked positive and negative feedback loops. Deterministic models of this network account for the occurrence of autonomous circadian oscillations and for their entrainment by light-dark cycles and are used to study the molecular bases of clock-related pathologies. Stochastic versions of these models take into consideration the molecular fluctuations that arise when the number of molecules involved in the regulatory mechanism is low. We previously studied the robustness of the oscillations with respect to noise using a simple model based on the core regulatory mechanism of the circadian clock. Here we extend this analysis to the more detailed model for circadian rhythms in mammals. Our numerical simulations show that robust oscillations can occur even when the number of mRNA and protein molecules oscillates with a maximum around 100. Interestingly, we notice that robust oscillations in some genes do not require that all clock components exhibit robust oscillations. We also show that, in presence of noise, entrainment by a light-dark cycle might be lost.

Index Terms—circadian rhythms, mammalian circadian network, stochastic simulations, robustness to molecular noise

I. INTRODUCTION

CIRCADIAN rhythms are observed in most living organisms, from cyanobacteria to plants, insects, and mammals [1]. These 24-hour rhythms allow living organisms to live in phase with the alternance of day and night and are observed at every level of the physiological and cellular organisation [2]. Alterations of the circadian clock have been linked with physiological disorders. In mammals, circadian rhythms are generated autonomously at the molecular level in the neurons of the suprachiasmatic nucleus of the hypothalamus, which constitutes the pacemaker of circadian oscillations [2,3]. In these cells, oscillations in the expression of the clock genes as

well as in the concentration of the corresponding proteins are generated through a complex gene regulatory mechanism involving interlocked positive and negative feedback loops [4].

The dynamics of the circadian clockwork has been investigated using mathematical models. One of the first molecular models showed that a delayed negative feedback loop involving a single gene was sufficient to generate self-sustained circadian oscillations [5]. This auto-regulation constitutes the core of the circadian clockwork. However, with the accumulation of molecular data, the network was shown to be more complex: it involves several clock genes and several regulatory feedback loops. Subsequent mathematical models were then developed to integrate this complexity [6,7]. These detailed models account for the occurrence of autonomous circadian oscillations and for their entrainment by light-dark cycles and were used to study the molecular bases of clock-related pathologies [8]. Based on ordinary differential equations, they are however of a deterministic nature and thus neglect the fluctuations due to molecular noise. Because of the low number of molecules involved in the molecular mechanism responsible for circadian rhythms, the effect of noise may impair the robustness of circadian oscillations [9,10]. To assess the impact of molecular noise on the dynamical behaviour of the system, it is thus needed to resort to stochastic simulations.

In a previous work, we performed stochastic simulations of Goldbeter's 5-variable model [11]. Numerical simulations of this model have been performed by means of the Gillespie method [12]. The results indicated that robust circadian oscillations can occur even when the numbers of molecules of mRNA and nuclear proteins involved in the oscillatory mechanism are reduced to a few hundreds. We further showed that entrainment by light-dark cycles and cooperativity in repression enhance the robustness of circadian oscillations with respect to molecular noise. Here, we extend this analysis to a detailed model for circadian rhythms in mammals.

II. MODEL AND SIMULATIONS

In this work we considered the detailed model developed by Leloup & Goldbeter [6]. The model incorporates the major components of the circadian clock, namely the genes *Per*, *Cry*, and the CLOCK/BMAL1 complex, and two interlocked regulatory feedback loops (fig. 1). A positive loop is based on the activation of the expression of the *Bmal1* gene by the

Manuscript received May 1, 2009.

Didier Gonze works at the Faculté des Sciences, CP 231, Université Libre de Bruxelles Bvd du Triomphe, 1050 Bruxelles, Belgium (e-mail: dgonze@ulb.ac.be).

This work was supported by grant n° 3.4607.99 from the *Fonds de la Recherche Scientifique Médicale* (F.R.S.M., Belgium), by the Belgian Federal Science Policy Office (IAP P6/25 "BioMaGNet": "Bioinformatics and Modeling- From Genomes to Networks"), and by the European Union through the Network of Excellence BioSim, Contract No. LSHB-CT-2004-005137.

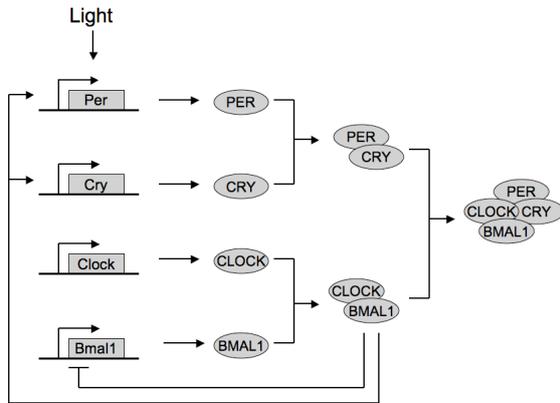


Fig. 1. Simplified scheme of the mammalian circadian oscillator

CLOCK-BMAL1 complex. In the negative loop, the products of the *Per* and *Cry* genes form a PER-CRY complex that can bind to and inactivates the CLOCK-BMAL1 activator, thereby preventing the transcription of *Per* and *Cry* genes. The model also takes into account the reversible phosphorylations of the clock proteins, as well as their nuclear transport. In an extended version of the model, an additional feedback loop involving *Rev-Erba* is included. Without and with the *Rev-Erba* feedback loop, the model counts 16 or 19 variables, respectively and their evolution is described by ordinary differential equations.

We then need to establish the stochastic version of the 16-variable model for the mammalian circadian clock. To this end we describe the reaction steps as birth and death processes. In our model the reaction steps consist of the synthesis and degradation of mRNA, and synthesis, degradation, phosphorylation/dephosphorylation, and nuclear transport of proteins, as well as protein complex formation. In total, there are 42 reaction steps. Note that we did not develop Michaelian and Hill kinetics into elementary reaction steps. This means that we compute the propensity of each reaction directly using the non-linear kinetic functions. In a previous work, we compared the developed and non-developed approaches for the core model for circadian rhythms and showed that both approaches yield largely similar results [13]. The system is then simulated by the exact algorithm proposed by Gillespie [12]. The list of the reactions steps and parameter values can be found in the Appendix.

III. RESULTS

In figure 2, we compare the deterministic (fig. 2A) and stochastic (fig. 2B) time series obtained by numerical simulation of the mammalian circadian model. For the default parameter values, we observe that the stochastic oscillations in the level of *Per* and *Cry* mRNA (variables M_P and M_C) are relatively robust, although the number of molecules does not exceed 100 molecules. The levels of the PER and CRY protein

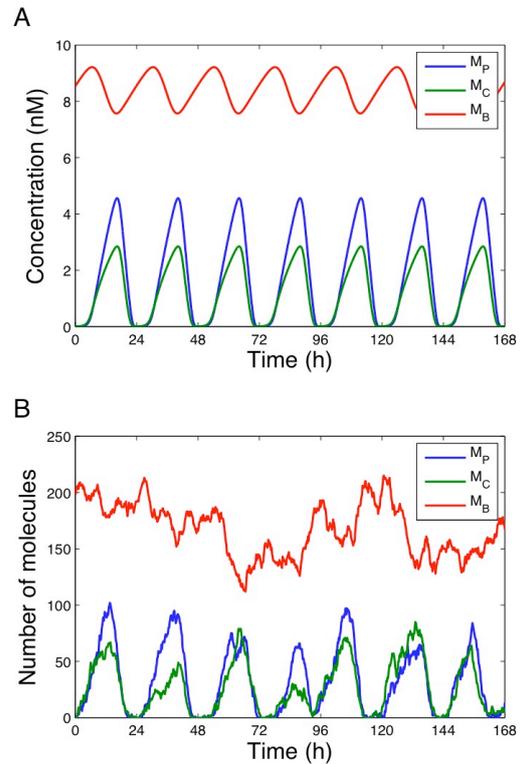


Fig. 2. (A) Deterministic and (B) stochastic oscillations obtained by numerical simulation of the mammalian circadian oscillator using the default parameter values.

also oscillate with a clear circadian period and their amplitudes are of the same magnitude as the mRNA levels (not shown). In contrast, in presence of noise, the level of *Bmal1* mRNA (variable M_B) does not display a clear periodicity. Yet, the corresponding deterministic time series predicts that this protein should oscillate. These results show that noise can obliterate oscillations and that robust oscillation in *Bmal1* mRNA is not required to observe robust oscillations in *Per* and *Cry* levels. This is consistent with the observation of Leloup & Goldbeter [8] who remarked that, even in absence of oscillations in the expression of the *Bmal1* gene, *Per* and *Cry* mRNA can still undergo self-sustained oscillations.

Leloup & Goldbeter showed that in absence of *per* expression (i.e. when its synthesis rate $v_{sp}=0$), the system converges towards a stable steady state (unless other parameters are changed) [6]. In figure 3, we show both the deterministic (fig. 3A) and stochastic (fig. 3B) time series obtained for the same parameter values as in figure 2 except $v_{sp}=0$. In this condition, the *Per* gene is not expressed and the PER protein is not produced. Looking at the transients of the deterministic evolution of the system, we can nevertheless notice a high amplitude peak of *Cry* and *Bmal1* mRNA levels before they undergo damped oscillations. This explains why in the stochastic time series, large amplitude oscillations in the level of *Cry* mRNA are observed. Because the noise

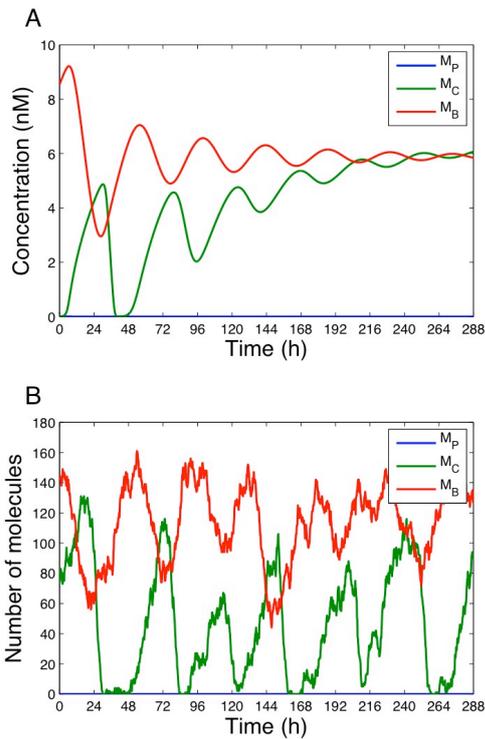


Fig. 3. (A) Deterministic and (B) stochastic time series obtained by numerical simulation of the mammalian circadian oscillator for the default parameter values except $v_{sp}=0$.

continuously moves the system out of its steady state, the evolution of *Cry* and *Bmal1* mRNA repetitively undergo large peaks. The amplitude and peak-to-peak intervals are highly variable and obviously do not reflect circadian oscillations.

In natural conditions, living organisms are subject to the alternance of day and night, and thus, in order to allow the organisms to stay in phase with the day, the circadian oscillator must be appropriately entrained by a light-dark cycle. In mammals, light was shown to induce the expression of the *Per* genes. This is in contrast with the situation encountered in the fly *Drosophila*, where light interacts with the circadian clock by inducing the degradation of one the clock protein. Using a periodic forcing of the parameter controlling *Per* expression, deterministic simulations have permitted Leloup & Goldbeter to identify the conditions in which the circadian oscillator is entrained and appropriately phase-locked [6]: the maximum of *Per* mRNA always occurs in the middle of the dark phase (fig. 4A).

We performed stochastic simulation of the system in the same conditions (fig. 4B). Interestingly, we observed that, in presence of noise, the system is not phase-locked any more. The time of the maximum of *Per* mRNA varies from one cycle to another. This lack of entrainment is better reflected by the rapid damping of the auto-correlation function (fig. 5). Contrarily to our previous results that featured a core model

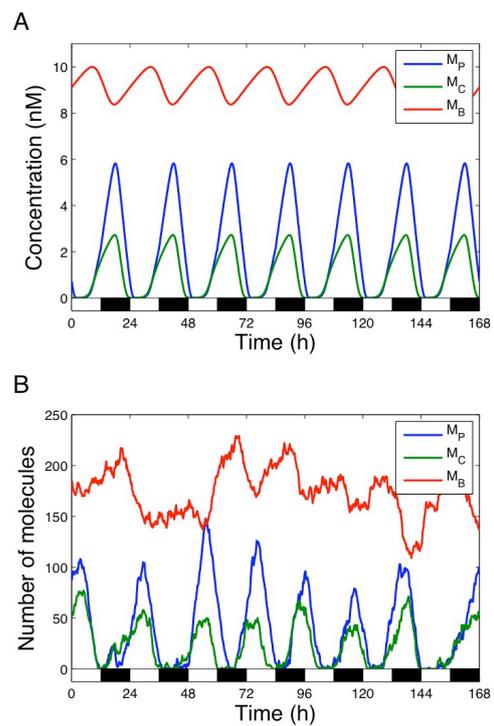


Fig. 4. (A) Deterministic and (B) stochastic oscillations obtained by numerical simulation of the mammalian circadian oscillator subject to a periodic forcing. Parameter v_{sp} varies as a square wave between 1.5 during the dark phase (represented by black rectangles) and $v_{sp}=1.8$ during the light phase (represented by white rectangles).

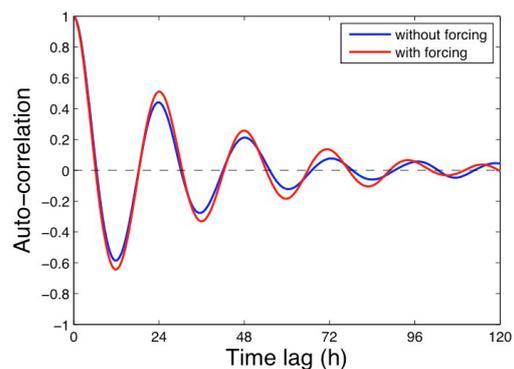


Fig. 5. Auto-correlation functions of the *Per* mRNA time series obtained for the autonomous system (blue curve) and for the periodically forced system (red curve).

for the circadian clock with a periodic forcing of a clock protein degradation as in *Drosophila* (and for which the auto-correlation function does not damp out) [11], the detailed mammalian circadian oscillator is not stably entrained by a periodic forcing.

IV. DISCUSSION AND PERSPECTIVES

Molecular noise, resulting from the low number of molecules involved in cellular processes, may affect the robustness of circadian oscillations [9,14]. To assess the impact of molecular noise on the mammalian circadian oscillator we considered here the stochastic version of the model of Leloup & Goldbeter [6]. Our results corroborate the conclusions of our previous work using a core model for circadian oscillations [11]: robust autonomous oscillations can still be observed when the maximum number of mRNA and protein molecules is around 100. This observation is also in agreement with the conclusions of Forger & Peskin [15], who performed stochastic simulation of another detailed mammalian circadian oscillator. The present simulations furthermore show that robust oscillations in one of the key variables, namely *Bmal1*, are not required to guarantee robust oscillations in the expression of the *Per* and *Cry* genes.

Deterministic simulations also predict that, upon the deletion of one gene, the system would evolve to a stable steady state. Here we have seen that, under the influence of the noise, the system may nevertheless display large-amplitude fluctuations. These results thus suggest that experimental data should be interpreted with caution: it is not because one variable of the system does not show clear oscillations that the whole system does not oscillate, and, conversely, it is not because one variable presents large peaks that the system oscillates. In this context, it should be noticed that some systems characterized by a relaxation behaviour may present noise-induced oscillations, i.e. periodic oscillations that occur only in presence of noise [16].

The present stochastic simulations also show that in presence of noise, entrainment by a light-dark cycle can be lost. We can nevertheless not rule out the possibility that for other parameter sets stable entrainment may occur. Using a core model for circadian clock, we previously showed that entrainment by a light-dark cycle has a stabilizing effect on the phase [11]. It is also possible that other regulatory mechanisms might play a role in the robustness of the oscillations upon entrainment.

In subsequent studies, it would be interesting to investigate further parameters responsible for the robustness of the oscillations, such as the binding/unbinding rates associated to the binding of the regulatory proteins to their target genes. This would imply to decompose the model into detailed reaction steps, as shown in [17]. The full clockwork is also known to involve additional feedback loops, including the RevErb α regulatory loop. It would be interesting to check if these additional feedback loops contribute to the robustness of the oscillations with respect to molecular noise. Mathematical and experimental analyses of a synthetic oscillator suggested that interlocked positive and negative feedback loops may be a means to increase the robustness of the oscillator to noise [18]. Finally, here we considered the oscillator in a single cell. It is known however that neurons in the SCN are coupled through neurotransmitters and gap junctions. This intercellular

coupling allows the synchronisation between the individual cellular oscillators, but may also provide the system with a higher robustness. Preliminary studies already showed that coupling indeed increases the robustness of the oscillations in *Drosophila* [19] and *Neurospora* [14]. These results could be extended to the mammalian systems. More generally, it would be interesting to understand if the complexity of the circadian network may have a role in its robustness to molecular noise.

APPENDIX

Reaction steps and parameters values are listed in Tables 1 and 2, respectively.

ACKNOWLEDGMENT

I would like to thank A. Goldbeter, J.-C. Leloup and J. Halloy for fruitful discussions and K. Faust for careful reading of the manuscript.

REFERENCES

- [1] M.W. Young and S.A. Kay, "Time zones: a comparative genetics of circadian clocks", *Nat Rev Genet.* 2:702-15, 2001.
- [2] S.M. Reppert and D.R. Weaver, "Coordination of circadian timing in mammals", *Nature* 418:935-41, 2002.
- [3] H. Dardente and N. Cermakian, "Molecular circadian rhythms in central and peripheral clocks in mammals", *Chronobiol Int.* 24:195-213, 2007.
- [4] P.E. Hardin, "Transcription regulation within the circadian clock: the E-box and beyond", *J. Biol. Rhythms.* 19:348-60, 2004.
- [5] A. Goldbeter, "A model for circadian oscillations in the *Drosophila* period protein (PER)", *Proc Biol Sci* 261:319-24, 1995.
- [6] J.-C. Leloup and A. Goldbeter, "Toward a detailed computational model for the mammalian circadian clock", *Proc Natl Acad Sci USA.* 100:7051-6, 2003.
- [7] D.B. Forger and C.S. Peskin, "A detailed predictive model of the mammalian circadian clock." *Proc Natl Acad Sci USA* 100:14806-11, 2003.
- [8] J.-C. Leloup and A. Goldbeter, "Modeling the circadian clock: from molecular mechanism to physiological disorders", *Bioessays.* 30:590-600, 2008.
- [9] N. Barkai and S. Leibler, "Circadian clocks limited by noise", *Nature.* 403:267-8, 2000.
- [10] A. Raj and A. van Oudenaarden, "Nature, nurture, or chance: stochastic gene expression and its consequences", *Cell* 135:216-26, 2008.
- [11] D. Gonze, J. Halloy, and A. Goldbeter, "Robustness of circadian rhythms with respect to molecular noise", *Proc Natl Acad Sci USA.* 99:673-8, 2002.
- [12] D.T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", *J. Phys. Chem.* 81:2340-61, 1977.
- [13] D. Gonze, J. Halloy, and A. Goldbeter, "Deterministic versus stochastic models for circadian rhythms". *J. Biol. Phys.* 28: 637-653, 2002.
- [14] D. Gonze and A. Goldbeter, "Circadian rhythms and molecular noise", *Chaos.* 16:026110, 2006.
- [15] D.B. Forger and C.S. Peskin, "Stochastic simulation of the mammalian circadian clock", *Proc Natl Acad Sci USA.* 102:321-4, 2005.
- [16] J.M. Vilar, H.Y. Kueh, N. Barkai, and S. Leibler, "Mechanisms of noise-resistance in genetic oscillators", *Proc Natl Acad Sci USA.* 99:5988-92, 2002.
- [17] D. Gonze, J. Halloy, and A. Goldbeter, "Emergence of coherent oscillations in stochastic models for circadian rhythms". *Physica A* 342:221-233, 2004.
- [18] J. Stricker, S. Cookson, M.R. Bennett, W.H. Mather, L.S. Tsimring, and J.A. Hasty, "Fast, robust and tunable synthetic gene oscillator", *Nature.* 456:516-9, 2008.
- [19] H.R. Ueda, K. Hirose, and M. Iino, "Intercellular coupling mechanism for synchronized and noise-resistant circadian oscillators", *J Theor Biol.* 216:501-12, 2002.

TABLE I
STOCHASTIC VERSION OF THE MAMMALIAN CIRCADIAN CLOCK

No	Reaction	Propensity
1	$\rightarrow M_P$	$w_1 = v'_{sp} \frac{B_N^n}{K'_{ap} + B_N^n}$
2	$M_P \rightarrow$	$w_2 = v'_{mp} \frac{M_P}{K'_{mp} + M_P} + k_{dn} M_P$
3	$\rightarrow M_C$	$w_3 = v'_{sc} \frac{B_N^n}{K'_{ac} + B_N^n}$
4	$M_C \rightarrow$	$w_4 = v'_{mc} \frac{M_C}{K'_{mc} + M_C} + k_{dn} M_C$
5	$\rightarrow M_B$	$w_5 = v'_{sb} \frac{K'^m_{ib}}{K'^m_{ib} + B_N^m}$
6	$M_B \rightarrow$	$w_6 = v'_{mb} \frac{M_B}{K'_{mb} + M_B} + k_{dn} M_B$
7	$\rightarrow P_C$	$w_7 = k_{sp} M_P$
8	$P_C \rightarrow P_{CP}$	$w_8 = v'_{1p} \frac{P_C}{K'_p + P_C}$
9	$P_{CP} \rightarrow P_C$	$w_9 = v'_{2p} \frac{P_{CP}}{K'_{dp} + P_{CP}}$
10	$P_{CC} \rightarrow P_C + C_C$	$w_{10} = k_4 P_{CC}$
11	$P_C + C_C \rightarrow P_{CC}$	$w_{11} = k'_3 P_C C_C$
12	$P_C \rightarrow$	$w_{12} = k_{dpp} P_C$
13	$\rightarrow C_C$	$w_{13} = k_{sc} M_C$
14	$C_C \rightarrow C_{CP}$	$w_{14} = v'_{1c} \frac{C_C}{K'_p + C_C}$
15	$C_{CP} \rightarrow C_C$	$w_{15} = v'_{2c} \frac{C_{CP}}{K'_{dp} + C_{CP}}$
16	$C_C \rightarrow$	$w_{16} = k_{dnc} C_C$
17	$P_{CP} \rightarrow$	$w_{17} = v'_{dpc} \frac{P_{CP}}{K'_d + P_{CP}} + k_{dn} P_{CP}$
18	$C_{CP} \rightarrow$	$w_{18} = v'_{dcc} \frac{C_{CP}}{K'_d + C_{CP}} + k_{dn} C_{CP}$
19	$P_{CC} \rightarrow P_{CCP}$	$w_{19} = v'_{1pc} \frac{P_{CC}}{K'_p + P_{CC}}$
20	$P_{CCP} \rightarrow P_{CC}$	$w_{20} = v'_{2pc} \frac{P_{CCP}}{K'_{dp} + P_{CCP}}$
21	$P_{CN} \rightarrow P_{CC}$	$w_{21} = k_5 P_{CN}$
22	$P_{CC} \rightarrow P_{CN}$	$w_{22} = k'_1 P_{CC}$
23	$P_{CC} \rightarrow$	$w_{23} = k_{dn} P_{CC}$
24	$P_{CN} \rightarrow P_{CNP}$	$w_{24} = v'_{3pc} \frac{P_{CN}}{K'_p + P_{CN}}$
25	$P_{CNP} \rightarrow P_{CN}$	$w_{25} = v'_{4pc} \frac{P_{CNP}}{K'_{dp} + P_{CNP}}$
26	$P_{CN} + B_N \rightarrow I_N$	$w_{26} = k'_7 B_N P_{CN}$
27	$I_N \rightarrow P_{CN} + B_N$	$w_{27} = k_8 I_N$
28	$P_{CN} \rightarrow$	$w_{28} = k_{dn} P_{CN}$
29	$P_{CCP} \rightarrow$	$w_{29} = v'_{dpc} \frac{P_{CCP}}{K'_d + P_{CCP}} + k_{dn} P_{CCP}$
30	$P_{CNP} \rightarrow$	$w_{30} = v'_{dpcn} \frac{P_{CNP}}{K'_d + P_{CNP}} + k_{dn} P_{CNP}$
31	$\rightarrow B_C$	$w_{31} = k_{sb} M_B$
32	$B_C \rightarrow B_{CP}$	$w_{32} = v'_{1b} \frac{B_C}{K'_p + B_C}$
33	$B_{CP} \rightarrow B_C$	$w_{33} = v'_{2b} \frac{B_{CP}}{K'_{dp} + B_{CP}}$

TABLE I (CONT.)
STOCHASTIC VERSION OF THE MAMMALIAN CIRCADIAN CLOCK

No	Reaction	Propensity
34	$B_C \rightarrow B_N$	$w_{34} = k_5 B_C$
35	$B_N \rightarrow B_C$	$w_{35} = k_6 B_N$
36	$B_C \rightarrow$	$w_{36} = k_{dn} B_C$
37	$B_{CP} \rightarrow$	$w_{37} = v'_{dbc} \frac{B_{CP}}{K'_d + B_{CP}} + k_{dn} B_{CP}$
38	$B_N \rightarrow B_{NP}$	$w_{38} = v'_{3b} \frac{B_N}{K'_p + B_N}$
39	$B_{NP} \rightarrow B_N$	$w_{39} = v'_{4b} \frac{B_{NP}}{K'_{dp} + B_{NP}}$
40	$B_N \rightarrow$	$w_{40} = k_{dn} B_N$
41	$B_{NP} \rightarrow$	$w_{41} = v'_{dnn} \frac{B_{NP}}{K'_d + B_{NP}} + k_{dn} B_{NP}$
42	$I_N \rightarrow$	$w_{42} = v'_{din} \frac{I_N}{K'_d + I_N} + k_{dn} I_N$

Note that in the stochastic version of the model the variables are numbers of molecules and not concentrations as in the deterministic model. Some parameter values should be adapted accordingly. They are noted by a prime. Their values are modulated by the system size Ω as defined in Table 2.

TABLE 2
PARAMETER VALUES

v'_{sp}	$1.5 \times \Omega$	molec/h	k_{dnc}	0.12	/h
K'_{ap}	$0.7 \times \Omega$	molec	v'_{dpc}	$0.7 \times \Omega$	molec/h
n	4		v'_{dcc}	$0.7 \times \Omega$	molec/h
v'_{mp}	$1.1 \times \Omega$	molec/h	K'_d	$0.3 \times \Omega$	molec
K'_{mp}	$0.31 \times \Omega$	molec	v'_{jpc}	$0.4 \times \Omega$	molec/h
k_{dn}	0.01	/h	v'_{2pc}	$0.1 \times \Omega$	molec/h
v'_{sc}	$1.1 \times \Omega$	molec/h	k_2	0.2	/h
K'_{ac}	$0.6 \times \Omega$	molec	k_l	0.4	/h
v'_{mc}	$1 \times \Omega$	molec/h	v'_{3pc}	$0.4 \times \Omega$	molec/h
K'_{mc}	$0.4 \times \Omega$	molec	v'_{dpc}	$0.1 \times \Omega$	molec/h
v'_{sb}	$1 \times \Omega$	molec/h	k_7	$0.5 / \Omega$	/h
K'_{ib}	$2.2 \times \Omega$	molec	k_8	0.1	/h
m	2		v'_{dpc}	$0.7 \times \Omega$	molec/h
v'_{mb}	$0.8 \times \Omega$	molec/h	v'_{dpcn}	$0.7 \times \Omega$	molec/h
K'_{mb}	$0.4 \times \Omega$	molec	k_{sb}	0.12	/h
k_{sp}	0.6	/h	v'_{1b}	$0.5 \times \Omega$	molec/h
v'_{1p}	$0.4 \times \Omega$	molec/h	v'_{2b}	$0.1 \times \Omega$	molec/h
K'_p	$0.1 \times \Omega$	molec	k_5	0.4	/h
v'_{2p}	$0.3 \times \Omega$	molec/h	k_6	0.2	/h
K'_{dp}	$0.1 \times \Omega$	molec	v'_{dbc}	$0.5 \times \Omega$	molec/h
k_d	0.2	/h	v'_{3b}	$0.5 \times \Omega$	molec/h
k_3	$0.4 / \Omega$	/molec/h	v'_{4b}	$0.2 \times \Omega$	molec/h
k_{sc}	1.6	/h	v'_{dnn}	$0.6 \times \Omega$	molec/h
v'_{1c}	$0.6 \times \Omega$	molec/h	v'_{din}	$0.8 \times \Omega$	molec/h
v'_{2c}	$0.1 \times \Omega$	molec/h			

For all the simulations presented in this paper, $\Omega=20$. Note also that in fig. 3, $v_{sp}=0$ and in fig. 4, $v_{sp}=1.5 \times \Omega$ (during the dark phases) and $1.8 \times \Omega$ (during the light phases).

Close Return Plots for a Analysing Patient-ventilator Interactions during Nocturn Ventilation

C. Letellier¹, D. Bounoiare², R. Naeck², U. S. Freitas¹, A. Portmann², A. Cuvelier² & J.-F. Muir²

Abstract—Biomedical systems as the cardio-respiratory system can be easily considered as a nonlinear dynamical system. Although the underlying determinism is still difficult to evidence, tools borrowed from the nonlinear dynamical systems theory are appropriate to investigate them. A good example is provided by recurrence plots when applied to noninvasive mechanical ventilation. Using a Shannon entropy computed from the recurrence plots built on the pressure maxima during ventilatory cycles and on total time duration, respectively, patient-ventilator interactions can be divided into four typical groups, one corresponding to optimal synchronization between the patient and its ventilator. Then the sleep fragmentation can be estimated by the Shannon entropy computed from the recurrence plots built from the EEG scoring into sleep stages.

Index Terms—Chaos, nonlinear dynamics, nonlinearity. Shannon entropy, Recurrence plots, cardio-respiratory system, Sleep

I. INTRODUCTION

AMONG the natural rhythms that are relevant for life, breathing is certainly one of the most often perceived as fluctuating depending on various parameters. We always alternate inspirations and expirations but such a cycling depends on our activity, our stress, etc. For short it depends on the surrounding world and on our behavior in this ambient world. It is known that breathing is not exactly periodic but a decisive proof for its chaotic nature is still lacking [1], [2], [3]. Such a feature does not prevent to use tools borrowed from the nonlinear dynamical systems theory to investigate — to characterize — its related dynamics. Biomedical dynamics are necessarily noise contaminated in the sense that the ambient world has always an impact on the subjects. Moreover, performing a measurement, even in a noninvasive manner, necessarily affects the subjects and, consequently, the dynamics underlying the mechanism under study. Often, the simple fact that a patient knows that something is measured suffices to change his cardiac or/and breathing rhythms. Such a noise contamination can contribute to develop — to make more complex — the dynamics and, unfortunately, this cannot be avoided.

It is thus required to develop techniques that are taking into account the most robust properties of the dynamics. Among others, recurrence properties are not too affected by noise contamination and are easily investigated using the so-called recurrence plots [4]. Recurrence plots analysis are often used in biomedicine [4], [5] and this work belongs to this paradigm.

¹ CORIA UMR 6614 — University of Rouen, France and ² GRHV EA 3830 — Rouen University Hospital, France, e-mail: Christophe.Letellier@coria.fr

Manuscript received May 22, 2009.

Two examples of applications of recurrence plot analysis will be discussed in the context of noninvasive mechanical ventilation. First, patient-ventilator interactions will be characterized using two Shannon entropies computed from recurrence plots built from pressure signal and time duration of respiratory cycles, respectively. Second, the sleep fragmentation will be quantified with a Shannon entropy based on a recurrence plots built from a sleep stage scoring. It will be shown how these quantities can bring some light to questions asked by the physicians.

II. CLOSE RECURRENCE PLOTS

Recurrence plots were introduced by Eckmann *et al* [6] and some quantifiers were later introduced to convert recurrence plots into a statistical analysis [7]. Among these quantifiers, it is possible to properly define a “Shannon entropy” which was found to be correlated to the largest Lyapunov exponent [8].

A *recurrence plot* R_{ij} is built as follows. Every point of the phase space trajectory $\{\mathbf{x}_i\}_{i=1}^N$ is tested whether it is close to another point \mathbf{x}_j of the trajectory, that is, whether the distance between these two points is less than a specified threshold ϵ . In this case, the point is said to be recurrent and is represented by a black dot. Otherwise, the point is not recurrent and is represented by a white dot. This can be described as a $N \times N$ array

$$R_{ij} = \theta(\epsilon - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (1)$$

where $\theta(\mathbf{x}_i)$ is the Heaviside function. When built using this definition, recurrence plots are symmetrical with respect to the diagonal and, consequently, they used twice the same information. In order to optimize the representation, Mindlin and Gilmore used a so-called “close return plots” defined as a $N \times \tau$ array

$$C_{ij} = \theta(\epsilon - \|\mathbf{x}_i - \mathbf{x}_{i-j}\|) \quad (2)$$

where $j \in [1, \tau]$ with τ defining the size of the window over which close returns are checked. These plots present the advantage to have horizontal and vertical lines as the main structuring axes which are, compared to diagonals, slightly easier to use for computations and interpretations. Thus, close return plots were used by Gabriel Mindlin and Robert Gilmore for extracting periodic orbits from experimental data recorded in a Belousov-Zhabotinski reaction [9] and by Claire Gilmore for showing that there is no obvious signature of deterministic chaos in financial data [10]. Two examples of these close return plots are shown in Fig. 1 for the Logistic map

$$x_{n+1} = \mu x_n(1 - x_n). \quad (3)$$

Periodic behaviors are easily identified by parallel horizontal lines spaced by the period of orbits (a period-4 limit cycle in the case of Fig. 1a). Chaotic behaviors present more jittery close return plots with few recurrent horizontal segments as a signature of the deterministic nature of the underlying dynamics (Fig. 1b).

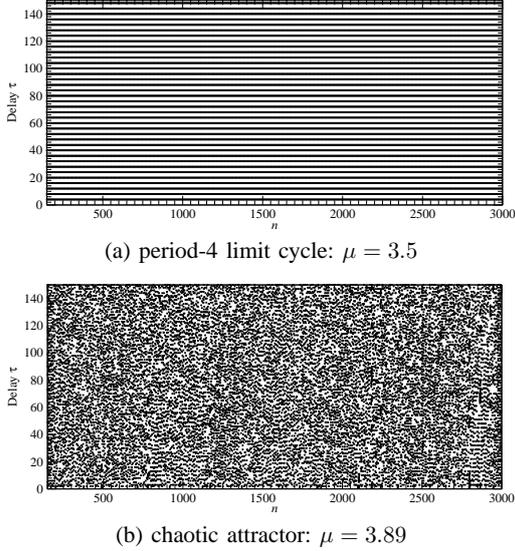


Fig. 1. Close return plots computed for the Logistic map.

As for the usual recurrence plots, it is possible to compute a Shannon entropy according to

$$S_{\text{RP}} = - \sum_{n=1}^H P_n \log(P_n). \quad (4)$$

where $P_n \neq 0$ corresponds to the number of non-recurrent horizontal segments with length $n > 0$ divided by the total number of recurrent points. If the number of recurrent points is not included in this definition, the entropy provides more or less a yes-or-no estimator [8]. With this definition, the Shannon entropy quantifies the complexity of the dynamics and is correlated to the largest Lyapunov exponent (Fig. 2). Such a feature suggests that such a Shannon entropy can behave as a Kolmogorov-Sinai entropy known to be correlated to the largest Lyapunov exponent as stated by the Pesin conjecture [11]. As for the recurrence plots, a working Shannon entropy can be defined and used as an estimation of the largest Lyapunov exponent.

As any estimation of a Shannon entropy or other dynamical invariants, computations may depend on some parameters. In the present case, estimations of the Shannon entropy depend on the number of points retained for computing the close return plots. Nevertheless, with a maximum time delay τ equal to 100 (the smallest reasonable value for τ_{max}), it is found that the Shannon entropy value does not depend too much on the number of points until this number is greater than 500 (Fig. 3). This feature tells us how long must be the data set to have a rather safe estimation of the Shannon entropy.

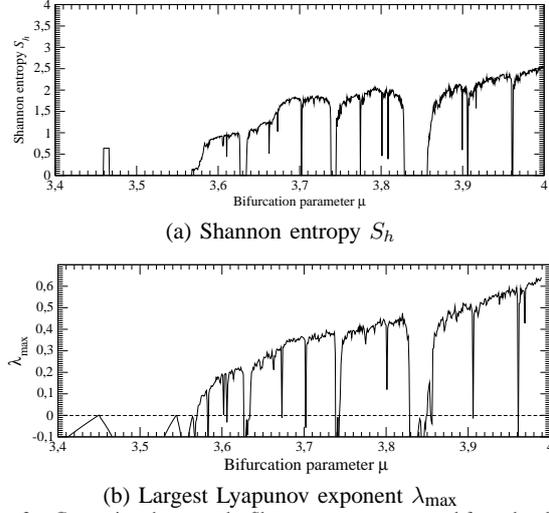


Fig. 2. Comparison between the Shannon entropy computed from the close return plots and the largest Lyapunov exponent for the Logistic map versus parameter μ . A close return plots of 3000×150 data points was used for computing the Shannon entropy for each μ -value.

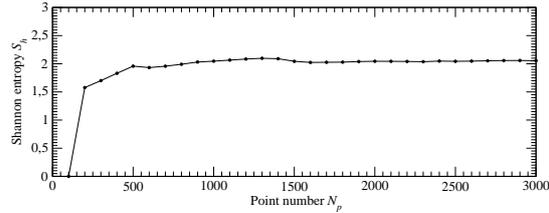


Fig. 3. Shannon entropy computed from the close return plots for the Logistic map versus the number of points. Parameter values $\mu = 3.89$.

III. PATIENT-VENTILATOR INTERACTIONS IN NONINVASIVE VENTILATION

Pressure support ventilation is a ventilatory mode where a preset inspiratory positive airway pressure (IPAP) and a preset expiratory positive airway pressure (EPAP) are superposed to spontaneous respiratory cycles. It unloads the respiratory muscles and, consequently, decreases the work of breathing in stable patients with obstructive diseases like chronic obstructive pulmonary disease (COPD) [12] or cystic fibrosis [13]. Pressure support ventilation is often qualified as a “physiological” ventilatory mode because it allows the patient to keep a control over his respiratory rate, inspiratory time and tidal volume. It is therefore not surprising that it has been found to be better tolerated than other ventilatory modes, especially when compared with volume-target ventilation [13], [14]. However, it requires an adequately titrated and performing ventilator in order to correctly superpose mechanical breaths to spontaneous respiratory efforts [15], that is, to optimize patient-ventilator synchronization [16].

Patient-ventilator asynchronies and especially ineffective inspiratory triggering efforts are regularly encountered when performing noninvasive mechanical ventilation. [17]. Ineffective inspiratory efforts under pressure support ventilation are

more frequent during sleep [18], [19] or when increasing the level of ventilatory assistance [20], [21]. Patient-ventilator asynchronisms including ineffective inspiratory efforts, may be clearly a cause of noninvasive ventilation intolerance and failure. Either in acute or chronic setting, the incidence of non triggered respiratory cycles and their consequences on noninvasive mechanical ventilation efficacy and comfort remains unknown.

During routine measurements of breathing pattern, respiratory flow (Q_v) was measured using a pneumotachograph connected to a pressure transducer. The pneumotachograph was inserted between the full face mask and the intentional leak. Airway pressure (P_{aw}) was measured with a differential pressure transducer near the pneumotachograph. Two typical excerpts of the data sets recorded during the protocol are shown in Fig. 4. They correspond to a healthy subject not very trained to noninvasive mechanical ventilation. When an antibacterial filter is inserted within the ventilatory circuit, this subject has some difficulties to trigger the ventilator and many inefficient inspiratory efforts lead to a high asynchrony event index (41%) (Fig. 4a). Basically, a non triggered cycle can be identified when a small amplitude oscillation of the airflow is associated with a low peak of pressure (about the EPAP value). In this case, it is sufficient to remove the antibacterial filter to greatly reduce the rate of non triggered cycles to 4.6% (Fig. 4b).

Twelve subjects (seven female and five male) with various health conditions were studied [5]. When applied in a Poincaré section close return plot analysis provides a more reliable characterization of the underlying dynamics [8]. Thus, the maximum of the airway pressure P_{max} during a respiratory cycle is used to build a “discrete” time series. The Shannon entropy is computed from these close return plots according to our new estimator. When computed from the maxima of the airway pressure, the Shannon entropy will be denoted as S_P . It is equal to 2.3 without the filter and equal to 0.4 with the filter. It has been found that the Shannon entropy S_P is strongly correlated to the rate of non triggered cycles. In a previous study [22], we found that an asynchrony event index less than 10% was not relevant for ventilatory comfort. Such a rate corresponds to a Shannon entropy slightly less than 1. Thus, a Shannon entropy S_P less than 1 corresponds to a situation where inefficient efforts are not clinically relevant on the subjects’ comfort.

Another dynamical characteristics relevant for the quality of the assisted mechanical ventilation is the rate of fluctuations of the total duration of the respiratory cycle. Since the subject is in a quiet seated position, the breathing rhythms should be regular. In particular, the patients very familiar with mechanical ventilation should be able to manage their ventilator for breathing in a regular way. On the other hand, we assume that more regular the dynamics is, better the comfort is. Moreover, the fluctuations over the total duration of the respiratory cycle, T_{tot} , are not necessarily correlated with the asynchrony index [5]. Recurrence plots are computed from T_{tot} using a threshold ϵ set to $\sqrt{d_E} \times 0.1 \times \bar{T}$ where \bar{T} is the “ideal” time duration cycle corresponding to a respiratory frequency equal to 12 breaths per min.

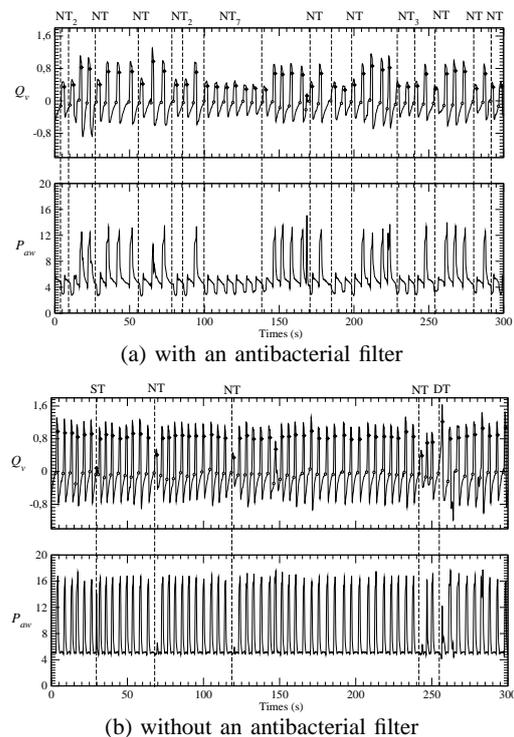


Fig. 4. Time series of the airflow and the pressure measured with a healthy subject not so well trained to noninvasive mechanical ventilation. Few types of asynchrony events were identified in these two examples. They are labelled as follows: NT = non-triggered cycle; ST= self-triggered cycle and DT= delayed triggered cycle. The beginnings of the respiratory cycle, corresponding to the inspiratory effort of the patient, are designated by circles and the end of the inspiratory phase (during which the ventilator applies a pressure support at the IPAP value) are marked by black diamonds on the airflow time series. Parameter setting: IPAP=16 mbar, EPAP=4 mbar and the inspiratory threshold is set to the most sensitive value, that is, 0.0167 l.s^{-1} .

The two Shannon entropies thus characterize two relevant dynamical properties for investigating the dynamics underlying patient-ventilator interactions. The Shannon entropy S_T was computed for the 69 data sets recorded and plotted versus the Shannon entropy S_P (fig. 5). These estimations of the Shannon entropies are computed over the whole 10 min data sets. Depending on the respiratory frequency, the number of data points is therefore between 91 and 337 as previously explained. Working at fixed number of data points is not natural in such a study. Typically, when more than one hundred of points are considered, estimations are not significantly dependent on the length of the data sets. Basically, four different regions are distinguished in this figure. First, the square defined by $S_T < 1$ and $S_P < 1$ corresponds to subjects who have fluctuations neither over P_{max} , nor over T_{tot} . There is no ambiguity for these subjects since they have almost no asynchrony event and their breathing rhythms are very regular.

Second, there is the rectangle defined by $S_T > 1$ and $S_P < 1$. These data sets correspond to subjects with less than 10% of non triggered cycles but quite significant fluctuations over the total duration of the ventilatory cycle. They corre-

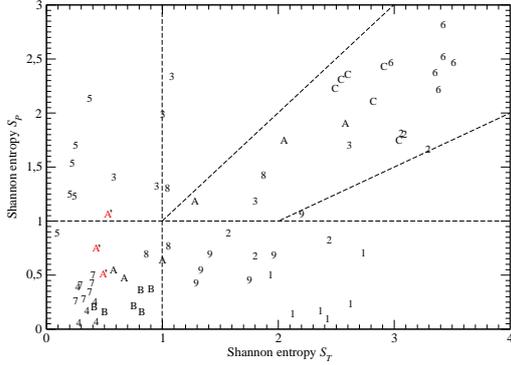


Fig. 5. Shannon entropy S_T versus Shannon entropy S_P for the 69 data sets recorded during our protocol. Integers i ($i \in [1; 9]$) designate subjects S_i for the six measurements at different IPAP values. Letters A, B and C designate subjects S_{10} , S_{11} and S_{12} , respectively.

spond to two subjects, S_1 and S_9 , not familiar with noninvasive mechanical ventilation. There is no OHS patient under these conditions. This could suggest that OHS patients would not display significant fluctuations over the total duration of the ventilatory cycle, T_{tot} . Indeed, obesity tends to reduce lung volumes and there is no longer possibility for varying the inspiratory volume and/or the respiratory time. Third, the rectangle defined by $S_T < 1$ and $S_P > 1$ corresponds to cases where there is many ineffective efforts although the breathing rhythm is regular. The fourth part of the graph shown in fig. 5 is associated with a sector such as $S_T > 1$ and $S_P > 1$. Most of the points located in this fourth sector correspond to subjects not familiar with a mechanical ventilation (only patient S_2 and S_6 are familiar). For all of the subjects being not familiar with mechanical ventilation there is an obvious correlation between S_P and S_T . For these patients, the fluctuations over the time duration T_{tot} results from asynchrony events.

IV. ESTIMATING SLEEP FRAGMENTATION

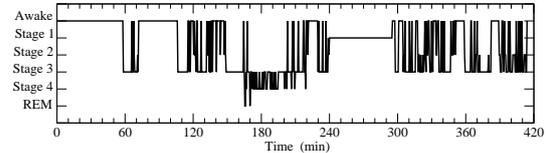
If patient-ventilator interactions are relevant for patient's comfort, it is still an open question to determine whether the quality of these interactions has an impact on the sleep quality. The main problem to address this question is to estimate the sleep quality from measurements. Usually, it is estimated using the Epworth sleepiness scale [23] that consists in a series of questions asked to the patients. In the other hand, a factor contributing to impaired daytime function and sleepiness is the sleep fragmentation [24], [25]. Many attempts have been made to estimate sleep quality from EEG. Among others, micro-arousals have been introduced to estimate sleep fragmentation [26] and number of arousals correlates significantly with nearly all sleep parameters in the direction of indicating decreased quantity and quality of sleep as arousals increases [27]. As any EEG scoring, identifying micro-arousals is a time-consuming method that requires a trained observer and manual edition. It therefore presents a high-inter-scorer variability [28], [29], [30]. As another measure of the sleep quality, a sleep fragmentation index has been introduced as a crude estimate of sleep disruption [31] presenting a good correlation

with micro-arousals. Unfortunately, EEG estimated arousals does not uniformly provide robust correlations with daytime sleepiness [32], although sometimes this index seems to be an accurate estimator of sleep fragmentation in patients with sleep disorders [33]. Consequently, the reliability of the sleep fragmentation index can be questioned.

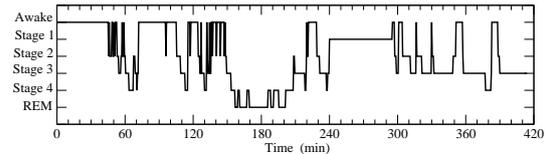
The sleep fragmentation index is defined as the total sleep stage shift plus the total number of awakening divided by the total sleep time [33]. Thus, this index does not take into account the time duration of each stage, a factor quite important since intervals of sleep must be 5-10 min to provide restoration, that is, to eliminate sleepiness [25], [34], [35]. In order to improve the sleep fragmentation index previously discussed, that is, to take into account the interval of sleep, we will use close return plots computed from the EEG scored using 30 s-windows into sleep stages (Fig. 6). The EEG is thus converted into a symbolic sequence using 6 symbols according to

$$\sigma_n = \begin{cases} 0 & \text{Awake} \\ 1 & \text{Stage 1} \\ 2 & \text{if EEG} \\ 3 & \text{corresponds to} \\ 4 & \text{Stage 2} \\ 5 & \text{Stage 3} \\ & \text{Stage 4} \\ & \text{REM sleep} \end{cases} \quad (5)$$

The EEG is thus transformed into a hypnogram as shown in Figs. 6. In this study each EEG was scored by two different scorers, namely A and B. Obvious departures are observed. They will be quantified by using close return plots.



(a) Scoring by neurologist A



(b) Scoring by neurologist B

Fig. 6. Hypnograms scored by two different neurologists from the same electroencephalograms.

Hypnograms, that is, symbolic sequences $\{\sigma_n\}_{n=1}^N$ were converted into a close return plots using

$$C_{ij} = \begin{cases} 1 & \text{if } \sigma_i = \sigma_{i-j} \\ 0 & \text{if } \sigma_i \neq \sigma_{i-j} \end{cases} \quad (6)$$

where $j \in [0, 50]$. The close return plots shown in Figs. 7a and 7b encode the recurrence properties of hypnograms scored by A and B, respectively. Black segments indicate intervals of a given sleep stage. More jittery the close return plot is, more fragmented the sleep is. The common part shared by these two close return plots is obtained by the product

$$C_{A \otimes B, ij} = 1 - C_{A, ij} \times C_{B, ij} \quad (7)$$

(Fig. 7c). White points thus correspond to the parts of EEG that were scored with the same type of sleep stage shift by both scorers. It thus provides the sleep structure identified by both scorers.

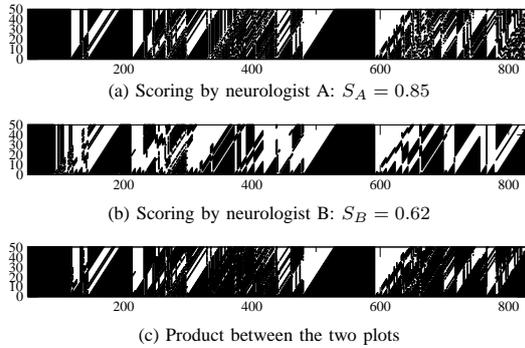


Fig. 7. Close return plots for two different sleep stages coded from the same electroencephalograms. Case where neurologist A found a sleep more fragmented than found by neurologist B.

From these close return plots, a Shannon entropy is computed to estimate the sleep fragmentation. The length of interval of sleep is naturally taken into account in a Shannon entropy computation. In the present case, the Shannon entropies resulting from the two scoring differs by around 25%. Here, scorer A estimated that the sleep was more fragmented than scorer B. In the case shown in Figs. 8, both scorers got the same entropy, although the two close return plots differ a little bit. In Fig. 9, contrary to what was observed in Figs. 7, scorer B estimated the sleep more fragmented than scorer A did. Shannon entropy S_A estimated from scoring A was thus less than the Shannon entropy S_B from scoring B. It has been observed that when $S_A < S_B$, this always corresponded to a situation where a quite long interval was scored in a rather fragmented way — with micro-arousals — by scorer B but uniformly scored in the awake state by scorer A. Such an example is shown in Fig. 9 for the interval $240 < t < 300$ min. The averaged recovering rate between the two close return plots is $(83 \pm 6.5)\%$.

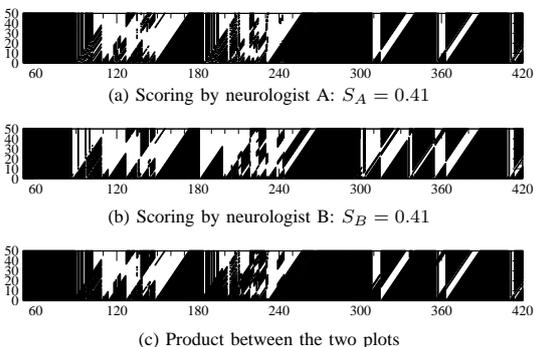


Fig. 8. Close return plots for two different sleep stages coded from the same electroencephalograms. Case where neurologist A found a sleep as fragmented as found by neurologist B.

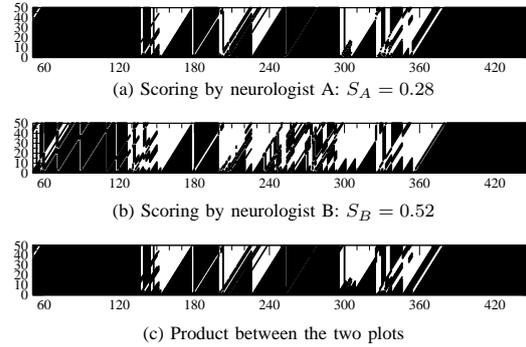


Fig. 9. Close return plots for two different sleep stages coded from the same electroencephalograms. Case where neurologist A found a sleep less fragmented than found by neurologist B.

Such a Shannon entropy was computed for each patient and plotted versus time duration of stages 1+2 (Fig. 10) and the number of micro-arousals (Fig. 11). In both cases, a significant correlation was obtained. These correlations are slightly better than those obtained with the sleep fragmentation index — defined as the total number of awakenings and shifts to stage 1 sleep divided by the total sleep time — used in [31]. The Shannon entropy is thus quite well correlated with two quantities that are known to quantify the sleep quality. The advantage it has compared to the usual sleep fragmentation index is that it takes into account how the stage switches are distributed over the night.

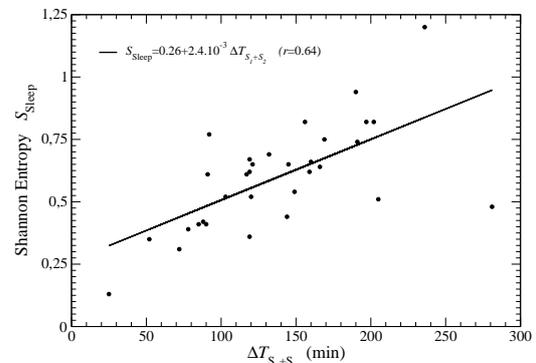


Fig. 10. Shannon entropy S_{Sleep} versus the time duration of stages 1 and 2.

V. CONCLUSION

Assessing the quality of patient-ventilator interactions is a major issue for the development of noninvasive mechanical ventilation. Indeed, determining such a quality contributes to understand why a ventilation is accepted or not by some patients. With two Shannon entropies estimated from close return plots built on maxima of pressure and time duration of respiratory cycles, it was possible to exhibit four classes of patient-ventilator interactions. One of the most promising feature is that one class corresponds to ideal interactions

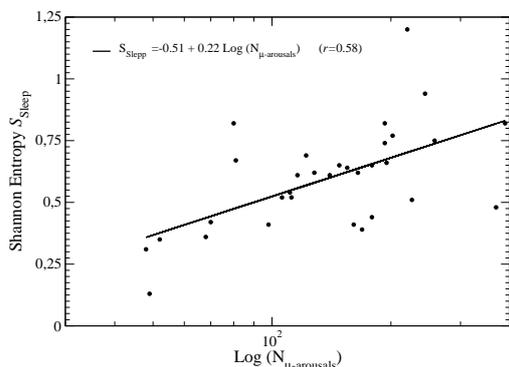


Fig. 11. Shannon entropy S_{Sleep} versus the number of micro-arousals.

and, consequently, constitutes a target to improve ventilator settings.

In a more prospective way, we introduced a novel manner to quantify sleep quality in computing Shannon entropy from a close return plot built from sleep stage scoring. It is promising because this Shannon entropy was correlated to some quantifiers that were commonly used to assess the sleep quality. Shannon entropies computed from close return plots are rather promising since they provide good measures of the complexity of the underlying dynamics. Moreover, their computations are quite robust against noise contamination.

REFERENCES

- [1] M. SMALL, J. JUDD, M. LOWE & S. STICK, Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep, *Journal of Applied Physiology*, **86**, 359-376, 1999.
- [2] C. AHLSTROM, A. JOHANSSON, P. HULT & P. ASK, Chaotic dynamics of respiratory sounds, *Chaos, Solitons & Fractals*, **29**, 1054-1062, 2006.
- [3] M. WYSOCKI, M.-N. FIAMMA, CH. STRAUS, C.-S. POON & T. SIMIŁOWSKI, Chaotic dynamics of resting ventilatory flow in humans assessed through noise titration, *Respiratory Physiology & Neurobiology*, **153** (1), 54-65, 2006.
- [4] N. MARWAN, N. WESSEL, U. MEYERFELDT, A. SCHIRDEWAN, & J. KURTHS, Recurrence-plot-based measures of complexity and their application to heart-rate-variability data, *Physical Review E*, **66**, 026702, 2002.
- [5] H. RABARIMANANTSOA, L. ACHOUR, C. LETELLIER, A. CUVELIER & J.-F. MUIR, Recurrence plots and Shannon entropy for a dynamical analysis of asynchronisms in mechanical non-invasive ventilation, *Chaos*, **013115**, 2007.
- [6] J.-P. ECKMANN, S. O. KAMPHORST & D. RUELLE, Recurrence plots of dynamical systems, *Europhysics Letters*, **4**, 973-977, 1987.
- [7] L. L. TRULLA, A. GIULIANI, J. P. ZBILUT & C. L. WEBBER JR., Recurrence quantification analysis of the logistic equation with transients, *Physica Letters A*, **223**, 255-260, 1996.
- [8] C. LETELLIER, Estimating the Shannon entropy: recurrence plots versus symbolic dynamics, *Physical Review Letters*, **96**, 254102, 2006.
- [9] G. B. MINDLIN & R. GILMORE, Topological analysis and synthesis of chaotic time series, *Physica D*, **58**, 229-242, 1992.
- [10] C. GILMORE, A new test for chaos, *Journal of Economic Behavior and Organization*, **22**, 209-237, 1993.
- [11] YA B. PESIN, *Dimension Theory in Dynamical Systems*, University of Chicago Press, 1998.
- [12] S. NAVA, N. AMBROSINO, F. RUBINI, C. FRACCHIA, C. RAMPULLA, G. TORRI & E. CALDERINI, Effect of nasal pressure support ventilation and external PEEP on diaphragmatic activity in patients with severe stable COPD, *Chest*, **103**, 143-150, 1993.
- [13] B. FAUROUX, J. PIGEOT, M. I. POLKEY, D. ISABEY, A. CLEMENT & F. LOFASO, In vivo physiologic comparison of two ventilators used for domiciliary ventilation in children with cystic fibrosis, *Critical Care Medicine*, **29**, 2097-2105, 2001.
- [14] M. VITACCA, F. RUBINI, K. FOGGIO, S. SCALVINI, S. NAVA & N. AMBROSINO, Non-invasive modalities of positive pressure ventilation improve the outcome of acute exacerbations in COLD patients, *Intensive Care Medicine*, **19**, 450-455, 1993.
- [15] L. BROCHARD, & F. LELLOUCHE, Pressure-support ventilation. In M. J. Tobin, editor. *Principles & practice of mechanical ventilation*, Second Edition ed. McGraw-Hill, Inc, New-York. 221-250, 2006.
- [16] J.-C. RICHARD, A. CARLUCCI, L. BRETON, N. LANGLAIS, S. JABER, S. MAGGIORE, S. FOUGERE, A. HARF & L. BROCHARD, Bench testing of pressure support ventilation with three different generations of ventilators, *Intensive Care Medicine*, **28**, 1049-1057, 2002.
- [17] S. NAVA, C. BRUSCHI, C. FRACCHIA, A. BRASCHI & F. RUBINI, Patient-ventilator interaction and inspiratory effort during pressure support ventilation in patients with different pathologies, *European Respiratory Journal*, **10**, 177-183, 1997.
- [18] S. PARTHASARATHY Effects of sleep on patient-ventilator interaction, *Respiratory Care Clinics of North America*, **11**, 295-305, 2005.
- [19] F. FANFULLA, M. DELMASTRO, A. BERARDINELLI, N. D. LUPO, & S. NAVA, Effects of different ventilator settings on sleep and inspiratory effort in patients with neuromuscular disease, *American Journal of Respiratory Critical Care Medicine*, **172**, 619-624, 2005.
- [20] P. LEUNG, A. JUBRAN & M. J. TOBIN, Comparison of assisted ventilator modes on triggering, patient effort, and dyspnea, *American Journal of Respiratory Critical Care Medicine*, **155**, 1940-1948, 1997.
- [21] E. GIANNOULI, K. WEBSTER, D. ROBERTS & M. YOUNES, Response of ventilator-dependent patients to different levels of pressure support and proportional assist, *American Journal of Respiratory Critical Care Medicine*, **159**, 1716-1725, 1999.
- [22] L. ACHOUR, C. LETELLIER, A. CUVELIER, E. VÉRIN & J.-F. MUIR, Asynchrony and cyclic variability in pressure support noninvasive ventilation, *Computer in Biology and Medicine*, **37**, 1308-1320, 2007.
- [23] M. W. JOHNS, A new method for measuring daytime sleepiness: the Epworth sleepiness scale, *Sleep*, **14**, 540-545, 1991.
- [24] M. A. CARSKADON, E. D. BROWN & W. C. DEMENT, Sleep fragmentation in the elderly: relationship to daytime sleep tendency, *Neurobiology Aging*, **3**, 321-327, 1982.
- [25] B. LEVINE, T. ROEHR, E. STEPANSKI, F. ZORICH & T. ROTH, Fragmenting sleep diminishes its recuperative value, *Sleep*, **10**, 590-599, 1987.
- [26] M. BONNET, D. CARLEY, M. CARSKADON, P. EASTON, C. GUILLEMINAULT, R. HARPER, B. HAYES, M. HIRSHKOWITZ, K. PERIKLIS, S. KEENAN, M. PRESSMAN, T. ROEHR, J. SMITH, J. WALSH, S. WEBER, & P. WESTBROOK, EEG arousals: scoring rules and examples: a preliminary report from the sleep disorders atlas task force of the American Sleep Disorders Association, *Sleep*, **15**, 173-184, 1992.
- [27] E. J. STEPANSKI, Improving the utility of interpreting sleep fragmentation, *Journal of Clinical Sleep Medicine*, **3** (3), 271-274, 2007.
- [28] J. S. LOREDO, J. L. CLAUSEN, S. ANCOLI-ISRAEL & J. E. DIMSDALE, Night-to-night arousal variability and inter-scoring reliability of arousal measurements, *Sleep*, **22**, 916-920, 1999.
- [29] C. W. WHITNEY, D. J. GOTTLIEB, S. REDLINE, R. G. NORMAN, R. R. DODGE, E. SHAHAR, S. SUROVEC, & F. J. NIETO, Reliability of scoring respiratory disturbance indices and sleep staging, *Sleep*, **21**, 749-757, 1998.
- [30] M. J. DRINNAN, A. MURRAY, C. J. GRIFFITHS & G. J. GIBSON, Inter-observer variability in recognizing arousal in respiratory sleep disorders, *American Journal of Respiratory Critical Care Medicine*, **158**, 358-362, 1998.
- [31] M. J. MORREL, L. FINN, H. KIM, P. E. PEPPARD, L. S. BARD & T. YOUNG, Sleep fragmentation, awake blood pressure, and sleep-disordered breathing in a population-based study, *American Journal of Respiratory Critical Care Medicine*, **158**, 358-362, 1998.
- [32] M. H. BONNET, K. DOGHRAMI, T. ROEHR, E. J. STEPANSKI, S. H. SHELDON, A. S. WALTERS, M. WISE, & A. L. CHESSON JR, The scoring of arousal in sleep: reliability, validity, and alternatives, *Journal of Clinical Sleep Medicine*, **2**, 175-180, 2006.
- [33] J. HABA-RUBIO, V. IBANEZ & E. SFORZA, An alternative measure of sleep fragmentation in clinical practice: the sleep fragmentation index, *Sleep Medicine*, **5**, 577-581, 2004.
- [34] R. DOWNEY & M. H. BONNET, Performance during frequent sleep disruption, *Sleep*, **10**, 354-363, 1987.
- [35] R. G. NORMAN, M. A. SCOTT, I. AYAPPA, J. A. WALSLIEBEN & D. M. RAPOPORT, Sleep continuity measured by survival curve analysis, *Sleep*, **29**, 1625-1631, 2006.



Christophe Letellier received B.S. degrees from the Physics Department (1991) at Paris VII and his Ph.D. from the Physics Department at Paris VII in 1994. He has been professor of physics at the University of Rouen since 1996. His research activities are devoted to the development of technique to characterize chaotic behaviours and to obtain global models from experimental data. He uses his know-how to investigate biomedical data, mainly from the cardio-respiratory system. He is the head of a national research group from CNRS on Dynamics and Control of Complex Systems (GdR 2984 DYCOEC) since 2004. He is responsible for a course in Biomedical Engineering since 2007.



Adriana E Portmann was received as a Medical Doctor in 1983, at the National University of Medicine of Rosario (Argentina). She continued the specialization in Neurology at the same place. She received in France the Inter-University Diploma of Sleep: "Wake/Sleep" and the Clinical Neurophysiology in 1993, the Physiology and Cardio-Respiratory Physiopathology of Sleep in 1994 and the M B degrees in Biological Sciences 1997, from the University of Rouen. She has been working at different laboratories of sleep at the Pneumological, Neurological, and Physiological Departments at the public hospitals in Paris, Montpellier and Rouen, respectively.



Dounia Bounoiare received a bachelors degree in 2002. Her first interest was for medical studies but she finally studied Physics. She was studying theoretical and applied Physics at the University of Rouen. Since 2007, she is under graduation in Fluid Mechanics, Solid Matter and Optics at the University of Rouen. During these years, she worked with Pr. Letellier to investigate Laser dynamics and biomedical data analysis.



Roomila Naeck received her Bachelor degree in 2002 in Mauritius. Then she obtained a master in Neurosciences at the University of Strasbourg (France). Since September 2008, she is spending a Ph'D under the supervision of Pr Christophe LETELLIER and Pr Antoine CUVELIER on sleep quality under non invasive ventilation , supported by ADIR Assistance.



Ubiratan S. Freitas received his Electrical Engineer degree in 2000 at the Federal University of Minas Gerais, Brazil, and his Master degree in Electrical Engineering in 2001 at the same university. In 2006, he received his PhD in Computer Science at Brazil's National Institute for Space Research. Since 2007 he has a Post-doc position at the University of Rouen, France.



Antoine Cuvelier received a Masters Degree in Respiratory Physiology from the Paris V University in 1991 and his PhD from the Rouen University in 2003. He is Professor of Pulmonary Medicine at Rouen University since 2008 and is specialized in the management of patients having chronic and/or acute-on-chronic respiratory failure. He has developed clinical works about noninvasive ventilation (NIV), including technical and monitoring issues of NIV in those patients requiring domiciliary mechanical ventilation.

Controlling the Interaction between Wild and Transgenic Mosquitoes

Marat Rafikov, *Member, IEEE*, Ana Paula P. Wyse, and Luiz Bevilacqua

Abstract— The development of transgenic mosquitoes, that are resistant to diseases, may provide a new and effective weapon of diseases control. Such approach relies on transgenic mosquitoes being able to survive and compete with wild-type populations. These transgenic mosquitoes carry a specific code that inhibits the plasmodium evolution in his organism. In actual paper, a nonlinear control strategy is proposed to indicate how the genetically modified mosquitoes should be introduced in the environment. The numerical simulations show the effectiveness of the proposed control.

Index Terms—Mathematical modeling, malaria, transgenic mosquitoes, optimal vector control.

I. INTRODUCTION

Vector borne diseases have affected many countries, mainly that very poor, but due to global warming, there is a real risk of these diseases appear in regions where they have already been eradicated or even in that where the environmental conditions would never have allowed its existence.

Some efforts have been made to control tropical diseases such as dengue fever, malaria and others. Even these diseases are not lethal at most of cases; the consequences of an epidemic are very serious.

Scientists are working in vaccines, new drugs, biological e chemical insecticide and other strategies to combat diseases and/or their intermediate hosts. Due to evolution of genetic studies, it has been possible to obtain genetically modified mosquitoes refractory to some diseases. These new insects should couple with wild mosquitoes and spread out the gene that intervenes in the transmission block. This research have advanced quickly and a chronological line of main results is shown in [1].

The development of transgenic mosquitoes, that are resistant to diseases, may provide a new and effective weapon of diseases control. Such approach relies on transgenic mosquitoes being

Manuscript received April 21, 2009. This work was supported in part by the Universidade Federal do ACB, Santo André, SP, Brazil. The authors thank the UFABC that provided financial support for this research.

M. Rafikov is with the Universidade Federal do ABC, Rua Santa Adelia, 166, Santo Andre, SP, Brazil, CEP 09210-170 (phone: +55112779-5409; e-mail: marat9119@yahoo.com.br).

A.P.P. Wyse is with the Universidade Federal do Amapa, Rod. Juscelino Kubitschek, KM-02, Macapá, AP, Brazil, CEP 68.902-280 (e-mail: anawyse@lncc.br).

L. Bevilacqua is with the Universidade Federal do ABC, Rua Santa Adelia, 166, Santo Andre, SP, Brazil, CEP 09210-170 (e-mail: bevi@lncc.br).

able to survive and compete with wild-type populations. These transgenic mosquitoes carry a specific code that inhibits the plasmodium evolution in his organism. It is said that this characteristics is hereditary and consequently the disease fades away after some time.

In the last years, the genetic modification of malaria vectors has been very prominent. The first *Anopheles* mosquitoes refractory to malaria were engineered in 2002 from a technique developed by Catteruccia et al [2]. Once transgenic mosquitoes are released, interactions between the two populations and inter-specific mating between the two types of mosquitoes take place.

The simple mathematical model for interacting wild and transgenic mosquito populations based on systems of difference equations was formulated in [3]. The generation overlapping and a variable environment were not considered in this model. In [4] a non-autonomous continuous-time mathematical model was presented. In that model the transgenic mosquitoes were considered to be in a single population and, as the wild population, its dynamics followed a seasonal pattern varying during the year. In [5] an optimal control problem was formulated and solved for this model, and the linear feedback control strategies indicated guidelines for the success of transgenic mosquitoes.

In actual paper, a nonlinear control strategy is proposed to indicate how the genetically modified mosquitoes should be introduced in the environment. The numerical simulations show the effectiveness of the proposed control.

II. MATHEMATICAL MODEL FOR INTERACTING WILD AND TRANSGENIC MOSQUITO POPULATIONS

The interactions between wild and transgenic mosquito populations consider that the following assumptions prevail:

- a) all transgenic mosquitoes, without distinguishing their zygosity, were considered as a single population group;
- b) the transgenic and wild mosquitoes have the same carrying capacity and fitness rate;
- c) in the absence of transgenic mosquitoes, population of wild mosquitoes is described by the solution of the model (1);
- d) in the absence of wild mosquitoes, population of transgenic mosquitoes is described by the solution of the model (1).

The mathematical model which considers the interactions between wild and transgenic mosquito populations in a variable environment has the following form [4]:

$$\begin{cases} \frac{dV}{dt} = \frac{r_{11}(t)V + r_{12}(t)T}{V+T} V \left(1 - \frac{V+T}{k(t)}\right) - \delta_1 V \\ \frac{dT}{dt} = \frac{r_{21}(t)V + r_{22}(t)T}{V(t)+T(t)} T \left(1 - \frac{V+T}{k(t)}\right) - \delta_1 T \end{cases} \quad (1)$$

where

V is the adult female mosquito density for time t ;

T is the transgenic adult female density for time t ;

$r_{ij}(t)$ is the difference between the recruitment rate of female mosquitoes through a mating with the same mosquito class into adult for time t and the density-dependent death rate;

$r_{ij}(t)$, $i \neq j$ is the difference between the recruitment rate of female mosquitoes through a mating with the another mosquito class into adult for time t and the density-dependent death rate; δ_1 are coefficient associated with death rate.

The practical achievement of the control program is the substitution of the wild mosquitoes by the transgenic ones. Introducing the control function $u(t)$ in the system (2) with the purpose of accomplishing this substitution we get:

$$\begin{cases} \frac{dV}{dt} = \frac{r_{11}(t)V + r_{12}(t)T}{V+T} V \left(1 - \frac{V+T}{k(t)}\right) - \delta_1 V \\ \frac{dT}{dt} = \frac{r_{21}(t)V + r_{22}(t)T}{V(t)+T(t)} T \left(1 - \frac{V+T}{k(t)}\right) - \delta_1 T + u(t) \end{cases} \quad (2)$$

Now, the control strategy has to be able to modify the population mixing – wild/transgenic mosquitoes – from an initial state where transgenic mosquitoes are absent into a final population where no wild mosquito is present. That is the control function $u(t)$ in the system (2) has to be able to carry the population density into the ideal final state given by:

$$\begin{cases} \tilde{V} = 0 \\ \frac{d\tilde{T}}{dt} = r_{22} \tilde{T} \left(1 - \frac{\tilde{T}}{k(t)}\right) - \delta_1 \tilde{T} \end{cases} \quad (3)$$

The control problem can be formulated as follow:

Determine the strategy associated to the introduction of transgenic species u which leads the non-linear system (4) with periodical coefficients from a given initial to a final state defined by:

$$x(t_f) = 0 \quad (4)$$

In [5] this optimal control problem was solved using the linear feedback control coupled to a nonlinear system.

In actual paper this problem is solved by using the nonlinear feedback control.

III. NONLINEAR OPTIMAL CONTROL DESIGN PROBLEM

Consider the general infinite-horizon, autonomous, nonlinear regulation (stabilization) problem where the system is full-state observable, in the state, and affine in the input, represented in the form

$$\dot{x} = f(x) + B(x)u \quad (5)$$

$$x(0) = x_0 \quad (6)$$

where $x \in R^n$ are state vector, $u \in R^m$ is input vector, and $t \in [0, \infty)$, with $C^1(R^n)$ functions $f: R^n \rightarrow R^n$ and $B: R^n \rightarrow R^n$, and $B(x) \neq 0$ for all x . The origin $x=0$ is assumed to be an equilibrium point, such that $f(0) = 0$. The infinite-time horizon nonlinear regulation problem is defined with the following (nonquadratic in x but quadratic in u) performance index:

$$V(x) = \min_{u(t)} \int_0^{\infty} [x^T Q(x)x + u^T R(x)u] dt, \quad (7)$$

where the state and input weighting matrices are assumed state-dependent such that $Q: R^n \rightarrow R^{n \times n}$ is positive semidefinite and $R: R^n \rightarrow R^{m \times m}$ is positive definite for all x . It is assumed that functions $f(x)$, $B(x)$, $Q(x)$ and $R(x)$ are sufficiently smooth so that value function defined by (3) is continuously differentiable. It can be shown that the optimal feedback controller for system (1) can be constructed as [7]

$$u = -\frac{1}{2} R^{-1}(x) B^T(x) \frac{\partial V}{\partial x}$$

where the $\frac{\partial V}{\partial x}$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation [7]

$$\frac{\partial V^T}{\partial x}(x) f(x) - \frac{1}{4} \frac{\partial V^T}{\partial x}(x) B(x) R^{-1}(x) B^T(x) \frac{\partial V}{\partial x} + x^T Q(x)x = 0 \quad (8)$$

The HJB equation provides the solution to the optimal control problem for general nonlinear system (5); however, in most cases, it is impossible to solve it analytically. This has led to many methods being proposed in the literature for ways to approximately obtain the solution to the HJB equation as well as obtain a suboptimal feedback control for general nonlinear dynamical models.

According Mracek and Cloutier [11] the SDRE approach for obtaining a suboptimal solution of problem (5)-(7) is:

1) use direct parameterization (factorization) to bring the nonlinear dynamics to the state-dependent coefficient (SDC) form

$$\dot{x} = A(x)x + B(x)u \quad (9)$$

where

$$f(x) = A(x)x ; \quad (10)$$

2) solve the state-dependent Riccati equation

$$P(x)A(x) + A^T(x)P(x) - P(x)B(x)R^{-1}(x)B^T(x)P(x) + Q(x) = 0 \quad (11)$$

to obtain matrix-valued function $P(x)$ which is positive definite for all x ;

construct the nonlinear feedback controller

$$u = -R^{-1}B^T P(x)x \quad (12)$$

Under the assumption $f(0) = 0$ and $f \in C^1(R^n)$, a continuous nonlinear matrix-valued function $A(x)$ always exists such that (5) is satisfied. It is obvious, that $A: R^n \rightarrow R^{n \times n}$ is found by mathematical factorization and is nonunique for $n > 1$.

IV. SDRE SOLUTION OF THE WILD POPULATION CONTROL

Introducing error variables:

$$x_1 = V - \tilde{V} = V$$

$$x_2 = T - \tilde{T}$$

the following error system is obtained:

$$\dot{x} = A(x)x + Bu \quad (13)$$

with

$$A(x) = \begin{bmatrix} a_{11} & a_{12} \\ -a_{21} & a_{22} \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad (14)$$

where

$$a_{11} = \frac{r_{11}x_1 + r_{12}\tilde{T}}{x_1 + x_2 + \tilde{T}} \left(1 - \frac{x_1 + x_2 + \tilde{T}}{k}\right) - \delta_1,$$

$$a_{12} = \frac{r_{12}x_1}{x_1 + x_2 + \tilde{T}} \left(1 - \frac{x_1 + x_2 + \tilde{T}}{k}\right)$$

$$a_{21} = \frac{r_{22}(x_2 + \tilde{T})}{x_1 + x_2 + \tilde{T}} + \frac{r_{21}}{k}(x_2 + \tilde{T}),$$

$$a_{22} = r_{22} - \frac{r_{22}(x_2 + 2\tilde{T})}{k} - \delta_1.$$

The SDRE control is obtained in nonlinear feedback form:

$$u = -p_{12}x_1 - p_{22}x_2 \quad (15)$$

where p_{12} and p_{22} are elements of the positive definite matrix $P(x)$ which is solution of (11).

We illustrate the application of the strategy which is proposed to indicate how the genetically modified mosquitoes should be introduced in the environment.

For numerical simulations of interactions between wild and transgenic mosquito populations were used the following values of model coefficients:

$$r_{11} = 2.5684, \quad r_{12} = 0.5979, \quad r_{21} = 2.5684, \quad r_{22} = 2.5684.$$

These values were obtained in [4] based on data from [1].

Choosing

$$Q = \begin{bmatrix} 0.5 & 0 \\ 0 & 50 \end{bmatrix} \text{ and } R = [1],$$

we obtain $P(x)$ from the solution of the Riccati equation (11). The evolution of the wild and transgenic mosquito population without control for initial conditions $V(0)=10$ and $T(0)=0$ is shown in Fig. 1. The dynamics of the controlled error system and control function are presented in Fig. 2 and 3, respectively. The evolution of the wild and transgenic mosquito populations subjected to the control function u are shown in Fig. 4.

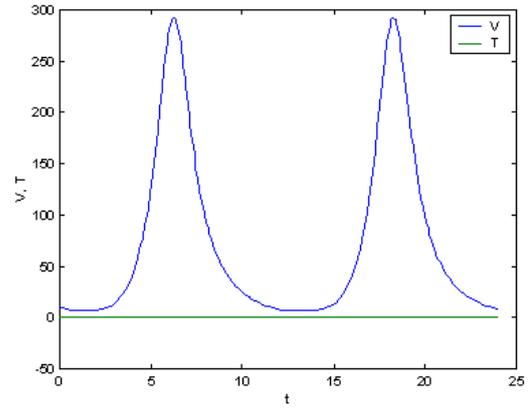


Fig. 1. Evolution of the wild and transgenic mosquito population without control

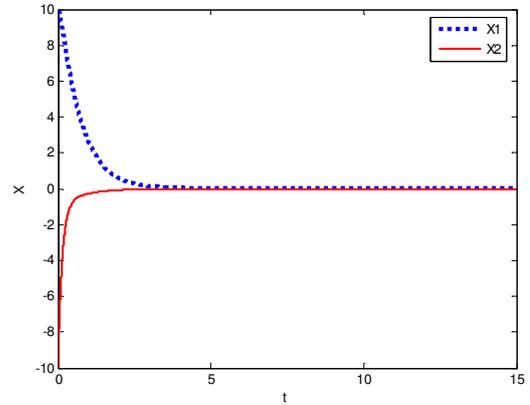


Fig. 2. Dynamics of the controlled error system

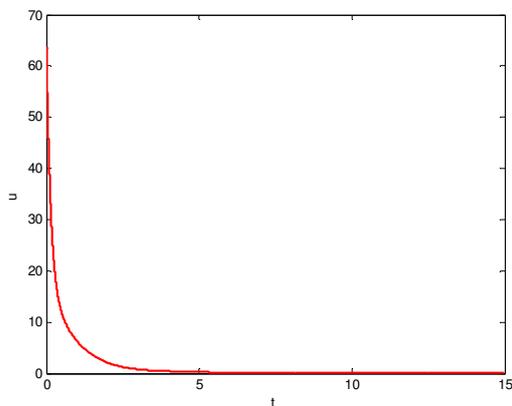


Fig. 3. Dynamics of the control function u

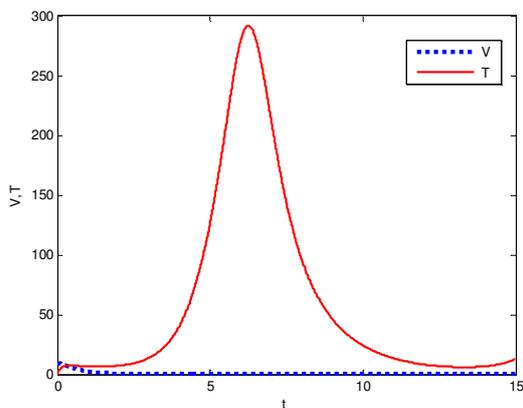


Fig. 4. Evolution of the wild and transgenic mosquito populations with control

V. CONCLUSION

The continuous variable model proposed to describe an interaction between wild and transgenic mosquitoes, does not consider the zigosity of the transgenic mosquito. We decided to start with a simpler model that is however supported by very strong evidence that the transgenic mosquito's gene is dominant. This fact makes the model meaningful to evaluate the performance of a control variable that could optimize the cost/benefit relation. The relevant question is: what is the best strategy to introduce transgenic mosquitoes in the environment in order to optimize costs.

The answer is to formulate and solve a vector control optimal control problem indicating how the genetically modified mosquitoes should be introduced in the environment.

It was shown that a nonlinear feedback control could steer the solution of the nonlinear system towards a desirable trajectory. The numerical simulations show the effectiveness of the proposed control system.

Finally one must be aware of the fact that the introduction of transgenic species in the environment is still very questionable. Biologists and ecologists play a critical role in answering the questions about secure introduction of transgenic mosquitoes. If there is sufficiently confidence that there is no harm by introducing this procedure then the proposed control can be considered as an economic and effective solution of the malaria vector control.

REFERENCES

- [1] L. A. Moreira, A. K. Ghosh, E. G. Abraham, M. Jacobs-Lorena "Genetic transformation of mosquitoes: a quest for malaria control," *International Journal for Parasitology* 32 (2002) 1599–1605.
- [2] F. Catteruccia, T. Nolan, T.G. Loukeris, C. Blass, C. Savakis, F.C. Kafatos and A. Crisanti, "Stable germline transformation of the malaria mosquito *Anopheles stephensi*," *Nature* 405 (2000) 959–962.
- [3] J. Li, "Simple mathematical models for interacting wild and transgenic mosquito populations," *Mathematical Biosciences* 189 (2004) 39-59.
- [4] A.P.P. Wyse, "*Optimal control for malaria vector for a seasonal mathematical model*," PhD Thesis, National Laboratory for Scientific Computing, Petropolis, RJ, Brazil, April 2007.
- [5] M. Rafikov, L. Bevilacqua and A.P.P. Wyse, "Optimal control strategy of malaria vector using genetically modified mosquitoes(Periodical style—Accepted for publication)," *Journal of Theoretical Biology*, to be published, doi: 10.1016/j.jtbi.2008.08.006
- [6] A.P.P. Wyse, L. Bevilacqua, M. Rafikov. Simulating malaria model for different treatment intensities in a variable environment, *Ecological Modelling* 206 (2007) 322-330.
- [7] B.D.O. Anderson and J.B. Moor. Optimal control: linear quadratic methods. NJ: Prentice-Hall, 1990.

TBA
Marc LEFRANC

Analysis of the sequential organisation of mice
trajectories
Philippe FAURE , Annick LESNE

Complexity in Soft and Condensed matter

Organizing Committee

- Arnaud PRIGENT
LOMC, University of Le Havre, France
arnaud.prigent@univ-lehavre.fr
- Daniel BONAMY
LNOSC, DSM/IRAMIS/SPCSI, CEA Saclay, Gif-sur-Yvette, France
daniel.bonamy@cea.fr
- François DAVIAUD
Groupe Instabilités & Turbulence, DSM/IRAMIS/SPEC, CEA Saclay, Gif-sur-Yvette, France
francois.daviaud@cea.fr
- Innocent MUTABAZI
LOMC, University of Le Havre, France
innocent.mutabazi@univ-lehavre.fr

Description

Complexity can be observed in many aspects of Soft and Condensed matter physics. Examples include turbulent flows in liquid, glassy dynamics, granular matter, complex fluids, damage and fracture of disordered materials, earthquakes, domain growth in ferroelectric or magnetic materials, superconductivity, dislocations in crystals, surface and interface dynamics in deposition problems, among other realisations. Despite their diversity, these systems share similar features: the emergence of generic collective behaviours from the interaction between the elementary constituents that cannot be understood or predicted simply by some averaging over the behaviour of individual components.

Statistical physics represents the privileged tool of physicists to study these systems. However understanding this complex matter necessitates new theoretical developments since the systems under consideration are generally far from equilibrium and often involve various relaxation processes, sometimes dissipative and irreversible, ranging over several time and length scales.

This special session of the ICCSA is intended (i) to provide some examples of physical complex systems illustrating the diversity of this theme in Soft and condensed matter physics, (ii) to present the various recent theoretical progresses to describe them and (iii) to draw possible orientation for future research in this field. It will be the occasion to bring together physicists from various domains, both experimentalists and theoreticians working on complex systems as well as researchers from other fields interested by physical applications or concepts.

Invited speaker

Hugues Chaté (Groupe Théorie des Systèmes Complexes, DSM/IRAMIS/SPEC, CEA Saclay)

Contents

Statistical analysis of nanoscale real space images for surface, interface and thin film studies and applications T. NGUYEN , O. BEZENCENET , A. BATAILLE , Daniel BONAMY , A. BARBIER , J. COUSTY , Luc BARBIER	29
Crackling noise in cracks Daniel BONAMY , S. SANTUCCI , L. PONSON	30
Aging and effective temperatures near a critical point Sylvain JOUBAUD , Artyom PETROSSYAN , Sergio CILIBERTO	31
Effective rheology and transport properties of an active suspension of E-Coli Eric CLEMENT , J. DAUCHET , G. MINO , T. DARNIGE , M. HOYOS , A. ROUSSELET	32
Dynamics and thermodynamics of the thermohaline circulation in a two-dimensional ocean model: theory Gilles COLLETTE , B. DUBRULLE , D. PAILLARD	33
Crack propagation and pinning in heterogeneous materials: effects of microscopic disorder and toughness pattern Davy DALMAS , E. BARTHEL , D. VANDEMBROUCQ	34
From cage jumps to dynamical heterogeneities in granular media Olivier DAUCHOT , R. CANDELIER	35
New approach for treating of spatio-temporal diagram: conditional averaging of two-dimensional fields Alexander EZERSKI , N. ABCHA , Innocent MUTABAZI	36
How herds move: a study of polar self-propelled particles É. BERTIN , M. DROZ , Guillaume GRÉGOIRE	37
Modulated spiral in a Couette-Taylor system submitted to a high radial temperature gradient Raphaël GUILLERM , Arnaud PRIGENT , Innocent MUTABAZI	38
Experimental investigation of the dynamical correlations between magnetic nanoparticles in isotropic and textured superspin-glasses K. KATSUYOSHI , Denis L'HÔTE , S. NAKAMAE , Y. TAHRI , Caroline THIBIERGE , E. VINCENT , V. MOSSER , A. KERLAIN , M. KONCZYKOWSKI , E. DUBOIS , V. DUPUIS , R. PERZYNSKI	39
Finite-size effects on sub-critical bifurcations in high-dimensional lattices of coupled logistic maps Paul MANNEVILLE	40
Oscillations in bidisperse fluidized suspensions A. DEBOEUF , G. GAUTHIER , Jérôme MARTIN , Dominique SALIN	41
Intermittent regime in flow oscillations investigated by means of symbolic dynamics Luc PASTUR , François LUSSEYRAN , Thierry FAURE , Christophe LETELLIER	42
Intermittency in granular flow Marc RABAUD , P. GONDERT , R. FISCHER	43
Evolution of the fabric tensor in amorphous silica: via molecular dynamics simulations Cindy ROUNTREE , M. TALAMALI , D. VANDEMBROUCQ , S. ROUX , E. BOUCHAUD	44
Lock exchange with autocatalytic reaction front I. BOU MALHAM , N. JARRIGE , Jérôme MARTIN , N. RAKOTOMALALA , L. TALON , Dominique SALIN	45
Dynamic fracture: how brittle are brittle amorphous solids? Julien SCHEIBERT , C. GUERRA , F. CÉLARIÉ , Davy DALMAS , Daniel BONAMY	46
Dynamical length scale in glass transition: experimental study Caroline THIBIERGE , Denis L'HÔTE , F. LADIEU , C. BRUN	47

Trajectories of sand grains in a wave boundary layer	
Tien DAT CHU , A. JARNO-DRUAUX , Alexander EZERSKI , F. MARIN	48
Elastic instability of polymer solutions in the Couette-Taylor system	
Fayçal KELAI , Olivier CRUMEYROLLE , Innocent MUTABAZI	49
The renormalized fragmentation equation and its exact solutions	
M.A. GOROKHOVSKI , V.L. SAVELIEV	50

Statistical analysis of nanoscale real space images for surface, interface and thin film studies and applications

T.T.T. Nguyen, O. Bezencenet, A.M. Bataille, D. Bonamy, A. Barbier, J. Cousty and L. Barbier

¹CEA, IRAMIS, SPCSI, F 91 191 Gif sur Yvette, France,

Controlling and tuning the morphology and texture of surfaces, interfaces and thin films at the nanoscale is of primary importance in term of technological applications (electronics, spintronics, chemical reactivity, photovoltaics, photonics...). Since the introduction of the Scanning Tunnelling Microscope (STM), progress in imagery at the nanoscale, by various techniques, is constant. Real space pictures give us a direct view of the topography, chemical composition, magnetisation... among others. At the nanoscale, the competition between thermal fluctuations and smoothening processes leads to complex morphologies that must be deciphered. In this respect, tools of "out of equilibrium" statistical physics were shown to be relevant. Morphological evolution of vicinal surfaces at thermal equilibrium and upon growth / smoothening, together with the investigation of ferromagnetic ($\text{Fe}_3\text{O}_4/\gamma\text{-Al}_2\text{O}_3$) and antiferromagnetic ($\alpha\text{-Fe}_2\text{O}_3/\text{Co}$) thin films will illustrate the potentialities of such approaches in the field of nanotechnologies.

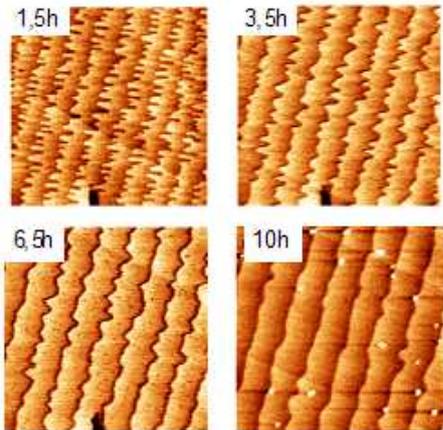


Fig.1 – AFM images of the topography of a vicinal surface of sapphire after various time of annealing at 1273K.

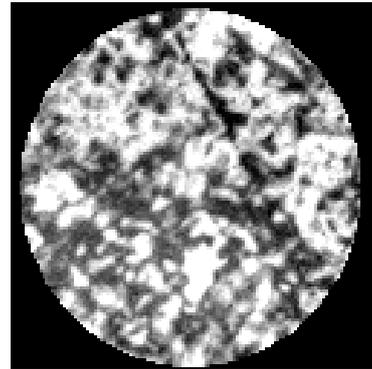


Fig.2 – 20 μm field of view X-PEEM images of an $\alpha\text{-Fe}_2\text{O}_3$ 20 nm thin film. The contrast arises from antiferromagnetic domains.

Crackling noise in cracks

D. Bonamy¹, S. Santucci^{2,3}, L. Ponson^{1,4}

¹CEA, IRAMIS, SPCSI, Grp. Complex Systems & Fracture, F-91191 Gif sur Yvette, France

²Fysisk Institutt, Universitetet i Oslo, P.O. Boks 1048 Blindern, N-0316 Oslo 3 Norway

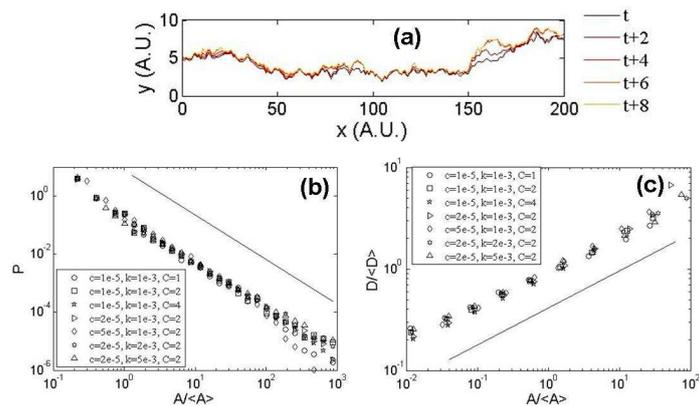
³Ecole Normale Supérieure de Lyon, Laboratoire de Physique, C.N.R.S. UMR5672, 46, Allée d'Italie, 69364 Lyon

⁴Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA.

Over the past century, the understanding of failure in brittle amorphous materials has been greatly improved. There exists now a coherent theoretical framework, the Linear Elastic Fracture Mechanics (LEFM) which allows describing precisely crack propagation in homogeneous brittle materials. On the other hand, the effect of materials heterogeneities onto their failure properties remains far from being understood. In particular, in heterogeneous materials under slow external loading, cracks growth often displays a jerky dynamics, with sudden jumps spanning over a broad range of length-scales, as also suggested from the acoustic emission accompanying the failure of various materials and - at much larger scale - the seismic activity associated to earthquakes. Presently, this intermittent “crackling” dynamics cannot be captured by standard Linear Elastic Fracture Mechanics (LEFM).

In this presentation, we will see how to extend LEFM to derive a stochastic description of quasi-static crack growth in disordered media [1]. Its predictions will then be confronted to experimental observations performed at University of Oslo [2] on the crack propagation within a transparent Plexiglas block. All the statistical features seem perfectly reproduced. This description suggests that the material failure is analogue to a critical transition between a stable phase where the crack remains pinned by the material heterogeneities and a moving phase where the mechanical energy available at the crack tip is sufficient to make the front propagate. While growing, the crack decreases its mechanical energy and gets pinned again. This mechanism exhibits universal – and to some extent predictable – statistical features, insensitive to the materials nature and the loading conditions.

Fig.1 – (a) In heterogeneous media, cracks progress through successive pinning/depinning phases, so-called avalanches. The distribution of avalanche area (b) and the scaling between the avalanche duration and area (c) form universal power-law independent of both the materials nature and loading conditions.



References

- [1] D. Bonamy, S. Santucci and L. Ponson, *Physical Review Letters* **101**, 045501 (2008).
 [2] K.-J. Maloy, S. Santucci, J. Schmittbuhl and R. Toussaint, *Physical Review Letters* **96**, 045501 (2006).

Aging and effective temperatures near a critical point

Joubaud Sylvain¹ Petrossyan Artyom² and Ciliberto Sergio²

¹*Physique of Fluids - University of Twente P.O. Box 217, 7522AE Enschede
The Netherlands,*

²*Ecole Normale Supérieure de Lyon, Laboratoire de Physique, C.N.R.S. UMR5672,
46, Allée d'Italie, 69364 Lyon*

The mean orientations of the fluctuations of the director of a nematic liquid crystal are measured using a sensitive polarization interferometer. When an electric field is applied perpendicularly to the initial alignment of the molecules, there is a critical point for which molecules try to align to the field. This is called the Fréedericksz transition which is expected to be second order phase transition. We report experimental evidence that, because of the critical slowing down, the LC presents, after a quench near the critical point, several properties of an aging system, such as power law scaling versus time of correlation and response functions. During this slow relaxation, a well defined effective temperature, much larger than the heat bath temperature, can be measured using the fluctuation dissipation relation. The results are in excellent agreement with the previous theoretical prediction for the effective temperature.

References

- [1] S. Joubaud, B. Percier, A. Petrosyan, and S. Ciliberto, *Phys.Rev. Lett.* 102, 130601 (2009)

Effective rheology and transport properties of an active suspension of E-Coli

E.Clement, J.Dauchet, G. Mino, T.Darnige, M.Hoyos, A.Rousselet

PMMH, ESPCI, CNRS-UMR 7636, 10, rue Vauquelin 75005 Paris, France

Contrasting with the rheology of standard dispersions of passive particles, bacterial fluids form an active system essentially due to the self-propelling properties of each swimming entity (the bacterium). To a large extent, the macroscopic rheological properties depend non only on the supplementary transfer of momentum due to the swimming activity but also, on the collective organization of the bacteria that can react to chemical signaling, long range hydrodynamic interaction and other complex properties due to the their biological nature.

Here we present experimental measurement of diffusion properties of a passive tracer (latex sphere) among a population of bacteria (E-Coli) that are concentrated close to a boundary. We change systematically the activity properties of the bacterium by changing the fraction of nutritive material in the fluid, the pH of the solution that directly act on the proton-motor rotation (bacterium velocity) and also the concentration of active bacteria in the surrounding of the tracer particles. We establish a relation between the effective diffusivity that can be much higher than the Brownian thermal value, as a function of active bacteria concentration and self propelling characteristics.

Dynamics and Thermodynamics of the thermohaline circulation in a two-dimensional ocean model: Theory

G. Collette¹, B. Dubrulle², D. Paillard¹

¹ CEA, LSCE, F 91 191 Gif sur Yvette, France,

²CEA, IRAMIS, SPEC, Grp. Instabilité et Turbulence, F 91 191 Gif sur Yvette, France,

We consider a simplified 2D ocean model. We use the thermohaline equations, under the classical Boussinesq approximation [1]. We show that the thermohaline equations for inviscid flows admit an infinite number of steady state solutions. We describe the statistical mechanics of such a system, and derive equilibrium states [2, 3]. We show that the stable steady state minimize a functional F while conserving an infinite number of conserved quantities. For a finite number of conserved quantities, simplest analytical solutions are calculated. These solutions describe well the thermohaline circulations. We furthermore derive relaxation equations which can be used as numerical algorithm to construct nonlinearly dynamically stable stationary solutions of thermohaline circulation.

References

- [1] O. Thual, J.C. McWilliams *Geophys. Astrophys. Fluid Dynamics* **64** (1991)
- [2] R. Robert, J. Sommeria, *J.Fluid Mech.* **229** (1991)
- [3] R. Robert, J. Sommeria, *Phys. Rev. Lett.* **69** (1992) 2776

Crack propagation and pinning in heterogeneous materials: Effects of microscopic disorder and toughness pattern.

D Dalmas¹, E. Barthel¹ and D. Vandembroucq²

¹Laboratoire SVI, UMR 125 CNRS/Saint-Gobain, Aubervilliers, France,

²Laboratoire PMMH, UMR 7636 CNRS/ESPCI/UPMC/Paris 7, Paris, France

Since the early eighties, fracture and damage of heterogeneous materials have motivated a large number of studies inspired by statistical physics [1]. In those disordered materials, the toughness varies from place to place which complicates the prediction and understanding of crack propagation. Due to randomness, the crack front can be distorted both in the out of plane and in-plane directions (cf. Fig.1). At the same time, this distortion of the crack front is responsible for elastic forces which tend to make it smooth. This competition between elasticity and disorder leads to a rich phenomenology. In particular a critical threshold appears for the stress intensity factor. Below threshold, the front propagate over a finite distance; above threshold, the crack front propagates freely at a highly fluctuating velocity [2,3]. Although promising, this depinning scenario appeared to give only a qualitative agreement with experimental results. In particular, the predicted roughness exponents are far from the experimental estimates. In the following we will try to discuss the validity of these models in brittle heterogeneous materials.

First we will precise the type of disorder to be considered in the context of fracture and show an experimental illustration which seems to validate this approach [4]. These results which contrasts with the previous experimental knowledge could be obtained via a fine control on the nature and scale of heterogeneities in a model material: phase separated glasses.

Second, an experimental study of purely interfacial crack propagation in the presence of very simple, macroscopic and fully described heterogeneous toughness field will be presented [5]. We will try to capture the individual contribution of each components of the toughness field by analyzing the crack front morphology (cf. Fig.2). We will show that designing specific heterogeneous interface allows modifying and controlling the crack front morphology. We will show that it is possible to describe quantitatively the crack front morphology in the framework of the pinning theory.

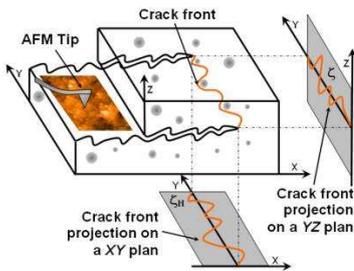


Fig.1 – Sketch of a propagating crack front pinned by heterogeneities.

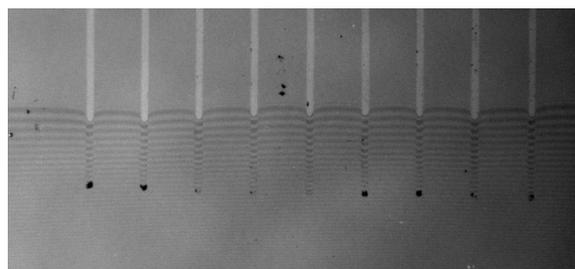


Fig.2 – Crack front morphology in presence of a patterned interface.

References:

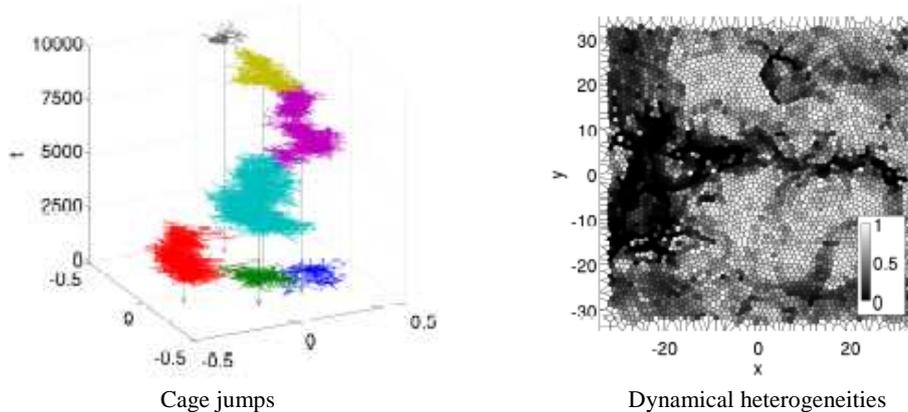
- [1] M.J. Alava, P.K.V.V Nukala and S. Zapperi, *Adv. Phys.* 55 (2006) 349--476
- [2] J. Schmittbuhl, S. Roux, J.P. Vilotte, and K.J. Maloy, *Phys. Rev. Lett.* 74 (1995) 1787-1790
- [3] S. Ramanathan, D.Ertas, and D.S. Fisher, *Phys. Rev. Lett.* 79 (1997) 873-876
- [4] D. Dalmas, A. Lelarge, and D. Vandembroucq, *Phys. Rev. Lett.* 101, 255501 (2008).
- [5] D. Dalmas, E. Barthel et D. Vandembroucq, *J. Mech. Phys. Solids* 57, 446 (2009).

From cage jumps to dynamical heterogeneities in granular media

O. Dauchot¹, R. Candelier¹

¹*Groupe Instabilities and Turbulence, DSM/IRAMIS/SPEC, CEA Saclay,
91 191 Gif sur Yvette, France,*

Unveiling the connection between the short term relaxation and the long term dynamical heterogeneities as observed near the glass transition in super-cooled liquids and the jamming transition in granular materials remains one of the major challenge in the physics of glassy systems.



On one hand, KCM models relate dynamical heterogeneities to a non trivial structure in the trajectory space, inherited from the local dynamical rules. [1]. On the other hand, recent studies of hard spheres close to isostaticity suggest that the collective aspect of the relaxation would stem from the extended character of the softest degrees of freedom [2].

There is still no direct experimental evidence in favour of one or the other mechanism in super-cooled liquids and whether they apply to dense granular media remains pure speculation. Also, in both cases one misses a clear picture of the microscopic mechanisms underlying the formation of such dynamical heterogeneities. Here we will show that for a dense granular layer under cyclic shear [4, 5] dynamical heterogeneities result from a two timescales process [6]. Short time *but already collective* events consist in clustered cage jumps, which concentrate most of the non affine displacements. Such clusters aggregate both temporally and spatially within an avalanche process, which ultimately builds the large scales dynamical heterogeneities.

References

- [1] Juan P. Garrahan and David Chandler, *Geometrical Explanation and Scaling of Dynamical Heterogeneities in Glass Forming Systems* Phys. Rev. Lett. 89, 035704 (2002)
- [2] Carolina Brito and Matthieu Wyart. *Heterogeneous dynamics, marginal stability and soft modes in hard sphere*, J. Stat. Mech. L08003 (2007)
- [4] G. Marty and O. Dauchot. *Subdiffusion and Cage Effect in a Sheared Granular Material*. Phys. Rev. Lett., 94(1):15701, (2005).
- [5] O. Dauchot, G. Marty, and G. Biroli. *Dynamical Heterogeneity Close to the Jamming Transition in a Sheared Granular Material*. Phys. Rev. Lett., 95 (26):265701, (2005).
- [6] R. Candelier, O. Dauchot, and G. Biroli, *Building Blocks of Dynamical Heterogeneities in Dense Granular Media*, Phys Rev Lett., 102 088001 (2009)

New approach for treating of spatio-temporal diagram: conditional averaging of two- dimensional fields

A.B. Ezersky¹, N. Abcha², and I. Mutabazi²

¹ M2C, UMR 6143 CNRS, Université de Caen, 2-4 rue des Tilleuls, 14000 Caen, France

² LOMC, FRE 3102 CNRS, Université du Havre, 53 rue de Prony, 76058 Le Havre, France

Space-time diagrams are widely used for investigation of spatio-temporal dynamics of nonlinear systems. To obtain a space-time diagram, values of a physical parameter I are measured in the points with coordinates (x_1, x_2, x_3, \dots) along a line in different moments of time (t_1, t_2, t_3, \dots) . In the obtained two dimensional field $I(x_i, t_k)$, it is possible to distinguish different structures (so called *building blocks*): sources and sinks, topological defects, holes Nozaki-Bekki. Such “blocks” play an important role in the pattern formation and spatio – temporal dynamics of extended nonlinear systems. In physical experiments identification of such blocks and measurements of their characteristics are not a trivial task. One of the encountered problems here is connected with influence of noise. In this report we propose the conditional averaging of spatio-temporal diagram to diminish the influence of noise distorting the properties of *building blocks*.

We have applied conditional averaging for investigation of defect appearing on the background of spiral pattern in Couette - Taylor flow between two rotating cylinders. We investigated structure of the field $I(x_i, t_k)$ in the neighborhood of a single defect. For conditional averaging we proceeded as follows. We used several spatio-temporal diagrams. In each diagram we determined position of the defect (x_{d_i}, t_{d_i}) as a point where amplitude of structure is zero $|A(x_{d_i}, t_{d_i})|=0$. For each diagram we introduced a local coordinate system, with the point (x_{d_i}, t_{d_i}) taken as an origin. After that we determined averaged amplitude fields over a sequence of diagram. During averaging these diagrams were superposed so that their local coordinate systems coincided.

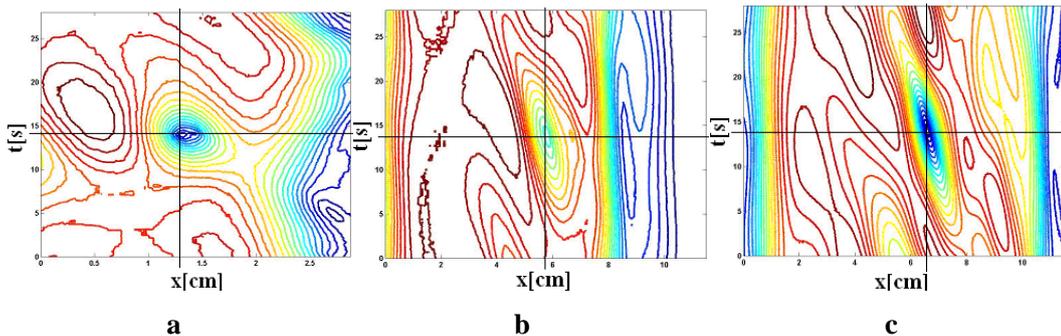


Fig.1 Results of conditional averaging of amplitude field for different supercriticalities. Intersection of black lines corresponds to the zero amplitude $|A(x_{d_i}, t_{d_i})|=0$ of structure. (a) $\varepsilon=0.038$; (b) $\varepsilon=0.040$, (c) $\varepsilon=0.042$

Obtaining of the averaged amplitude field allowed us to develop simple theoretical model to explain the characteristics of topological defects generated on the background of periodic vortex structure in Couette-Taylor flow.

How herds move: a study of polar self-propelled particles

É. Bertin ¹, M. Droz ², G. Grégoire ³

¹ *Laboratoire de Physique de l'ENS Lyon, CNRS UMR 5672, France*

² *University of Geneva, CH-1211 Geneva 4, Switzerland*

³ *Laboratoire Matière et Systèmes Complexes (MSC), UMR 7057, CNRS-Université Paris-Diderot*

Collective motion in animal groups is a ubiquitous phenomenon. Thus, during the two last decades, statistical physicists show a strong interest on the question of the emergence of such a new ordered state. A minimal model of moving vectors on a plane has been introduced by T. Vicsek *et al* [1] and has been studied mainly by simulations. But there are few concepts to understand out-of-equilibrium systems.

We will give an overview of the numerical results on two-dimensional collective motion [2,3]. Then we will show how we obtained a coarse-grained description of such a population [4]. Studying the properties of those modified Navier-Stokes equations allow us to give a better interpretation of the agent-based simulations.

References

- [1] T. Vicsek *et al*, *Phys. Rev. Lett.* **75** (1995), 1226.
- [2] J. Toner, Y. Tu, and S. Ramaswamy, *Annals of Physics* **318** (2005), 170.
- [3] H. Chaté *et al*, *Phys. Rev. E* **77** (2008), 046113.
- [4] É. Bertin, M. Droz, G. Grégoire, *Phys. Rev. E* **74** (2006), 022101.

Modulated spiral in a Couette-Taylor system submitted to a high radial temperature gradient

R. Guillerm, A. Prigent, I. Mutabazi ¹

¹LOMC, FRE 3102 CNRS, University of Le Havre, 53 rue de Prony, 76058 Le Havre, France

We investigate experimentally the flow produced in a narrow gap and large aspect ratio Couette-Taylor system submitted to a high radial temperature gradient. Demineralised water is confined in the five millimetres gap between a rotating inner cylinder and a fixed outer cylinder. The geometrical parameters are fixed with aspect ratio and radius ratio respectively equal to 111.8 and 0.8. The control parameters of the system are the Grashof number Gr , related to the radial temperature gradient and the Taylor number Ta , related to the rotation of the inner cylinder. For this study, Gr is fixed and Ta is gradually increased. The flow is characterized by Kalliroscope flakes visualizations and velocity and temperature fields measurements using encapsulated thermochromic liquid crystals [1-2].

For small values of Ta , the base flow is composed of the circular Couette flow due to the rotation of the inner cylinder and a vertical flow induced by the radial temperature gradient. Above a critical value of Ta , the destabilization of the base flow gives rise to a wavy pattern described by the temperature, the axial velocity component and the vorticity fields of figures 1 (a, b, c) [3-4]. This pattern has a spiral shape. It spreads on the whole length of the system and corresponds to alternate laminar and spiral areas (Fig.1d). It rotates as a whole at the mean angular velocity of the flow. By a suitable change of variable $t \rightarrow t'$ followed by a z -average, this modulated spiral pattern appears as a wave-packet and shares strong similarities with solitary waves (Fig.2).

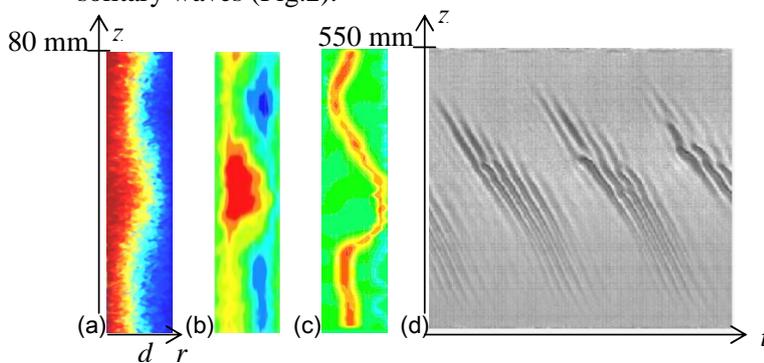


Fig.1 – (a) Temperature field, (b) axial velocity component field, (c) vorticity field and (d) light intensity space-time diagram of the modulated spiral pattern obtained for $Gr = 2405$ and $Ta = 11.2$

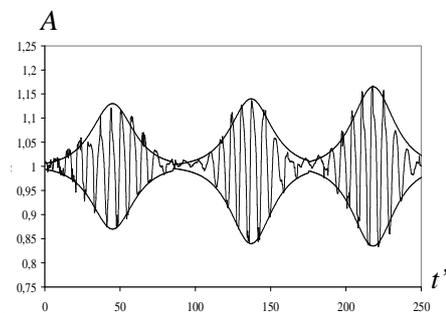


Fig.2 – z -averaged light intensity amplitude as a function of t' for $Gr = 4257$, $Ta = 12.5$.

References

- [1] SNYDER, H.A. et KARLSSON, S.K.F., *Phys. Fluids* **7** (10), (1964), 1696-1706.
- [2] LEPILLER, V., GOHARZADEH, A., PRIGENT, A. et MUTABAZI, I., *Euro. Phys. J.*, **B 61**, (2008), 445-455.
- [3] AKINO, N., KUNUGI, T., UEDA, M. et KUROSAWA, A., *Transport phenomena in turbulent flows (New York: Hemisphere)*, (1989), 807-827.
- [4] HAY, J.L., et HOLLINGSWORTH, D.K., *Experimental thermal and fluid science* **18**, pp. (1998), 251-257.

Experimental investigation of the dynamical correlations between magnetic nanoparticles in isotropic and textured superspin-glasses

K. Katsuyoshi¹, D. L'Hôte¹, S. Nakamae¹, Y. Tahri¹, C. Thibierge¹, E. Vincent¹, V. Mosser², A. Kerlain², M. Konczykowski³, E. Dubois⁴, V. Dupuis⁴, and R. Perzynski⁴

¹*Service de Physique de l'Etat Condensé, CEA Saclay, 91191 Gif/Yvette Cedex, France*

²*ITRON SAS, 76 av. P. Brossolette, 92240 Malakoff, France.*

³*Laboratoire des Solides Irradiés, Ecole Polytechnique, Palaiseau, France*

⁴*Lab. Liquides Ioniques et Interfaces Chargés, Univ. Pierre et Marie Curie, Paris, France*

A frozen and dense assembly of interacting magnetic nanoparticles exhibits a transition to a disordered and frustrated state, the superspin-glass at low temperature. Such a non-equilibrium system evolves continuously with time, and a still open question is the connexion of this aging with the growing of a dynamical correlation length. We have studied such a system made of maghemite (Fe₂O₃) nanoparticles suspended in glycerol, with a strong anisotropy due to “texturing”, i.e. freezing the ferrofluid in a strong magnetic field [1]. By using a global method based on the relaxation of the Zero-Field-Cooled Magnetization at various fields [2], we extracted the time dependence of the number of correlated superspins. Another method based on the local measurement of the superspin-glass magnetization fluctuations has also been used [3]. It is based on Hall microprobes operated with the spinning current technique [4]. Its first results will be presented.

References

- [1] S. Nakamae, Y. Tahri, C. Thibierge, D. L'Hôte, E. Vincent, V. Dupuis, E. Dubois, and R. Perzynski, Proc. 53rd Conference on Magnetism and Magnetic Materials (Austin, TX 10-14 Nov, 2008), *J. Appl. Phys.* 105, 07E318 (2009).
- [2] E. Wandersman, V. Dupuis, E. Dubois, R. Perzynski, S. Nakamae, E. Vincent, *Europhys. Lett.* **84**, 37011 (2008).
- [3] D. L'Hôte, S. Nakamae, F. Ladieu, V. Mosser, A. Kerlain, and M. Konczykowski, *J. Stat. Mech.: Theory & Exp.*, P01027 (2009).
- [4] A. Kerlain and V. Mosser, *Sensors & Actuators A* **142**, 528–532 (2008).

Finite-size effects on sub-critical bifurcations in high-dimensional lattices of coupled logistic maps

Paul Manneville

LadHyx, École Polytechnique, 91 128 Palaiseau, France.

Large assemblies of interacting nonlinear units give examples of complex systems in which collective behaviour emerges, characterised by the ordered evolution of averaged quantities in spite of strong chaos persisting at the scale of the individual units. Sub-critical transitions between collective regimes T2 and T4, with respective periods 2 and 4 not connected *via* a sub-harmonic bifurcation, have been observed in a 4-dimensional hyper-cubic lattice of coupled logistic maps $X_j^{t+1} = [1/(2d+1)] \sum_{\langle ij \rangle} r X_i^t (1-X_i^t)$, where $\langle ij \rangle$ means that node i belongs to the neighborhood of node j , and $d=4$ [1]. At finite and moderate size, intermittent switching between the two regimes takes place in the upper part of the bifurcation diagram (fig.1) while systematic decay with exponentially distributed transient lifetimes is observed in the lowest part of the coexistence region (fig.2). Statistical findings can be interpreted in terms of the average *activity* $M^t = \langle X_j^t \rangle$ escaping from potential wells in the presence of noise [2], the noise being here induced by chaos in the dynamics of local variables X_j . A consistent picture arises from the consideration of an effective stochastic dynamics with a deterministic part associated to the trend towards well bottoms counteracted by noise, the intensity of which varies as the inverse square root of the number of sites in the lattice. An analogy with sub-critical bifurcations between different turbulent regimes observed in some hydrodynamic systems is discussed.

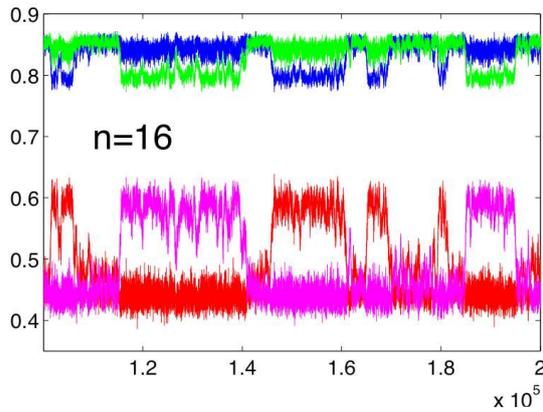


Fig.1 – Time series of the average X^t of the local variable X_i^t for $r = 4$ in a system of size n^4 , here $n=16$. T4 regime with two possible phases can be observed, interrupted by short sequences of T2 regime, e.g. for $t \sim 1.1 \cdot 10^5$, $1.45 \cdot 10^5$, or $1.7 \cdot 10^5$. The switching frequency decreases as n increases. For $n > 27$, the system can stay stuck in one or the other regime over more than 10^7 iterations when appropriately prepared.

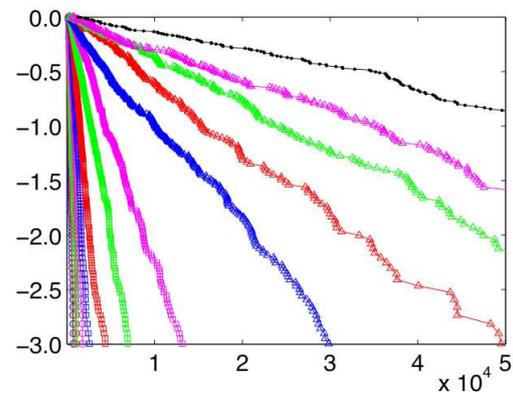


Fig.2 – Distribution of transient lifetimes t_{tr} , $P(t_{tr} > t)$, as functions of t (semi-log scale) in quench experiments from $r_i=3.98$ to various values of r_f from 3.905 (upper curve with black dots) down to 3.985: $P(t_{tr} > t) \sim \exp(t/\tau)$ with $\tau \sim 1/(r_{lim}-r)^4$ and $r_{lim} \sim 3.91$ for a system of size 27^4 .

References

- [1] H. Chaté and P. Manneville, *Prog. Theor. Phys.* 87 (1992)1-60.
- [2] N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland,1983).

Oscillations in bidisperse fluidized suspensions

A. Deboeuf, G. Gauthier, J. Martin and D. Salin

*Laboratoire FAST, Universités P&M Curie & Paris Sud & CNRS, UMR7608,
Bâtiment 502, campus universitaire, 91405 ORSAY CEDEX, France,*

We study experimentally the fluidization of bidisperse suspensions of macroscopic particles, at low Reynolds number. The suspensions are made of the same amount of large glass beads of $190\mu\text{m}$ and smaller ones, the size of which has been varied from 110 to $160\mu\text{m}$.

We follow the segregation process, when an initially mixed bidisperse suspension is fluidized at a fixed flow rate. The time evolution of the composition the suspension along the column is obtained with the help of an acoustic scanner, which records the composition-dependent attenuation of the sound propagation, measured at an appropriate frequency of 2MHz .

For a large enough size contrast (for small particles smaller than $150\mu\text{m}$), we observe a segregation process (Fig.1, left), induced by the different settling velocities. The process results in a stationary segregated state: a monodisperse suspension of small particles fluidized on top of a monodisperse suspension of large particles, with a transition zone slightly enlarged by the mixing of particles due to hydrodynamic dispersion.

For smaller size contrasts, however, the segregation process is interrupted by a fast mixing of the suspension, which is followed by a new segregation process, etc. As a result, the bidisperse fluidized suspension undergoes quasi periodic oscillations between segregated and mixed states (Fig.2, right). Although unexplained yet, the onset of the mixing could be caused by a columnar instability already observed in experiments of bidisperse sedimenting [1, 2].

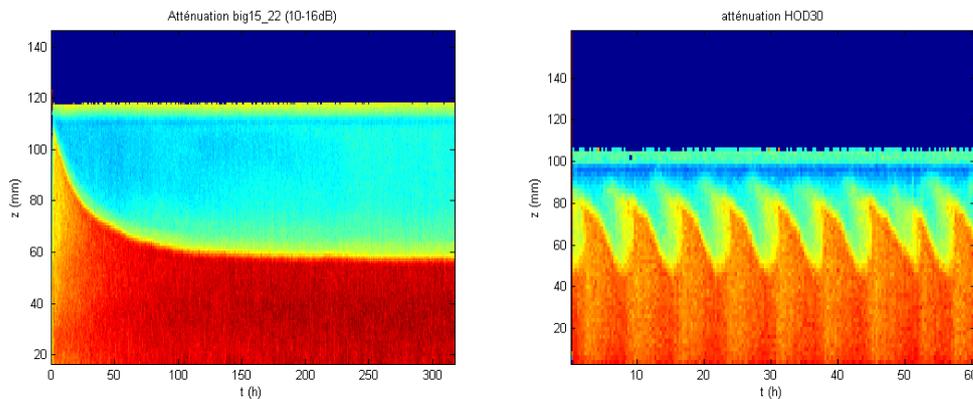


Fig.1 – Spatio-temporal diagram of sound attenuation, measured through a bidisperse suspension fluidized at a constant flow rate (at mean concentration $C\approx 40\%$). Whereas the suspension of beads sizes $190\mu\text{m}/150\mu\text{m}$ reaches a stationary segregated state (left figure), the one of beads sizes $190\mu\text{m}/155\mu\text{m}$ alternatively segregates and mixes, in a quasi periodical way.

References

- [1] On instabilities arising during sedimentation of two-component mixture of solids. R. H. Weiland, Y. P. Fessas, and B. V. Ramarao, *J. Fluid. Mech.* **141** (1984) 338.
- [2] Structure formation in bidisperse sedimentation, G. K; Batchelor and R. W. Jansa Van Rensburg, *J. Fluid. Mech.* **166** (1986) 379.

Intermittent regime in flow oscillations investigated by means of symbolic dynamics

Luc Pastur^{1,2}, François Lusseyran¹, Thierry Faure^{1,3}, Christophe Letellier⁴

¹ *LIMSI-CNRS, BP 133, F-91403 Orsay cedex, France*

² *Université Paris Sud 11, F-91405 Orsay Cedex, France*

³ *Université Pierre et Marie Curie, 75252 Paris Cedex 05, France*

⁴ *CORIA UMR 6614 - Université de Rouen, F-76801 Saint-Etienne du Rouvray Cedex, France*

A cavity flow driven by a laminar shear layer was investigated in the incompressible regime. Such flows are characterized by strong shear layer self-sustaining oscillations, due to complex feedback loop mechanisms [1]. Over some range of the working parameters (cavity length and height; inlet flow velocity), a mode-switching phenomenon is observed between two shear layer dominant modes, say f_1 and f_2 [2]. This regime is here investigated in terms of symbolic dynamics [3]. To that aim, velocity is measured at a single spatial point in the flow, past the cavity, using laser Doppler velocimetry (LDV). From the obtained time series, a space is reconstructed using principal components. The estimation of the attractor correlation dimension suggest that a space whose dimension should be at least 10 would be required

In spite of that, a first useful analysis is performed using a Poincaré section defined in the plane projection spanned by the first two components. Defining an angle with respect of the centre of the Poincaré section and an arbitrary axis, an angular first-return map is computed. Using the minima of the probability of visits of the angle interval to define a partition, a symbolic dynamics is introduced. Thus, each intersection of the trajectory with the Poincaré section is encoded by 0 or 1, depending on the angle associated with the orbit intersection. The two symbols, 0 or 1, reveal to be essentially associated to f_1 and f_2 , namely. Therefore, using a symbolic dynamics, based on an angular first-return map, provides an efficient tool for quantifying the observed intermittent regime. Moreover, it helps us to investigate the dynamics at time-scales of the order of a single period of the shear layer oscillations, that is, at a better resolution than common spectral-based methods, such as Fourier analysis, would do [4].

References

- [1] D. Rockwell, E. Naudascher, *Annual Review of Fluid Mechanics* **11** (1979) 67-94
- [2] L.R. Pastur, F. Lusseyran, T. M. Faure, Y. Fraigneau, R. Pethieu, P. Debesse, *Experiments in Fluids* **44** (2008) 597-608
- [3] J. Godelle, C. Letellier, *Physical Review E* **62** (2000) 7973-7981
- [4] F. Lusseyran, L. R. Pastur, C. Letellier, *Physics of Fluids* **20** (2008) 114101

Intermittency in granular flow

M. Rabaud, P. Gondret, R. Fischer

Lab. FAST, CNRS, Univ Paris-Sud 11, Université Pierre et Marie Curie-Paris6,
bât. 502, 91 405 Orsay Cedex, France

When a granular material is tilted above a maximum angle of stability, a surface avalanche starts that decreases the pile angle until the angle of repose where the flow stops. If during the avalanche the container rotates, two regimes can be observed: for slow rotation rates compared to the decrease rate of the pile angle during the avalanche, a succession of discrete avalanches are observed, when for larger rotation rate a continuous surface flow is observed.

We recently study experimentally and theoretically the transition between these two regimes in rotating cylinder geometry. Under the action of the continuous forcing imposed by the rotation of the tank we observed such transition between a limit cycle (discrete avalanche regime) and a fixed-point dynamics (continuous flow). In an intermediate range of angular velocity the system exchanges intermittently between the two regimes (figure 1). The mean fraction of continuous flow increases continuously from 0 to 1 in the transition range. The experimental results are recovered by a simple model equation of this dissipative system when a noise term is added [1].

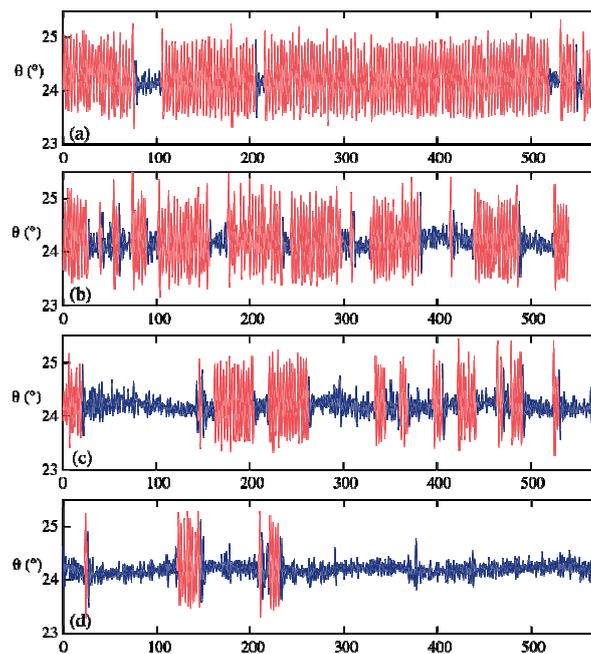


Fig.1 – Time evolution of the instantaneous pile angle for 4 increasing rotating velocity. In red, large oscillations corresponding to the discrete avalanche regime, in blue small oscillations of the pile angle corresponding to the continuous flow regime.

References

[1] R. Fischer, P. Gondret, M. Rabaud, “*Transition by intermittency from discontinuous avalanches to continuous granular flow*” preprint (2009)

Evolution of the Fabric Tensor in Amorphous Silica: via Molecular Dynamics Simulations

C. L. Rountree¹, M. Talamali², D. Vandembroucq², S. Roux³, E. Bouchaud¹

¹ CEA, IRAMIS, SPCSI, Grp. Complex Systems & Fracture, F-91191 Gif-sur-Yvette

² PMMH, ESPCI, 10 rue Vauelin 75231 Paris Cedex 05

³ LMT-Cachan, ENS de Cachan, CNRS-UMR 8535, Université Paris 6, 61 avenue du Président Wilson, F-94235 Cachan Cedex

AFM (Atomic Force Microscopy) experiments and MD (Molecular Dynamics) simulations have revealed a process zone (PZ) near the crack tip in amorphous silica (a-SiO₂). Within this process zone pores nucleate and coalesce with one another up to 20 nm ahead of the crack tip [1-4]. After which the cavities merge with the advancing crack to cause mechanical failure. Similarly, when a-SiO₂ sample is nanoindented one finds permanent damage under the indenter in the form of densified silica [6].

To shed light on the origin of irreversible deformation in amorphous media where the notion of dislocations is irrelevant, MD simulations have been performed in a-SiO₂ systems which are subjected to (1) a cyclic loading and unloading of the hydrostatic pressure and (2) a shearing force at room temperature. In particular, the so-called fabric tensor commonly used in granular physics is computed and allows to evidence anisotropy setting in the structure silica [7].

References

- [1] C.L. Rountree, S. Prades, D. Bonamy, E. Bouchaud, R. Kalia, C. Guillot. *J Alloy Compd.* 2007. 434: 60-63.
- [2] L. Van Brutzel, C.L. Rountree, R.K. Kalia, A. Nakano, P. Vashishta. *Mat Res Soc Symp Proc.* 2002. 703:V3.9.1- V3.9.6.
- [3] F. Celarie, S. Prades, D. Bonamy, L. Ferrero, E. Bouchaud, C. Guillot, C. Marliere. *Phys Rev Lett.* 2003. 90:075504.
- [4] S. Prades, D. Bonamy, D. Dalmas, E. Bouchaud C. Guillot. *Int J Sol Struct.* 2005. 42:637-645.
- [5] A. Perriot, D. Vandembroucq, E. Barthel, V. Martinez, L. Grosvalet, C. Martinet, B. Champagnon. *J Am Ceram Soc.* 2006. 89:596-601.
- [6] MF Ashby, DRH Jones. *Engineering Materials 1: An Introduction to Their Properties & Applications.* 2nd Edition. Butterworth-Heinmann. 1998.
- [7] C. L. Rountree, D. Vandembroucq, M. Talamali, E. Bouchaud, and S. Roux. *Phys Rev Lett.* 2009. 102:195501.

Lock exchange with autocatalytic reaction front

I. Bou Malham¹, N. Jarrige¹, J. Martin¹, N. Rakotomalala¹, L. Talon¹ & D.Salin^{1,2}

¹Laboratoire FAST, Universités P&M Curie & Paris Sud & CNRS, UMR7608,
Bâtiment 502, campus universitaire, 91405 ORSAY CEDEX, France,

²Institut Universitaire de France (IUF)

Autocatalytic reaction front between reacted and unreacted species may propagate as solitary waves, namely at a constant velocity and with a stationary concentration profile, resulting from a balance between molecular diffusion and chemical reaction. When a hydrodynamic flow field is forced on this reaction, the front still propagate as a solitary wave but with an enhanced propagation velocity and with a shape front reminiscent of front flame [1]. In most system, the fluid left behind the front has a different density leading to buoyancy driven flows and instabilities. In the present we revisit, with chemical reaction, the classical situation of lock exchange in gravity currents [2]. The latter refers to the reciprocal exchange of two fluids of different density in a horizontal channel; as the result the front spreads as the square root of time [2].

We present both experimental and numerical simulations analysis of buoyancy effects on the shape and the velocity of the Iodate Arsenous Acid (IAA) autocatalytic reaction fronts, propagating in a horizontal thin rectangular channel (Hele-Shaw cell). To complement the experiments, we use 2D lattice BGK numerical simulations of gap-averaged equations for the flow and the concentration, namely a Stokes-Darcy equation coupled with an advection-diffusion-reaction equation. We do observe stationary-shaped fronts, spanning the width of the cell and propagating along the cell axis. We analyze the scaling law for the shape and the travelling velocity of the fronts.

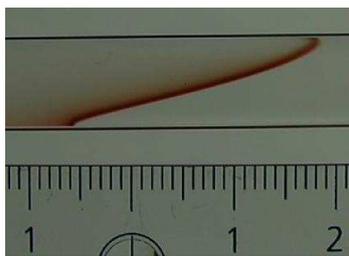


Fig.1 –Chemical gravity current in an 800 μm thick and 8 mm wide Hele-Shaw cell. The front travels from left to right leaving a lighter fluid behind. The gravity is vertical.

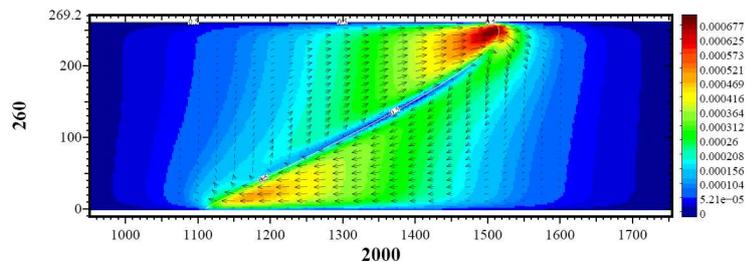


Fig.2 –Lattice BGK numerical simulations of the viscous chemical gravity current. Velocity field : the color scale gives the intensity of the velocity . We clearly see the vortex traveling with the front.

References

- [1] Pattern of reaction diffusion front in laminar flows. M. Leconte, J. Martin , N. Rakotomalala and D. Salin, *Phys. Rev. Letter.*, **90** (2003) 128302 .
- [2] H.E. Huppert, *J. Fluid. Mech.* **121** (1982) 43

Dynamic fracture: how brittle are brittle amorphous solids?

J. Scheibert^{1,2}, C. Guerra^{1,3}, F. Célarie^{1,2}, D. Dalmas², D. Bonamy¹

¹CEA, IRAMIS, SPCSI, Grp. Complex Systems & Fracture, F-91191 Gif sur Yvette, France

²Unité Mixte CNRS/Saint-Gobain "Surface du Verre et Interfaces" 39 Quai Lucien Lefranc, 93303 Aubervilliers cedex, France

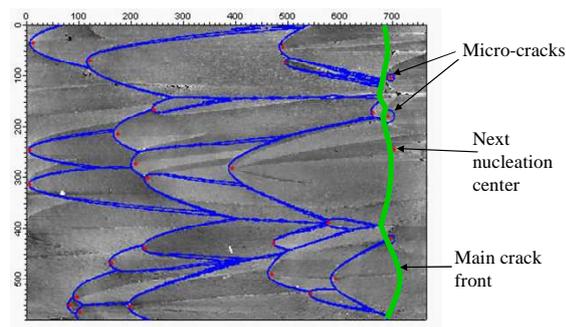
³Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León, Ave. Universidad, S/N, Ciudad Universitaria, C.P. 66450, San Nicolás de los Garza, NL, Mexico

Linear Elastic Fracture Mechanics [1] quantitatively describes crack propagation in brittle amorphous solids for crack velocities v smaller than v_b , a critical velocity governing the onset of micro-branching instability [2]. In this range, all dissipation processes, irrespective of their physical origin, can be gathered in one material-dependent quantity, the fracture energy Γ . Intriguingly, even below v_b , Γ was found much higher than at crack initiation.

Experiments in an extensively investigated brittle amorphous material, PMMA, allowed us to measure the velocity dependence of the dynamic fracture energy between 0 and v_b . The latter reveals a new critical velocity, v_a , well below v_b , at which Γ exhibits an abrupt 3-folds increase. The nature of this transition is unravelled by a post-mortem fractographic analysis of the samples: v_a corresponds exactly to the appearance, on the fracture surfaces, of conic patterns. In many materials, these are the signature of damage spreading through the nucleation and growth of micro-cracks ahead of the propagating crack [3-4]. Nominally brittle amorphous solids therefore behave as quasi-brittle for crack velocities above v_a .

Following [3], we consider a simple geometrical model to describe the nucleation and growth of micro-cracks. By feeding this model with the actual locations and nucleation thresholds of the micro-cracks, we successfully reconstruct the complex patterns observed (Fig.1), in the whole velocity range $v_a - v_b$, with both micrometer and microsecond resolutions. We present statistical characterizations of the velocity-dependent geometrical parameters of the population of micro-cracks, the origin of which is left as an open question. More generally, this system emerges as a good experimental model system to test descriptions of damage processes in amorphous materials, including rate dependence and spatial interactions.

Fig.1 – Reconstruction of crack propagation from *post-mortem* fracture surface images in the micro-cracking regime.



References

- [1] L. B. Freund, *Dynamic Fracture Mechanics* (Cambridge University Press, 1990)
- [2] E. Sharon and J. Fineberg, *Nature* **397** (1999) 333
- [3] K. Ravi-Chandar and B. Yang, *J. Mech. Phys. Solids* **45** (1997) 535
- [4] D. Hull, *Fractography: Observing, Measuring and Interpreting Fracture Surfaces Topography* (Cambridge University Press, 1999)

Dynamical length scale in glass transition: experimental study

C. Thibierge, D. L'Hôte, F. Ladieu, C. Brun

Groupe Magnétisme, frustration et désordre, DSM/DRECAM/SPEC, CEA Saclay, 91 191 Gif sur Yvette, France,

The vitreous slowing down of supercooled liquids at the glass transition is still poorly understood. Many models intend to describe its physical mechanism. A particularly interesting scenario is the one of dynamical heterogeneities: timescales are spatially distributed in the supercooled liquid with slow or fast relaxing regions. Finally slow regions become fast and vice versa. They are called dynamical heterogeneities. Our issue is to measure the number of correlated molecules of these regions. When the temperature decreases the relaxation time of the liquid increases and this may be linked to an increase of these regions' length scale and also to an increase of the energy needed to relax. This image refers to phase transition concepts. Dynamical heterogeneities have been shown in simulations or granular experiments but this number of correlated molecules has never been directly measured in supercooled liquids.

A method to measure this number of correlated molecules has been proposed by Bouchaud and Biroli [1]. It is based on the measure of the polarisation P third harmonics. $P = \chi_1 E + \chi_2 E^2 + \chi_3 E^3$ with E the electric field. $\chi_2 E^2 = 0$ for symmetry reasons and χ_3 is proportional to the number of correlated molecules. We have developed an original experiment to measure this non-linear susceptibility by dielectric techniques in supercooled glycerol near the glass temperature [2]. We succeeded in following the temperature evolution of this number of molecules as in phase transitions. This number as also been evaluated from linear susceptibilities [3] and the two results are compared.

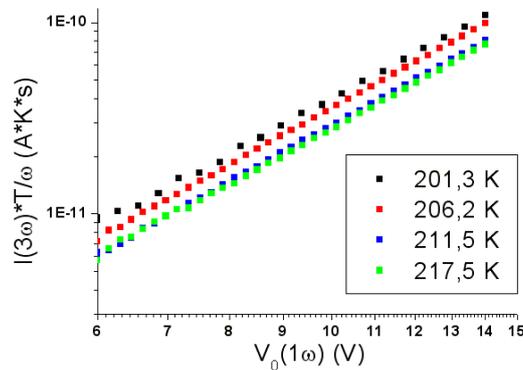


Fig.1 – The growing of a quantity proportional to $\chi_3 E^3$, with $I(3\omega)$ the measured third harmonic current as a function of the source voltage $V = E \times$ sample thickness.

References

- [1] Bouchaud and Biroli *PRB* **72** 064204 (2005)
- [2] Thibierge, L'Hôte, Ladieu and Tourbot, *Rev. Sci. Instrum.* **79** 103905 (2008)
- [3] Dalle-Ferrier *et al.* *PRE* **76** 041510 (2007)

Trajectories of sand grains in a wave boundary layer

T-D. Chu¹, A. Jarno-Druaux¹, A. Ezersky² & F. Marin¹

¹LOMC, FRE 3102 CNRS, 53 rue de Prony, BP 540, 76058 Le Havre Cedex, France

²M2C, UMR CNRS 6143, 2-4 rue des tilleuls, 14000 Caen Cedex, France

Sea beds are often covered by small scale sedimentary structures in the coastal zone [1, 2] (typical wave length ~ 10 cm); these structures result from complex interactions between the flow and the sediments. Many experimental and theoretical works have been conducted on the morphology of these structures which are generally called ripples, due to their significant effect on sediment transport, wave attenuation and boundary layer structure. Most of these studies have been carried out considering size homogeneous sediments; however, the *in-situ* works generally show the presence of a large variety of sediment size, and segregation phenomena [3]. The aim of the present work is to bring a contribution to the study of this segregation; experiments are carried out in a wave flume where ripples are generated, and the trajectories of sediments in suspension close to the bed are recorded. Figure 1 shows a typical view of the grain segregation in the case of bi-dispersed sediments [4]. In order to explain this segregation, the bedforms are fixed and a few sediment grains are dropped in the flow, above the rippled bed. The wave maker is switched off and the grain trajectories (Fig 2) are recorded using a high resolution video camera. These trajectories are compared with the results predicted by a theoretical model, and an explanation of the observed segregation based on the existence of a mean flow towards the ripple crests is proposed.

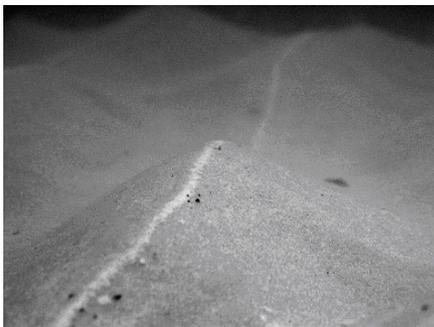


Fig 1. Typical view showing sediments segregation on a rippled bed (light grains in white and heavy grains in grey).

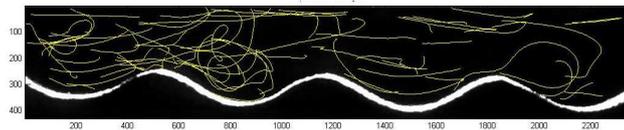


Fig 2. Grain trajectories in suspension above the rippled bed beneath waves.

References:

- [1] R. A Bagnold, Motion of waves in shallow water. Interaction between waves and sand bottom, with an additional note by Sir G. I Taylor: Proceedings of the Royal Society of London, Series A 187: 1-15.
- [2] J. Leunetel-Levaslot, Dynamique de formation des réseaux de rides de sable en canal à houle. PhD, University of Le Havre.
- [3] Pieter C. Roos, Suzanne J. M. H Hulscher, Fenneke Van der Meer, Thaiëne A. G. P van Dijk, Irene G. M Wientjes and Joris van den Berg, Grain size sorting over offshore sandwaves: Observations and modelling. River, Coastal and Estuarine Morphodynamics: RCEM 2007- Dohmen –Janssen & Hulscher (eds). © 2008 Taylor & Francis Group, London, ISBN 978-0-415-45363-9, p 649 – 656.
- [4] A.B Ezersky and F. Marin, Segregation of sedimenting grains of different densities in an oscillating velocity field of strongly nonlinear surface waves. Physical Review E, Vol.78, No 2, doi: 10.1103/PhysRevE.78.022301, 2008.

Elastic instability of polymer solutions in the Couette-Taylor system

Fayçal Kelai, Olivier Crumeyrolle and Innocent Mutabazi

Laboratoire Ondes et Milieux Complexes, FRE 3102, CNRS-Université du Havre 53, rue de Prony, B.P. 540, F-76058 Le Havre Cedex, France

The study of transition to turbulence in Newtonian flows is governed by the well-known Navier-Stokes equations. In case of viscoelastic flows, no such universal equations exist and the transition to turbulence in such fluids is much more difficult to predict without any theoretical guidelines [1]. Viscoelastic liquids intervene in many industrial processes in which production rates or final product quality can be limited by flow instabilities. Therefore strong motivations exist for experimental investigation of viscoelastic flow behavior.

The present work report the behavior of the viscoelastic Couette-Taylor flow with dilute solutions of high-molecular-mass polyethyleneoxide in viscous solvents obtained by dissolving low-molecular-mass polyethyleneglycol in water. The solution viscosity varied from 41.5 mPa.s to 113.3 mPa.s and did not depend on shear rate. The solution was sheared in vertical Couette-Taylor cell with inner radius $a = 4$ cm and outer radius $b = 5$ cm. The gap width is $d = 1$ cm and its height is $L = 45.9$ cm leading to aspect ratio $\Gamma = H/d = 45.9$ and radius ratio $\eta = a/b = 0.8$. Experiments were conducted at room temperature. Kalliroscope flakes were used for flow visualization together with a laser light sheet. Rotation speeds of both cylinders were controlled, but for a given experiment only one of the cylinders is rotating with angular velocity Ω . Above a critical rotation rate Ω_c , the flow bifurcated to flow patterns typical of purely elastic regime (Fig. 1). These *elastic instability*-induced patterns exhibited disordered domains separated by strong radial flows [2]. The critical value of the onset of this purely elastic instability agrees well (Fig. 2) with theoretical criterion from [3] when only the inner cylinder is rotated, while preliminary result obtained with rotation of the outer cylinder exhibited a significant hysteresis.

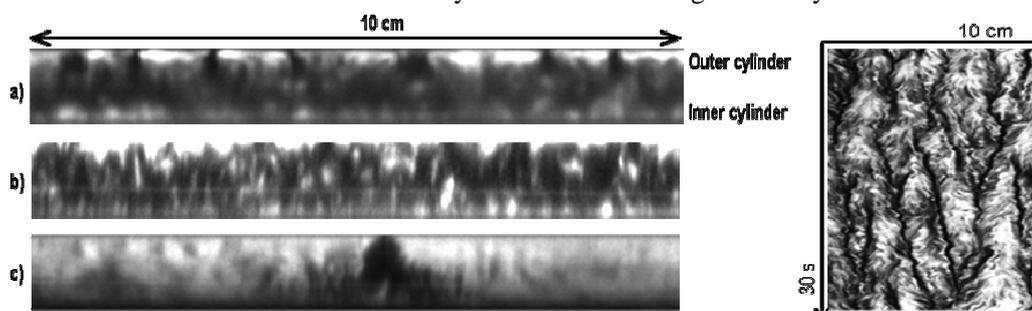


Fig. 1: Typical flow patterns for : a) inner rotating cylinder, $\Omega = \Omega_c$; b) outer rotating cylinder, $\Omega = \Omega_c$; c) outer rotating cylinder, $\Omega = 0.88\Omega_c$; d) space-time plot extracted from pattern (a) at $r = 0.25 + a$

References

- [1] R. G. Larson, E. S. G. Shaqfeh & S. J. Muller, *J. Fluid Mech.* **218**, 573 (1990).
- [2] A. Groisman & V. Steinberg, *Phys. Fluids*, **10**, 2451 (1998).
- [3] G. H. McKinley, A. Oztekin, J. A. Byars, and R. A. Brown. *J. Fluid Mech.* **285**, 123 (1995).

The renormalized fragmentation equation and its exact solutions

M. A. Gorokhovski¹ and V. L. Saveliev²

¹ LMFA, CNRS-UMR 5509, Ecole Centrale de Lyon, 36 av. Guy de Collongue, 69134 Ecully,

² Institute of Ionosphere, National CSRT, Kamenskoe Plato, 050020 Almaty, Kazakhstan

The fragmentation equation describes the time evolution of particles system, when particles break up. Spray atomization, turbulent eddy decay, solid particle decomposition, rock crushing, polymer degradation, impact of meteorites and network branching represent the examples of such fragmentation. In this paper, the renormalized form of the fragmentation equation is obtained, with the spectrum and the frequency of fragmentation, as arbitrary functions. If the fragmentation frequency is a power function of size, a simple exact solution to this equation is obtained, providing for stationary mass flux, from large scales towards zero scales. This work is a further development of results obtained in [1-3].

1. The fragmentation equation has the following form:

$$\frac{\partial F(r)}{\partial t} = (q_0 \hat{I}_+^p - 1) \nu(r) F(r) \quad (1); \quad \hat{I}_+^p F = \int_0^1 \frac{d\alpha}{\alpha} Q(\alpha) F\left(\frac{r}{\alpha}\right); \quad \int_0^1 Q(\alpha) d\alpha = 1; \quad q_0 = \frac{1}{\langle \alpha^3 \rangle_Q} \quad (2)$$

where $F(r)$ is the number distribution function, $Q(\alpha)$ is the probability density function of break-up, q_0 is the mean number of particles formed by a break-up and $\nu(r)$ is the break-up frequency.

2. The mass distribution function. With evolution in time, the total mass is conserved but not the total number of particles. This motivates to introduce the energy distribution function $f(r)$, instead of using $F(r)$; the norm of the former one will not change. The relation between both functions, $f(r) = r^3 F(r)$, leads to the following fragmentation equation:

$$\frac{\partial f(r)}{\partial t} = (\hat{I}_+ - 1) \nu(r) f(r) \quad (3); \quad \hat{I}_+ f = \int_0^1 \frac{d\alpha}{\alpha} q(\alpha) f\left(\frac{r}{\alpha}\right); \quad \int_0^1 q(\alpha) d\alpha = 1; \quad q(\alpha) = \frac{\alpha^3 Q(\alpha)}{\langle \alpha^3 \rangle_Q} \quad (4)$$

3. Exact solutions for arbitrary fragmentation spectrum $q(\alpha)$. The first term in (2) represents difference between two terms; each of them can be very large, and even infinite (if to give up the normalization requirement of $q(\alpha)$). In order to overcome this problem, the renormalization of (3) was obtained, describing evolution of an infinite system of particles in the form of continuity equation. When $\nu(r) = cr^\mu$, a simple solution to this equation, verifies to be:

$$f(r, t) = \frac{j_0}{\langle -\ln \alpha \rangle} \cdot r^{-(\mu+1)} + t \cdot j_0 \delta(r), \quad 0 \leq r < \infty \quad (5) \quad \text{where} \quad \langle (-\ln \alpha)^n \rangle = \int_0^1 (-\ln \alpha)^n q(\alpha) d\alpha.$$

With accumulation of particles of zero size, this solution provides for stationary mass flux j_0 , from infinity towards zero size. Other self-similar solutions to the renormalized fragmentation

equation have the following form: $f(r, t) = |\eta|^{\frac{1}{\mu}} \phi\left(|\eta|^{\frac{1}{\mu}} r, t\right); \quad \eta(t) = 1 + \tau^{-1} t; \quad \tau < 0; \quad 0 \leq t < \infty \quad (6)$

where $\phi(r)$ is defined from the solution of the equation $\frac{\partial}{\partial r} r \left[-\frac{1}{\mu\tau} + \langle -\ln \alpha \rangle \hat{I}_+^{(1)} \nu \right] \phi(r) = 0 \quad (7)$

References

- [1] M.A. Gorokhovski and V.L. Saveliev, *J. Phys. D: Appl. Phys.* **41**, (2008), 085405
- [2] V.L. Saveliev and M.A. Gorokhovski, *Phys. Rev E* **72**, (2005), 016302
- [3] V.L. Saveliev and M.A. Gorokhovski, *Phys. Scr.* **T132**, (2008) 014005

Combinatorics, Physics and Complexity

Organizing Committee

- Gérard H.E. DUCHAMP
LIPN, University of Paris XIII, France
`Gerard.Duchamp@lipn.univ-paris13.fr`

Description

The (new) science of Complex Systems likes to consider the notion of "level of description". Many tools from Combinatorics, Computer Science and Mathematics are there needed, be it at the level where "things do happen" or at higher levels where statistical exactly solved models of a striking precision can be experimented and set. This special session is intended to be a contribution to the exposition of these tools/results. Includes (not exhaustive)

- Mathematical and Combinatorial local rules
- Theoretical Computer Science for Complex Systems
- Statistical Physics
- Exactly solved models
- Combinatorial models for and of emergence
- Explicit schemes for analysis of dynamical systems: finite (combinatorial) and infinite (based on functional analysis)

Contents

Asymptotic analysis and convergence of some leader election algorithms	
Christian LAVAUT	53
Learning self organizing maps as a Markov mixture	
Mustapha LEBBAH , Younès BENNANI , Nicoleta ROGOVSCHI	53
Statistics on graphs, exponential formula and combinatorial physics	
G�rard H.E. DUCHAMP , Laurent POINSOT , Silvia GOODENOUGH , Karol A. PENSON	60
Exponential random graphs as models of overlay networks	
Moez DRAIEF	64
Coding rhombus tilings by multidimensional words: a first attempt	
Damien JAMET	64
Synchronization of countable cellular systems, localization of quasi-periodic solutions of autonomous differential systems	
Laurent GAUBERT	64
Drawing solution curve of a differential equation	
Farida BENMAKROUHA , Christiane HESPEL , Edouard MONNIER	75
Generating series: a combinatorial computation	
Farida BENMAKROUHA , Christiane HESPEL	79
On approximation of nonlinear generating series by rational series	
Mikhail V. FOURSOV , Christiane HESPEL	84
Counting rooted and unrooted triangular maps	
Samuel VIDAL , Michel PETITOT	91
About a group of Drinfel'd associators	
Hoang Ngoc MINH	97
Combinatorics & Schelling's model	
Cyril BANDERIER , Hanane TAFAT	97
Systemic modelling	
Daniel KROB	97

Asymptotic analysis and convergence of some leader election algorithms

Christian LAVALT

We start with a set of n players. With some probability $P(n, k)$, we kill $n - k$ players; the other ones stay alive, and we repeat with them. What is the distribution of the number X_n of rounds before getting only one player? We present a probabilistic analysis of this very general algorithm (expectation and moments of X_n), under some conditions on the probability distributions $P(n, k)$ (if any), including stochastic monotonicity and the assumption that roughly a fixed proportion of the players survive in each round. Applications of the general result include asymptotic analysis of leader election algorithms where players are eliminated by independent coin tosses.

Learning Self-Organizing Maps as a Mixture Markov Models

Mustapha Lebbah, Younès Bennani and Nicoleta Rogovschi

Abstract—This paper describes a new algorithm to learn Self-Organizing map as Markov Mixture Models. Our model realizes an unsupervised learning using unlabelled evolutionary data sets, namely those that describe sequential data. The new formalism that we present is valid for all structure of graphical model. We use E-M (Expectation-Maximisation) standard algorithm to maximize the likelihood. The graph structure is integrated in the parameter estimation of Markov model using a neighborhood function to learn a topographic clustering of not i.i.d data set. The new approach provides a self-organizing Markov model using an original learning algorithm.

Index Terms—Self-Organizing clustering, Markov Models, Sequential data, Expectation-Maximisation, graphical model

I. INTRODUCTION

Since many years, temporal and spatial sequences have been the subject of investigation in many fields, such as statistics, pattern recognition, web mining, and bioinformatic. The easiest way to treat sequential data would be simply to ignore the sequential aspects and treat the observations as independent and identically distribution (i.i.d) in the first stage. For many applications, the i.i.d assumption will be a poor one. Often in many application the treatment is decomposed in two steps; the first one is the clustering task with i.i.d assumption. In second stage the result of clustering is used to learn a probabilistic model by relaxing the i.i.d. assumption, and one of the simplest ways to do this is to consider a Markov model.

Hidden Markov Models (HMMs) are the most well-known and practically used extension of Markov model. They offer a solution to this problem introducing, for each state, an underlying stochastic process that is not known (hidden) but could be inferred through the observations it generates. In fact the probabilistic graphical modelling motivates different graphical structures based on the HMM [1], [2]. Another variant of the HMM worthy of mention is the factorial hidden Markov model [3], in which there are multiple independent Markov chains of latent variables, and the distribution of the observed variable at a given time step is conditional on the states of all of the corresponding latent variables at that same time step. Many related models, such as hybrids of HMMs with artificial neural networks [4], [5], [6]. Clearly, there are many possible probabilistic structures that can be constructed according to the needs of particular applications. Graphical models provide a general formalism for motivating, describing, and analysing such structures. Therefore, it will be very important to have algorithms able to infer from a

data set of sequences not only the probability distributions but also the topological structure of the model, i.e., the number of states and the transitions interconnecting them. Unfortunately, this task is very difficult and only partial solutions are today available [7], [8], [9]. In order to overcome the limitations of HMMs, in [9] the author proposes a novel and an original machine learning paradigm, which is titled topological HMM, that embeds the nodes of an HMM state transition graph in Euclidian space. This approach models the local structure of HMM and extract their shape by defining a unit of information as a shape formed by a group of symbols of a sequence.

Others attempts have been made for combining HMMs and SOMs (Sel-Organizing Map of Kohonen) to form hybrid models that contain the clustering power of SOM with the sequential time series aspect of HMMs [6]. In many of these hybrid architectures, SOM models are used as front-end processors for vector quantization, and HMMs are then used in higher processing stages. In [10], [11], a vector sequence is associated with a node of SOM using DTW (dynamic time warping) model. Others works exist and differ in the manner of combination [12], [13], [14]. In [14] the authors propose an original combined model which is the offspring of a crossover between the SOM algorithm and the HMM theory. The model's core consists of a novel unified/hybrid SOM-HMM algorithm where each cell of SOM map presents an HMM. The model is coupled with a sequence data training method, that blends together the SOM unsupervised learning and the HMM dynamic programming algorithms. Of course, there is a lot more litterature on HMMs and their applications than can be covered here, but this survey wants to be representative of the issues addressed here. However, in the other the organization process are not integrated explicitly in HMM approach.

The aim of this paper is to built a new model for automating and self-organizing the construction of a statistical generative model of a data set of spatial sequences. In our model 3M-SOM (Self-Organizing Maps as a Mixture Markov Models), we consider that we have one Markov chain forming a grid. The generation of the observed variable at a given time step is conditional on the neighborhood states at that same time step. Thus, a high proximity implies a high probability to contribute to generation. This proximity is quantified using neighborhood function. The same principle is used by Kohonen algorithm for i.i.d data set [15]. In our case we focus about not i.i.d observations. We use Expectation-Maximization (EM) algorithm to maximize the likelihood. The formalism that we present is valid for all structure of graph model. In our case we prefer to define the HMM architecture as map (grid).

This paper is organized as follows. In section II we present the

Mustapha Lebbah, Younès Bennani and Nicoleta Rogovschi are with LIPN-UMR 7030 - CNRS, Université Paris 13. 99, av. J-B Clément F-93430 Villetaneuse e-mail: firstname.second-name@lipn.univ-paris13.fr.

model we propose 3M-SOM. In section III we discuss the self-organizing process integrated in HMM. Finally, conclusions and some future works are provided.

II. SELF-ORGANIZING MARKOV MODEL

We assume that the HMM architecture is a lattice \mathcal{C} , which has a discrete topology (discrete output space) defined by an undirect graph. Usually, this graph is a regular grid in one or two dimensions. We denote the number of cells (nodes, state) in \mathcal{C} as K . For each pair of cells (c, r) on the graph, the distance $\delta(c, r)$ is defined as the length of the shortest chain linking cells r and c . The architecture of 3M-SOM model is inspired from probabilistic topographic clustering of i.i.d observations using a self-organizing map model of Kohonen [16], [17], [18].

A. Mixture model and Self-Organizing

We assume that each element \mathbf{x}_n of sequence observation $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N\}$ is generated by the following process: We start by associating to each cell (state) $c \in \mathcal{C}$ a probability $p(\mathbf{x}_n/c)$ where \mathbf{x}_n is a vector in the data space. Next, we pick a cell c^* from \mathcal{C} according to the prior probability $p(c^*)$. For each cell c^* , we select an associated cell $c \in \mathcal{C}$ following the conditional probability $p(c/c^*)$. All cells $c \in \mathcal{C}$ contribute to the generation of an element \mathbf{x}_n with $p(\mathbf{x}_n/c)$ according to the proximity to c^* described by the probability $p(c/c^*)$. Thus, a high proximity to c^* implies a high probability $p(c/c^*)$, and therefore the contribution of state c to the generation of \mathbf{x}_n is high.

Let us introduce a K -dimensional binary random variable as latent variable \mathbf{z}_n and \mathbf{z}_n^* having a 1-of- K representation in which a particular element z_{nk} and z_{nk}^* is equal to 1 and all other elements are equal to 0. Each component z_{nk}^* and z_{nk} indicate a couple of state responsible for the generation of an element of the observation. Using this notation we can rewrite:

$$p(\mathbf{x}_n/c) \equiv p(\mathbf{x}_n/z_{nc} = 1) \equiv p(\mathbf{x}_n/\mathbf{z}_n)$$

and

$$p(c/c^*) = p(z_{nc} = 1/z_{nc^*}^* = 1) \equiv p(z_{nc}/z_{nc^*}^*) \equiv p(\mathbf{z}_n/\mathbf{z}_n^*)$$

is assumed to be known. To introduce the self-organizing process in the mixture model learning, we assume that $p(z_{nc}/z_{nc^*}^*)$ can be defined as:

$$p(z_{nc}/z_{nc^*}^*) = \frac{\mathcal{K}^T(\delta(c, c^*))}{\sum_{r \in \mathcal{C}} \mathcal{K}^T(\delta(r, c^*))},$$

where \mathcal{K}^T is a neighbourhood function depending on the parameter T (called temperature): $\mathcal{K}^T(\delta) = \mathcal{K}(\delta/T)$, where \mathcal{K} is a particular kernel function which is positive and symmetric ($\lim_{|x| \rightarrow \infty} \mathcal{K}(x) = 0$). Thus \mathcal{K} defines for each state $z_{nc^*}^*$ a neighbourhood region in the graph \mathcal{C} . The parameter T allows the control of the size of the neighbourhood influencing a given cell on the map \mathcal{C} . As with the Kohonen algorithm for i.i.d observations, we decrease the value of T between two values T_{max} and T_{min} .

For the better understanding we have used similar notations

as in the book [19, chap. 13]. We denote the set of all latent variables by \mathbf{Z}^* and \mathbf{Z} , with a corresponding row \mathbf{z}_n^* and \mathbf{z}_n associated to each sequence element \mathbf{x}_n . Now assume that, for each sequence observation in X , corresponds the couple of latent variable \mathbf{Z} and \mathbf{Z}^* . We denote by $\{\mathbf{X}, \mathbf{Z}, \mathbf{Z}^*\}$ the complete data set, and we refer to the observed data \mathbf{X} as incomplete.

The set of all model parameters is denoted by θ , the likelihood function is obtained from the joint distribution by marginalizing over the latent variables \mathbf{Z}^* and \mathbf{Z}

$$p(\mathbf{X}; \theta) = \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}, \mathbf{Z}^*; \theta) \quad (1)$$

Because the joint distribution $p(\mathbf{X}, \mathbf{Z}, \mathbf{Z}^*; \theta)$ does not factorize over n , we cannot treat each of the summations over \mathbf{z}_n^* and \mathbf{z}_n independently.

An important concept for probability distributions over multiple variables is that of conditional independence [20]. We assume that the conditional distribution of \mathbf{X} , given \mathbf{Z}^* and \mathbf{Z} , is such that it does not depend on the value of \mathbf{Z}^* . Often this assumption is used for graphical model, so that $p(\mathbf{X}/\mathbf{Z}, \mathbf{Z}^*) = p(\mathbf{X}/\mathbf{Z})$. Thus the joint distribution of the sequence observations is equal to:

$$p(\mathbf{X}, \mathbf{Z}^*, \mathbf{Z}) = p(\mathbf{Z}^*)p(\mathbf{Z}/\mathbf{Z}^*)p(\mathbf{X}/\mathbf{Z})$$

thus we can rewrite the marginal distribution as

$$p(\mathbf{X}; \theta) = \sum_{\mathbf{Z}^*} p(\mathbf{Z}^*) \sum_{\mathbf{Z}} p(\mathbf{Z}/\mathbf{Z}^*)p(\mathbf{X}/\mathbf{Z}) \quad (2)$$

We note that

$$p(\mathbf{X}/\mathbf{Z}^*) = \sum_{\mathbf{Z}} p(\mathbf{Z}/\mathbf{Z}^*)p(\mathbf{X}/\mathbf{Z}) \quad (3)$$

B. Cost function and optimization

Considering a map \mathcal{C} as Markov model, we allow the probability distribution of \mathbf{z}_n^* to depend on the state of the previous latent variable \mathbf{z}_{n-1}^* through a conditional distribution $p(\mathbf{z}_n^*|\mathbf{z}_{n-1}^*)$. Because the latent variables are K -dimensional binary variables, this conditional distribution corresponds to a table of probabilities that we denote by \mathbf{A} . The elements of \mathbf{A} are known as transition probabilities denoted by $A_{jk} = p(z_{nk}^* = 1/z_{n-1,j}^* = 1)$, with $\sum_k A_{jk} = 1$. So the matrix \mathbf{A} has maximum of $K(K-1)$ independent parameters. In our case the number of transitions are limited by the grid (map). We can then write the conditional distribution explicitly in the form

$$p(\mathbf{z}_n^*/\mathbf{z}_{n-1}^*, \mathbf{A}) = \sum_{k=1}^K \sum_{j=1}^K A_{jk}^{z_{n-1,j}^* z_{nk}^*}$$

All of the conditional distributions governing the latent variables share the same parameters \mathbf{A} .

The initial latent state \mathbf{z}_1^* is special in that it does not have a parent cell, and so it has a marginal distribution $p(\mathbf{z}_1^*)$ represented by a vector of probabilities π with elements $\pi_k = p(z_{1k}^* = 1)$, so that $p(\mathbf{z}_1^*|\pi) = \prod_{k=1}^K \pi^{z_{1k}^*}$, where $\sum_k \pi_k = 1$.

The model parameters are completed by defining the conditional distributions of the observed variables $p(\mathbf{x}_n/\mathbf{z}_n; \phi)$, where ϕ is a set of parameters governing the distribution which is known as emission probabilities in HMM model.

Because \mathbf{x}_n is observed, the distribution $p(\mathbf{x}_n/\mathbf{z}_n; \phi)$ consists, for a given value of ϕ , of a vector of K numbers corresponding to the K possible states of the binary vector \mathbf{z}_n . We can represent the emission probabilities in the form

$$p(\mathbf{x}_n/\mathbf{z}_n; \phi) = \prod_{k=1}^K p(\mathbf{x}_n; \phi_k)^{z_{nk}}$$

The joint probability distribution over sequence observed variables and both latent \mathbf{Z} and \mathbf{Z}^* is then given by

$$p(\mathbf{X}, \mathbf{Z}^*, \mathbf{Z}; \theta) = p(\mathbf{Z}^*; \mathbf{A}) \times p(\mathbf{Z}/\mathbf{Z}^*) \times p(\mathbf{X}/\mathbf{Z}; \phi)$$

$$\begin{aligned} p(\mathbf{X}, \mathbf{Z}^*, \mathbf{Z}; \theta) &= \left[p(\mathbf{z}_1^*|\pi) \prod_{n=2}^N p(\mathbf{z}_n^*/\mathbf{z}_{n-1}^*; \mathbf{A}) \right] \\ &\times \left[\prod_{i=1}^N p(\mathbf{z}_i/\mathbf{z}_i^*) \right] \\ &\times \left[\prod_{m=1}^N p(\mathbf{x}_m/\mathbf{z}_m; \phi) \right] \end{aligned} \quad (4)$$

$\theta = \{\pi, \mathbf{A}, \phi\}$ denotes the set of parameters governing the model. Most of our discussion of the self organizing Markov model will be independent of the particular choice of the emission probabilities. It's not obvious to maximize the likelihood function, because we obtain complex expressions with no closed-form solutions. Hence, we use the expectation maximization algorithm to find parameters for maximizing the likelihood function. EM algorithm starts with some initial selection for the model parameters, which we denote by θ^{old} . In the E step, we take these parameter values and find the posterior distribution of the latent variables $p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old})$. We then use this posterior distribution to evaluate the expectation of the logarithm of the complete-sequence data likelihood function (4), as a function of the parameters θ , to give the function $Q(\theta, \theta^{old})$ defined by:

$$\begin{aligned} Q(\theta, \theta^{old}) &= \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) \ln p(\mathbf{X}, \mathbf{Z}^*, \mathbf{Z}; \theta) \\ Q(\theta, \theta^{old}) &= \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) \ln p(\mathbf{Z}^*; \pi, \mathbf{A}) \\ &+ \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) \ln p(\mathbf{X}/\mathbf{Z}; \phi) \\ &+ \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) \ln p(\mathbf{Z}/\mathbf{Z}^*) \\ Q(\theta, \theta^{old}) &= Q_1(\pi, \theta^{old}) + Q_2(\mathbf{A}, \theta^{old}) \\ &+ Q_3(\phi, \theta^{old}) + Q_4 \end{aligned} \quad (5)$$

where

$$Q_1(\pi, \theta^{old}) = \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} \sum_{k=1}^K p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) z_{1k}^* \ln \pi_k$$

$$Q_2(\mathbf{A}, \theta^{old}) = \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} \sum_{n=2}^N \sum_{k=1}^K \sum_{j=1}^K p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) z_{n-1,j}^* z_{n,k}^* \ln(A_{jk})$$

$$Q_3(\phi, \theta^{old}) = \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} \sum_{n=1}^N \sum_{k=1}^K p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) z_{nk} \ln(p(\mathbf{x}_n; \phi_k))$$

$$Q_4 = \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) \ln p(\mathbf{Z}/\mathbf{Z}^*)$$

At this point, we introduce some notation. We will use $\gamma(\mathbf{z}_n^*, \mathbf{z}_n)$ to denote the marginal posterior distribution of a latent variable \mathbf{z}_n^* and \mathbf{z}_n , and $\xi(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*) = p(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*/\mathbf{X}, \theta^{old})$ to denote the joint posterior distribution of successive latent variables, so that

$$\gamma(\mathbf{z}_n^*, \mathbf{z}_n) = p(\mathbf{z}_n^*, \mathbf{z}_n|\mathbf{X}; \theta^{old})$$

thus

$$\gamma(\mathbf{z}_n^*) = \sum_{\mathbf{z}} p(\mathbf{z}_n^*, \mathbf{z}_n|\mathbf{X}; \theta^{old})$$

$$\gamma(\mathbf{z}_n) = \sum_{\mathbf{z}^*} p(\mathbf{z}_n^*, \mathbf{z}_n|\mathbf{X}; \theta^{old})$$

$$\begin{aligned} \gamma(z_{nk}^*) &= \mathbf{E}[z_{nk}^*] \\ &= \sum_{\mathbf{z}^*} \sum_{\mathbf{z}} \gamma(\mathbf{z}_n^*, \mathbf{z}_n) z_{nk}^* \\ &= \sum_{\mathbf{z}^*} \gamma(\mathbf{z}_n^*) z_{nk}^* \end{aligned}$$

We observe that the objective function (5) $Q(\theta, \theta^{old})$ is defined as a sum of four terms. The first term $Q_1(\pi, \theta^{old})$ depends on initial probabilities; the second term $Q_2(\mathbf{A}, \theta^{old})$ depends on transition probabilities \mathbf{A} ; the third term $Q_3(\phi, \theta^{old})$ depends on ϕ , and the fourth term is constant. Maximizing $Q(\theta, \theta^{old})$ with respect to $\theta = \{\pi, \mathbf{A}, \phi\}$ can be performed separately.

1) Maximization of $Q_1(\pi, \theta^{old})$: Initial probabilities:

$$\begin{aligned} Q_1(\pi, \theta^{old}) &= \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} \sum_{k=1}^K p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) z_{1k}^* \ln \pi_k \\ &= \sum_{\mathbf{Z}^*} \sum_{k=1}^K p(\mathbf{Z}^*/\mathbf{X}; \theta^{old}) z_{1k}^* \ln \pi_k \\ &= \sum_{k=1}^K \gamma(z_{1k}^*) \ln \pi_k \end{aligned}$$

The update parameter is computed as follows:

$$\pi_k = \frac{\gamma(z_{1k}^*)}{\sum_{j=1}^K \gamma(z_{1j}^*)} \quad (6)$$

2) Maximization of $Q_2(\mathbf{A}, \theta^{old})$: Probability transitions :

$$\begin{aligned} Q_2(\mathbf{A}, \theta^{old}) &= \\ \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} \sum_{n=2}^N \sum_{k=1}^K \sum_{j=1}^K p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) z_{n-1,j}^* z_n^* \ln(A_{jk}) \\ &= \sum_{\mathbf{Z}^*} \sum_{n=2}^N \sum_{k=1}^K \sum_{j=1}^K p(\mathbf{Z}^*/\mathbf{X}; \theta^{old}) z_{n-1,j}^* z_n^* \ln(A_{jk}) \\ &= \sum_{n=2}^N \sum_{k=1}^K \sum_{j=1}^K \xi(z_{n-1,j}^*, z_n^*) \ln(A_{jk}) \end{aligned}$$

The update parameter is computed as follows:

$$A_{jk} = \frac{\sum_{n=2}^N \xi(z_{n-1,j}^*, z_{nk}^*)}{\sum_{l=1}^K \sum_{n=2}^N \xi(z_{n-1,j}^*, z_{nl}^*)} \quad (7)$$

where

$$\xi(z_{n-1,j}^*, z_{nk}^*) = \mathbf{E}[z_{n-1,j}^* z_{nk}^*] = \sum_{\mathbf{z}^*} \gamma(\mathbf{z}^*) z_{n-1,j}^* z_{nk}^*$$

3) Maximization of $Q_3(\phi, \theta^{old})$: Emission probabilities:

$$\begin{aligned} Q_3(\phi, \theta^{old}) &= \\ \sum_{\mathbf{Z}^*} \sum_{\mathbf{Z}} \sum_{n=1}^N \sum_{k=1}^K p(\mathbf{Z}^*, \mathbf{Z}/\mathbf{X}; \theta^{old}) z_{nk} \ln p(\mathbf{x}_n; \phi_k) \\ &= \sum_{\mathbf{Z}} \sum_{n=1}^N \sum_{k=1}^K p(\mathbf{Z}/\mathbf{X}; \theta^{old}) z_{nk} \ln p(\mathbf{x}_n; \phi_k) \\ &= \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \ln p(\mathbf{x}_n; \phi_k) \end{aligned}$$

In the case of spherical Gaussian emission densities we have $p(\mathbf{x}/\phi_k) = \mathcal{N}(\mathbf{x}; \mathbf{w}_k, \sigma_k)$, defined by its "mean" \mathbf{w}_k , which have the same dimension as input data, and its covariance matrix, defined by $\sigma_k^2 \mathbf{I}$ where σ_k is the standard deviation and \mathbf{I} is the identity matrix. The maximization of the function $Q_3(\phi, \theta^{old})$ provides:

$$\mathbf{w}_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n}{\sum_{n=1}^N \gamma(z_{nk})} \quad (8)$$

$$\sigma_k^2 = \frac{\sum_{n=1}^N \gamma(z_{nk}) \|\mathbf{x}_n - \mathbf{w}_k\|^2}{d \sum_{n=1}^N \gamma(z_{nk})} \quad (9)$$

where d is the dimension of the element \mathbf{x} .

The EM algorithm requires initial values for the parameters of the emission distribution. One way to set these is first to treat the data initially as i.i.d. and fit the emission density by maximum likelihood, and then use the resulting values to initialize the parameters for EM.

C. The forward-backward algorithm: E-step

Next we seek an efficient procedure for evaluating the quantities $\gamma(\mathbf{z}_n^*)$, $\gamma(\mathbf{z}_n)$ and $\xi(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*)$, corresponding to the E step of the EM algorithm. In the particular context of the hidden Markov model, this is known as the forward-backward

algorithm [21], or the Baum-Welch algorithm [22], [23]. In our case it can be renamed topological forward-backward algorithm, because we use the graph structure to organize the sequential data. Some formula are similar if we don't use the graph structure. We will use the notations $\alpha(z_{nk}^*)$ and $\alpha(z_{nk})$ to denote the value of $\alpha(\mathbf{z}^*)$ and $\alpha(\mathbf{z})$ when $z_{nk}^* = 1$, $z_{nk} = 1$ with an analogous notations of β .

$$\begin{aligned} \gamma(\mathbf{z}_n^*) &= p(\mathbf{z}_n^*/\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{z}_n^*)p(\mathbf{z}_n^*)}{p(\mathbf{X})} \\ &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}_n^*)p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N/\mathbf{z}_n^*)}{p(\mathbf{X})} \\ \gamma(\mathbf{z}_n) &= \frac{\alpha(\mathbf{z}_n^*)\beta(\mathbf{z}_n^*)}{p(\mathbf{X})} \end{aligned}$$

Using the similar decomposition we obtain

$$\gamma(\mathbf{z}_n) = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

The values of $\alpha(\mathbf{z}_n^*)$ and $\alpha(\mathbf{z}_n)$ are calculated by forward recursion as follows:

$$\begin{aligned} \alpha(\mathbf{z}_n^*) &= \left[\sum_{\mathbf{z}} p(\mathbf{x}_n/\mathbf{z}_n) p(\mathbf{z}_n/\mathbf{z}_n^*) \right] \\ &\times \sum_{\mathbf{z}_{n-1}^*} \alpha(\mathbf{z}_{n-1}^*) p(\mathbf{z}_n^*|\mathbf{z}_{n-1}^*) \end{aligned} \quad (10)$$

and

$$\begin{aligned} \alpha(\mathbf{z}_n) &= p(\mathbf{x}_n|\mathbf{z}_n) \sum_{\mathbf{z}_n^*} p(\mathbf{z}_n/\mathbf{z}_n^*) \\ &\left[\sum_{\mathbf{z}_{n-1}^*} \alpha(\mathbf{z}_{n-1}^*) p(\mathbf{z}_n^*|\mathbf{z}_{n-1}^*) \sum_{\mathbf{z}_{n-1}} p(\mathbf{z}_{n-1}|\mathbf{z}_{n-1}^*) \right] \end{aligned} \quad (11)$$

where $p(\mathbf{z}_n/\mathbf{z}_n^*) = p(z_{nc} = 1/z_{nc}^* = 1) = \frac{K^T(\delta(c,c^*))}{\sum_{r \in \mathcal{C}} K^T(\delta(r,c^*))}$. To start this recursion, we need an initial condition that is given by

$$\begin{aligned} \alpha(\mathbf{z}_1^*) &= p(\mathbf{x}_1, \mathbf{z}_1^*) = p(\mathbf{z}_1^*) \left[\sum_{\mathbf{z}_1} p(\mathbf{x}_1/\mathbf{z}_1) p(\mathbf{z}_1/\mathbf{z}_1^*) \right] \\ \alpha(\mathbf{z}_1) &= p(\mathbf{x}_1, \mathbf{z}_1) = p(\mathbf{x}_1/\mathbf{z}_1) \left[\sum_{\mathbf{z}_1^*} p(\mathbf{z}_1^*) p(\mathbf{z}_1/\mathbf{z}_1^*) \right] \end{aligned}$$

The value of $\beta(\mathbf{z}_n^*)$, are calculated by backward recursion as follows:

$$\beta(\mathbf{z}_n^*) = \sum_{\mathbf{z}_{n+1}^*} \beta(\mathbf{z}_{n+1}^*) p(\mathbf{x}_{n+1}/\mathbf{z}_{n+1}^*) p(\mathbf{z}_{n+1}^*/\mathbf{z}_n^*) \quad (12)$$

$$\begin{aligned} \beta(\mathbf{z}_n) &= \frac{1}{p(\mathbf{z}_n)} \sum_{\mathbf{z}_n^*} p(\mathbf{z}_n^*) p(\mathbf{z}_n/\mathbf{z}_n^*) \sum_{\mathbf{z}_{n+1}} \sum_{\mathbf{z}_{n+1}^*} \\ &p(\mathbf{z}_{n+1}/\mathbf{z}_{n+1}^*) \beta(\mathbf{z}_{n+1}^*) p(\mathbf{x}_{n+1}/\mathbf{z}_{n+1}^*) p(\mathbf{z}_{n+1}^*/\mathbf{z}_n^*) \end{aligned} \quad (13)$$

where

$$p(\mathbf{x}_{n+1}/\mathbf{z}_{n+1}^*) = \left[\sum_{\mathbf{z}} p(\mathbf{x}_{n+1}/\mathbf{z}_{n+1}) p(\mathbf{z}_{n+1}/\mathbf{z}_{n+1}^*) \right]$$

$$p(\mathbf{z}_n) = \sum_{\mathbf{z}_n^*} p(\mathbf{z}_n^*)p(\mathbf{z}_n/\mathbf{z}_n^*)$$

and

$$\begin{aligned} p(\mathbf{z}_{n+1}/\mathbf{z}_{n+1}^*) &= p(z_{n+1,c} = 1/z_{n+1,c}^* = 1) \\ &= \frac{K^T(\delta(c, c^*))}{\sum_{r \in C} K^T(\delta(r, c^*))} \end{aligned}$$

Again we need a starting condition for the recursion, a value for $\beta(\mathbf{z}_N^*) = 1$ and $\beta(\mathbf{z}_N) = 1$. This can be obtained by setting $n = N$ in (expression 10).

Next we consider the evaluation of the quantities $\xi(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*)$ which correspond to the values of the conditional probabilities $p(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*/\mathbf{X})$ for each of the $K \times K$ settings for $(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*)$. Using the applying Bayes theorem, we obtain

$$\begin{aligned} \xi(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*) &= \frac{p(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*/\mathbf{X})}{p(\mathbf{X})} \\ &= \frac{p(\mathbf{X}/\mathbf{z}_{n-1}^*, \mathbf{z}_n^*)p(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*)}{p(\mathbf{X})} \\ \xi(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*) &= \frac{\alpha(\mathbf{z}_{n-1}^*) [\sum_{\mathbf{z}} p(\mathbf{x}_n/\mathbf{z}_n)p(\mathbf{z}_n/\mathbf{z}_n^*)]}{p(\mathbf{X})} \\ &\times \frac{p(\mathbf{z}_n^*/\mathbf{z}_{n-1}^*)\beta(\mathbf{z}_n^*)}{p(\mathbf{X})} \end{aligned}$$

If we sum both sides of $\alpha(\mathbf{z}^*)$ over \mathbf{z}_N , we obtain $p(\mathbf{X}) = \sum_{\mathbf{z}_N} \alpha(\mathbf{z}_N)$. Then we compute the forward α recursion and the backward β recursion and use the results to evaluate γ and $\xi(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*)$. We use these results to compute a new parameter model θ using the M-step equations (6, 7, 8, 9). These both steps are repeated until some convergence criterion is satisfied.

III. DISCUSSION ABOUT TOPOLOGICAL MARKOV MODEL ORGANIZATION

The 3M-SOM model allows us to estimate the parameters maximizing the log-likelihood function for a fixed T . As in the topological clustering algorithm, we have to decrease the value of T between two values T_{max} and T_{min} , to control the size of the neighbourhood influencing a given state of HMM on the graph (grid) and at same time. For each T value, we get a likelihood function Q^T , and therefore the expression varies with T . When decreasing T , the model of 3M-SOM will be defined in the following way:

- The first step corresponds to high T values. In this case, the influencing neighbourhood of each state \mathbf{z}^* on the HMM graph (grid) is important and corresponds to higher values of $K^T(\delta(c, r))$. Formulas use a high number of state and hence high number of observations to estimate model parameters. This step provides the topological order of Markov model.
- The second step corresponds to small T values. The number of observations in formulas is limited. Therefore, the adaptation is very local. The parameters are accurately computed from the local density of the data. In this case we can consider that we converge to traditional HMM (without using neighborhood). Recall that clustering based on mixture model for i.i.d. observations is a special case of the HMM [19, chap 9].

IV. CONCLUSION

In this paper, we presented an original model that could be applied to more advanced/complex data set (not i.i.d observations, time series). We provides here the mathematical formulation of our model. We present one way to estimate the parameter using EM algorithm with Baum-Welch algorithm. Visualization techniques and refined graphic displays can be developed to illustrate the power of 3M-SOM to explore the not i.i.d data. As has been stressed, the 3M-SOM unsupervised topographic learning algorithm is purely batch learning. An extension to an on-line mode version is quite straightforward. Finally, providing an equivalent to the 3M-SOM for applications requiring Bernoulli emission probability density functions should be interesting task.

REFERENCES

- [1] Yoshua Bengio and Paolo Frasconi. An input output hmm architecture. In *NIPS*, pages 427–434, 1994.
- [2] Samy Bengio and Yoshua Bengio. An EM algorithm for asynchronous input/output hidden Markov models. In L. Xu, editor, *International Conference On Neural Information Processing*, pages 328–334, Hong-Kong, 1996.
- [3] Zoubin Ghahramani and Michael I. Jordan. Factorial hidden markov models. *Mach. Learn.*, 29(2-3):245–273, 1997.
- [4] Herve A. Boullard and Nelson Morgan. *Connectionist Speech Recognition: A Hybrid Approach*. Kluwer Academic Publishers, Norwell, MA, USA, 1993.
- [5] Marina Meila and Michael I. Jordan. Learning fine motion by markov mixtures of experts. Technical report, Cambridge, MA, USA, 1995.
- [6] Catherine Recanati, Nicoleta Rogovschi, and Younès Bennani. The structure of verbal sequences analyzed with unsupervised learning techniques. In *LTC'07, Language Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics. October 5-7, 2007, Poznan, Poland*, 2007.
- [7] Djamel Bouchaffra and Jun Tan. Structural hidden markov model and its application in automotive industry. In *ICEIS (2)*, pages 155–164, 2003.
- [8] D. Bouchaffra and J. Tan. Structural hidden markov models: An application to handwritten numeral recognition. *Intell. Data Anal.*, 10(1):67–79, 2006.
- [9] Djamel Bouchaffra. Embedding hmm's-based models in a euclidean space: The topological hidden markov models. In *ICPR08*, pages 1–4, 2008.
- [10] Panu Somervuo and Teuvo Kohonen. Self-organizing maps and learning vector quantization for feature sequences. *Neural Process. Lett.*, 10(2):151–159, 1999.
- [11] Panu Somervuo. Competing hidden markov models on the self-organizing map. *Neural Networks, IEEE - INNS - ENNS International Joint Conference on*, 3:3169, 2000.
- [12] Khalid Benabdeslem Arnaud Zeboulon, Youns Bennani. Hybrid connectionist approach for knowledge discovery from web navigation patterns. In *ACS/IEEE International Conference on Computer Systems and Applications*, pages 118–122, 2003.
- [13] Christos Ferles and Andreas Stafylopatis. A hybrid self-organizing model for sequence analysis. In *ICTAI '08: Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence*, pages 105–112, Washington, DC, USA, 2008. IEEE Computer Society.
- [14] C. Ferles and A. Stafylopatis. Sequence clustering with the self-organizing hidden markov model map. *Bioinformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*, pages 1–7, Oct. 2008.
- [15] T. Kohonen. *Self-organizing Maps*. Springer Berlin, 2001.
- [16] Fatiha Anouar, Fouad Badran, and Sylvie Thiria. Self-organizing map, a probabilistic approach. In *Proceedings of WSOM'97-Workshop on Self-Organizing Maps, Espoo, Finland June 4-6*, pages 339–344, 1997.
- [17] C. M. Bishop, M. Svensén, and C. K. I. Williams. GTM: The generative topographic mapping. *Neural Comput.*, 10(1):215–234, 1998.
- [18] Mustapha Lebbah, Nicoleta Rogovschi, and Younès Bennani. Besom : Bernoulli on self organizing map. In *International Joint Conferences on Neural Networks. IJCNN 2007, Orlando, Florida, August 12-17, 2007*.

- [19] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), 2006.
- [20] S. P. Luttrel. A bayesian analysis of self-organizing maps. *Neural Computing*, 6:767 – 794, 1994.
- [21] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [22] L. E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3:1–8, 1972.
- [23] Edoardo M Airoidi. Getting started in probabilistic graphical models. *PLoS Comput Biol*, 3(12):e252, 12 2007.

Statistics on Graphs, Exponential Formula and Combinatorial Physics

G erard H. E. Duchamp, Laurent Poinot, Silvia Goodenough and Karol A. Penson

Abstract—This paper concerns a famous combinatorial formula known as the “exponential formula” which occurs quite naturally in many physical contexts. Its meaning is the following: the exponential generating function of a whole structure is equal to the exponential of those of connected substructures. Keeping this descriptive statement as a guideline, we develop a general framework to handle many different situations in which the exponential formula can be applied.

Index Terms—Combinatorial physics, Exponential generating function, Partial semigroup, Experimental mathematics.

I. INTRODUCTION

Applying the exponential paradigm one can feel sometimes uncomfortable wondering whether “one has the right” to do so (as for example for coloured structures). The following paper is aimed at giving a rather large framework where this formula holds.

Exponential formula can be traced back to works by Touchard and Ridell & Uhlenbeck [16], [13]. For an other exposition, see for example [2], [4], [7], [15].

We are interested to compute various examples of EGF for combinatorial objects having (a finite set of) nodes (i.e. their set-theoretical support) so we use as central concept the mapping σ which associates to every structure, its set of (labels of its) nodes.

We need to draw what could be called “square-free decomposable objects” (SFD). This version is suited to our needs for the “exponential formula” and it is sufficiently general to contain, as a particular case, the case of multivariate series.

II. PARTIAL SEMIGROUPS

Let us call partial semigroup a semigroup with a partially defined associative law (see for instance [3] for usual semigroups and [1], [11], [14] for more details on structures with a partially defined binary operation). More precisely, a *partial semigroup* is a pair $(S, *)$ where S is a set and $*$ is a (partially defined) function $S \times S \rightarrow S$ such that the two (again partially defined) functions $S \times S \times S \rightarrow S$

$$(x, y, z) \mapsto (x * y) * z \text{ and } (x, y, z) \mapsto x * (y * z) \quad (1)$$

Manuscript received April 06, 2009. This work was supported by the French Ministry of Science and Higher Education under Grant ANR PhysComb.

G. H. E. Duchamp, L. Poinot and S. Goodenough are affiliated to Laboratoire d’Informatique Paris Nord, Universit  Paris-Nord 13, CNRS UMR 7030, 99 av. J.-B. Cl ment, F 93430 Villetaneuse, France (emails: {ghed, laurent.poinot}@lipn-univ.paris13.fr, goodenou@iutv.univ-paris13.fr).

K. A. Penson is affiliated to Laboratoire de Physique Th orique de la Mati re Condens e, Universit  Pierre et Marie Curie, CNRS UMR 7600, Tour 24 - 2e  t., 4 pl. Jussieu, F 75252 Paris cedex 05, France (email: penson@lptl.jussieu.fr).

coincide (same domain and values). Using this requirement one can see that the values of the (partially defined) functions $S^n \rightarrow S$

$$(x_1, \dots, x_n) \mapsto E_T(x_1, \dots, x_n) \quad (2)$$

obtained by evaluating the expression formed by labelling by x_i (from left to right) the i th leaf of a binary tree T with n nodes and by $*$ its internal nodes, is independant of T . We will denote $x_1 * \dots * x_n$ their common value. In this paper we restrict our attention to *commutative* semigroups. By this we mean that the value $x_1 * \dots * x_n$ does not depend on the relative order of the x_i . A nonempty partial semigroup $(S, *)$ has a (*two-sided and total*) unit $\epsilon \in S$ if, and only if, for every $\omega \in S$, $\omega * \epsilon = \omega = \epsilon * \omega$. Using associativity of $*$, it can be easily checked that if S has a unit, then it is unique.

Example 2.1: Let F be a set of sets (resp. which contains \emptyset as an element) and which is closed under the disjoint sum \sqcup , i.e., if $A, B \in F$ such that $A \cap B = \emptyset$, then $A \cup B (= A \sqcup B) \in F$. Then (F, \sqcup) is a partial semigroup (resp. partial semigroup with unit).

III. SQUARE-FREE DECOMPOSABLE PARTIAL SEMIGROUPS

Let $2^{\mathbb{N}^+}$ be the set of all finite subsets of the positive integers \mathbb{N}^+ and (S, \oplus) be a partial semigroup with unit equipped with a mapping $\sigma : S \rightarrow 2^{\mathbb{N}^+}$, called the (*set-theoretic*) support mapping. Let D be the domain of the \oplus . The triple (S, \oplus, σ) is called *square-free decomposable* (SFD) if, and only if, it fulfills the two following conditions.

- *Direct sum* (DS):

- 1) $\sigma(\omega) = \emptyset$ iff $\omega = \epsilon$;
- 2) $D = \{(\omega_1, \omega_2) \in S^2 : \sigma(\omega_1) \cap \sigma(\omega_2) = \emptyset\}$;
- 3) For all $\omega_1, \omega_2 \in S$, if $(\omega_1, \omega_2) \in D$ then $\sigma(\omega_1 \oplus \omega_2) = \sigma(\omega_1) \cup \sigma(\omega_2)$.

- *Levi’s property* (LP): For every $\omega_1, \omega_2, \omega^1, \omega^2 \in S$ such that $(\omega_1, \omega_2), (\omega^1, \omega^2) \in D$ and $\omega_1 \oplus \omega_2 = \omega^1 \oplus \omega^2$, there are $\omega_i^j \in S$ for $i = 1, 2, j = 1, 2$ such that $(\omega_i^1, \omega_i^2), (\omega_1^j, \omega_2^j) \in D$, $\omega_i = \omega_i^1 \oplus \omega_i^2$ and $\omega^j = \omega_1^j \oplus \omega_2^j$ for $i = 1, 2$ and $j = 1, 2$.

Note 3.1: The second and third conditions of (DS) imply that $\sigma(\omega_1 \oplus \omega_2) = \sigma(\omega_1) \sqcup \sigma(\omega_2)$ whenever $(\omega_1, \omega_2) \in D$ (which means that $\sigma(\omega_1) \cap \sigma(\omega_2) = \emptyset$), where \sqcup denotes the disjoint sum.

Example 3.2: As example of this setting we have:

- 1) The positive square-free integers, $\sigma(n)$ being the set of primes which divide n , the atoms being the prime numbers.

- 2) All the positive integers ($S = \mathbb{N}^+$), under the usual integer multiplication, $\sigma(n)$ being the set of primes which divides n .
- 3) Graphs, hypergraphs, (finitely) coloured, weighted graphs, with nodes in \mathbb{N}^+ , $\sigma(G)$ being the set of nodes and \oplus the juxtaposition (direct sum) when the set of nodes are mutually disjoint.
- 4) The set of endofunctions $f : F \rightarrow F$ where F is a finite subset of \mathbb{N}^+ .
- 5) The (multivariate) polynomials in $\mathbb{N}[X]$, $X = \{x_i : i \in I\}$, with $I \subseteq \mathbb{N}^+$, being a nonempty set of (commuting or not) variables, with $\sigma(P) = \text{Alph}(P)$ the set of indices of variables that occur in a polynomial P , and $\oplus = +$.
- 6) For a given finite or denumerable field, the set of irreducible monic polynomials is denumerable. Arrange them in a sequence $(P_n)_{n \in \mathbb{N}^+}$, then the square-free monic (for a given order on the variables) polynomials is SFD, $\sigma(P) := \{n \in \mathbb{N}^+ : P_n \text{ divides } P\}$ and \oplus being the multiplication.
- 7) Rational complex algebraic curves; $\sigma(V)$ being the set of monic irreducible bivariate polynomials vanishing on V .

In what follows we write $\oplus_{i=1}^n \omega_i$ instead of $\omega_1 \oplus \dots \oplus \omega_n$ (if $n = 0$, then $\oplus_{i=1}^n \omega_i = \epsilon$) and we suppose that (S, \oplus, σ) is SFD for the two following easy lemmas.

Lemma 3.3: Let $\omega_1, \dots, \omega_n \in S$ such that $\oplus_{i=1}^n \omega_i$ is defined. Then for every $i, j \in \{1, \dots, n\}$ such that $i \neq j$, it holds that $\sigma(\omega_i) \cap \sigma(\omega_j) = \emptyset$. In particular, if none ω_k is equal to ϵ , then $\omega_i \neq \omega_j$ for every $i, j \in \{1, \dots, n\}$ such that $i \neq j$. Moreover $\sigma(\oplus_{i=1}^n \omega_i) = \bigcup_{i=1}^n \sigma(\omega_i)$.

Lemma 3.4: Let $(\omega_i)_{i=1}^n$ be a finite family of elements of S with pairwise disjoint supports. Suppose that for $i = 1, \dots, n$, $\omega_i = \oplus_{k=1}^{n_i} \omega_i^k$, where $(\omega_i^k)_{k=1}^{n_i}$ is a finite family of elements of S . Then $\oplus_{i=1}^n \omega_i = \oplus_{i=1}^n (\oplus_{k=1}^{n_i} \omega_i^k)$.

These lemmas are useful to define the sum of two or more elements of S using respective sum decompositions.

Now, an *atom* in a partial semigroup with unit S is any object $\omega \neq \epsilon$ which cannot be split, formally

$$\omega = \omega_1 \oplus \omega_2 \implies \epsilon \in \{\omega_1, \omega_2\} . \quad (3)$$

The set of all atoms is denoted by $\text{atoms}(S)$. Whenever the square-free decomposable semigroup S is not *trivial*, i.e., reduced to $\{\epsilon\}$, $\text{atoms}(S)$ is not empty.

Example 3.5: The atoms obtained from examples 3.2:

- 1) The atoms of 3.2.2 are the primes.
- 2) The atoms of 3.2.3 are connected graphs.
- 3) The atoms of 3.2.4 are the endofunctions for which the domain is a singleton.
- 4) The atoms of 3.2.5 are the monomials.

The prescriptions (DS,LP) imply that decomposition of objects into atoms always exists and is unique.

Proposition 3.6: Let (S, \oplus, σ) be SFD. For each $\omega \in S$ there is one and only one finite set of atoms $A = \{\omega_1, \dots, \omega_n\}$ such that $\omega = \oplus_{i=1}^n \omega_i$. One has $A = \emptyset$ iff $\omega = \epsilon$.

IV. EXPONENTIAL FORMULA

In this section we consider (S, \oplus, σ) as a square-free decomposable partial semigroup with unit.

In the set S , objects are conceived to be “measured” by different parameters (data in statistical language). So, to get a general purpose tool, we suppose that the statistics takes its values in a (unitary) ring R of characteristic zero that is to say which contains \mathbb{Q} (as, to write exponential generating series it is convenient to have no trouble with the fractions $\frac{1}{n!}$). Let then $c : S \rightarrow R$ be the given statistics. For F a finite set and each $X \subseteq S$, we define

$$X_F := \{\omega \in X : \sigma(\omega) = F\} . \quad (4)$$

In order to write generating series, we need

- 1) that the sums $c(X_F) := \sum_{\omega \in X_F} c(\omega)$ exist for every finite set F of \mathbb{N}^+ and every $X \subseteq S$;
- 2) that $F \rightarrow c(X_F)$ would depend only of the cardinality of the finite set F of \mathbb{N}^+ , for each fixed $X \subseteq S$;
- 3) that $c(\omega_1 \oplus \omega_2) = c(\omega_1).c(\omega_2)$.

We formalize it in

(LF) *Local finiteness.* — For each finite set F of \mathbb{N}^+ , the subset S_F of S is a finite set.

(Eq) *Equivariance.* —

$$\text{card}(F_1) = \text{card}(F_2) \implies c(\text{atoms}(S)_{F_1}) = c(\text{atoms}(S)_{F_2}) . \quad (5)$$

(Mu) *Multiplicativity.* —

$$c(\omega_1 \oplus \omega_2) = c(\omega_1).c(\omega_2) . \quad (6)$$

Note 4.1: a) In fact, (LF) is a property of the set S , while (Eq) is a property of the statistics. In practice, we choose S which is locally finite and choose equivariant statistics for instance

$$c(\omega) = x^{(\text{number of cycles})} y^{(\text{number of fixed points})}$$

for some variables x, y .

b) More generally, it is typical to take integer-valued partial (additive) statistics $c_1, \dots, c_i, \dots, c_r$ (for every $\omega \in S$, $c_i(\omega) \in \mathbb{N}$) and set $c(\omega) = x_1^{c_1(\omega)} x_2^{c_2(\omega)} \dots x_r^{c_r(\omega)}$.

c) The set of example 3.2.2 is not locally finite, but other examples satisfy (LF): for instance 3.2.3 if one asks that the number of arrows and weight is finite, 3.2.1.

A multiplicative statistics is called *proper* if $c(\epsilon) \neq 0$. It is called *improper* if $c(\epsilon) = 0$. In this case, for every $\omega \in S$, $c(\omega) = 0$. Indeed $c(\omega) = c(\omega \oplus \epsilon) = c(\omega)c(\epsilon) = 0$.

If R is an integral domain and if c is proper, then $c(\epsilon) = 1$ because $c(\epsilon) = c(\epsilon \oplus \epsilon) = c(\epsilon)^2$, therefore $1 = c(\epsilon)$. Note that for each $X \subseteq S$,

$$c(X_\emptyset) = \sum_{\omega \in X_\emptyset} c(\omega) = \begin{cases} c(\epsilon) & \text{if } \epsilon \in X \\ 0 & \text{if } \epsilon \notin X \end{cases} . \text{ For every}$$

finite subset X of S , we also define $c(X) := \sum_{\omega \in X} c(\omega)$, then

we have in particular $c(\emptyset) = 0$ (different from $c(S_\emptyset) = c(\{\epsilon\})$ if c is proper). The requirement (LF) implies that for every $X \subseteq S$ and every finite set F of \mathbb{N}^+ , $c(X_F)$ is defined as a sum of a finite number of terms because $X_F \subseteq S_F$, and therefore X_F is finite.

Now, we are in position to state the exponential formula as it will be used throughout the paper. Let us recall the usual *exponential formula* for formal power series in $R[[z]]$ (see [10], [15] for more details on formal power series). Let $f(z) = \sum_{n \geq 1} f_n z^n$. Then we have

$$e^f = \sum_{n \geq 0} a_n \frac{z^n}{n!} \quad (7)$$

where

$$a_n = \sum_{\pi \in \Pi_n} \prod_{p \in \pi} f_{\text{card}(p)} \quad (8)$$

with Π_n being the set of all partitions of $[1..n]$ (in particular for $n = 0$, $a_0 = 1$) and $e^z = \sum_{n \geq 0} \frac{z^n}{n!} \in R[[z]]$.

In what follows $[1..n]$ denotes the interval $\{j \in \mathbb{N}^+ : 1 \leq j \leq n\}$, reduced to \emptyset when $n = 0$. Let (S, \oplus, σ) be locally finite SFD and c be a multiplicative equivariant statistics. For every subset X of S one sets the following *exponential generating series*

$$\mathbf{EGF}(X; z) = \sum_{n=0}^{\infty} c(X_{[1..n]}) \frac{z^n}{n!}. \quad (9)$$

Theorem 4.2 (exponential formula): Let S be a locally finite SFD and c be a multiplicative equivariant statistics. We have

$$\mathbf{EGF}(S; z) = c(\epsilon) - 1 + e^{\mathbf{EGF}(\text{atoms}(S); z)}. \quad (10)$$

In particular if $c(\epsilon) = 1$ (for instance if c is proper and R is an integral domain),

$$\mathbf{EGF}(S; z) = e^{\mathbf{EGF}(\text{atoms}(S); z)}. \quad (11)$$

Proof — Let $n = 0$. Then the unique element of S_\emptyset is ϵ . Therefore $c(S_\emptyset) = c(\epsilon)$. Now suppose that $n > 0$ and let $\omega \in S_{[1..n]}$. According to proposition 3.6, there is a unique finite set $\{\alpha_1, \dots, \alpha_k\} \subseteq \text{atoms}(S)$ such that $\omega = \bigoplus_{i=1}^k \alpha_i$. By lemma 3.3, $\{\sigma(\alpha_i) : 1 \leq i \leq k\}$ is a partition of $[1..n]$ into k blocks. Therefore $\omega \in \text{atoms}(S)_{P_1} \oplus \dots \oplus \text{atoms}(S)_{P_k}$ where $P_i = \sigma(\alpha_i)$ for $i = 1, \dots, k$. We can remark that $\alpha_1 \oplus \dots \oplus \alpha_k$ is well-defined for each $(\alpha_1, \dots, \alpha_k) \in \text{atoms}(S)_{P_1} \times \dots \times \text{atoms}(S)_{P_k}$ since the supports are disjoint. Now, one has, thanks to the partitions of $[1..n]$

$$S_{[1..n]} = \bigsqcup_{\pi \in \Pi_n} \bigoplus_{p \in \pi} \text{atoms}(S)_p \quad (12)$$

$$c(S_{[1..n]}) = \sum_{\pi \in \Pi_n} \prod_{p \in \pi} c(\text{atoms}(S)_p) \quad (13)$$

as, for disjoint (finite) sets F and G of \mathbb{N}^+ , it is easy to check that $c(X_F \oplus X_G) = c(X_F)c(X_G)$ for every $X \subseteq S$ and because the disjoint union as only a finite number of

factors. Therefore due to equivariance of c on sets of the form $\text{atoms}(S)_F$, one has

$$c(S_{[1..n]}) = \sum_{\pi \in \Pi_n} \prod_{p \in \pi} c(\text{atoms}(S)_{[1..\text{card}(p)]}). \quad (14)$$

But $c(\text{atoms}(S)_{[1..\text{card}(p)]})$ is the $\text{card}(p)$ th coefficient of the series $\mathbf{EGF}(\text{atoms}(S); z)$. Therefore due to the usual exponential formula, $\mathbf{EGF}(S; z) = c(\epsilon) - 1 + e^{\mathbf{EGF}(\text{atoms}(S); z)}$. Now if $c(\epsilon) = 1$, then we obtain $\mathbf{EGF}(S; z) = e^{\mathbf{EGF}(\text{atoms}(S); z)}$. \square

V. TWO EXAMPLES

The examples provided here pertain to the class of labelled graphs where the ‘‘classic’’ exponential formula applies, namely *Burnside’s Classes*¹ $Burn_{a,b}$, defined, for $0 \leq a < b$ two integers, as the class of graphs of numeric endofunctions f such that

$$f^a = f^b \quad (15)$$

where f^n denotes the n th power with respect to functional composition. Despite of its simplicity, there are still (enumerative combinatorial) open problems for this class and only $B_{1,\ell+1}$ gives rise to an elegant formula [5], [15] (see also [8], for the idempotent case: $\ell = 1$ and compare to exact but non-easily tractable formulas in [2] for the general case in the symmetric semigroup, and in [9] for their generalization to the wreath product of the symmetric semigroup and a finite group).

The second example: the class of finite partitions which can be (and should here) identified as graphs of equivalence relations on finite subsets $F \subseteq \mathbb{N}^+$. Call this class ‘‘Stirling class’’ as the number of such graphs with support $[1..n]$ and k connected components is exactly the Stirling number of the second kind $S_2(n, k)$ and, using the statistics $x^{(\text{number of points})}y^{(\text{number of connected components})}$, one obtains

$$\sum_{n,k \geq 0} S_2(n, k) \frac{x^n}{n!} y^k = e^{y(e^x - 1)}. \quad (16)$$

Examples of this kind bring us to the conclusion that bivariate statistics like $Burn_{a,b}(n, k)$, $S_2(n, k)$ or $S_1(n, k)$ (Stirling numbers of the second and first kind) are better understood through the notion of one-parameter group, conversely such groups naturally arising in Combinatorial Physics lead to such statistics and new ones some of which can be interpreted combinatorially.

VI. GENERALIZED STIRLING NUMBERS IN COMBINATORIAL PHYSICS

Many tools of Quantum Mechanics bail down to the consideration of creation and annihilation operators which will be

¹The name is related to the notion of *free Burnside semigroups*, namely the quotient of the free semigroup A^+ , where A is a finite alphabet, by the the smallest congruence that contains the relators $\omega^{n+m} = \omega^n$, $\omega \in A^+$. For more details see [12].

here denoted respectively a^\dagger and a . These two symbols do not commute and are subject to the unique relation

$$[a, a^\dagger] = 1 . \quad (17)$$

The complex algebra generated by these two symbols and this unique relation, the *Heisenberg-Weyl algebra*, will be here denoted by $HW_{\mathbb{C}}$. The consideration of evolution (one-parameter) groups $e^{\lambda\Omega}$ where $\Omega = \sum_{\omega \in HW_{\mathbb{C}}} \alpha(\omega)\omega$ is an element of $HW_{\mathbb{C}}$, with all - but a finite number of them - the complex numbers $\alpha(\omega)$ equal to 0, and ω a word on the alphabet $\{a, a^\dagger\}$ leads to the necessity of solving the *Normal Ordering Problem*, i.e., the reduction of the powers of Ω to the form

$$\Omega^n = \sum \beta_{i,j}(a^\dagger)^i a^j . \quad (18)$$

In the sequel, $Normal(\Omega^n)$ denotes such a sum. This problem can be performed with three indices in general and two in the case of homogeneous operators that is operators for which the "excess" $e = i - j$ is constant along the monomials $(a^\dagger)^i a^j$ of the support (for which $\beta_{i,j} \neq 0$). Thus, for

$$\Omega = \sum_{i-j=e} \beta_{i,j}(a^\dagger)^i a^j \quad (19)$$

one has, for all $n \in \mathbb{N}$,

$$Normal(\Omega^n) = (a^\dagger)^{ne} \sum_{k=0}^{\infty} S_{\Omega}(n, k)(a^\dagger)^k a^k \quad (20)$$

when $e \geq 0$, and

$$Normal(\Omega^n) = \left(\sum_{k=0}^{\infty} S_{\Omega}(n, k)(a^\dagger)^k a^k \right) a^{n|e|} \quad (21)$$

otherwise. It turns out that, when there is only one annihilation, one gets a formula of the type (x, y are formal commutative variables)

$$\sum_{n,k \geq 0} S_{\Omega}(n, k) \frac{x^n}{n!} y^k = g(x) e^y \sum_{n \geq 1} S_{\Omega}(n, 1) \frac{x^n}{n!} \quad (22)$$

which is a generalization of formula (16). A complete study of such a procedure and the details to perform the solution of the normal ordering problem may be found in [6].

VII. CONCLUSION

In this paper, we have broadened the domain of application of the exponential formula, a tool originated from statistical physics. This broadening reveals us together with the essence of "why this formula works" a possibility of extension to denominators other than the factorial and also a link with one-parameter groups whose infinitesimal generators are (formal) vector fields on the line. The general combinatorial theory of the correspondence (vector fields \leftrightarrow bivariate statistics) is still to be done despite the fact that we have already a wealth of results in this direction.

REFERENCES

- [1] R. H. Bruck, *A survey of binary systems*, Ergebnisse der Mathematik und ihrer Grenzgebiete, new series, vol. 20, Berlin-Göttingen-Heidelberg, Springer, 1958.
- [2] A. Dress and T. Müller, "Decomposable functors and the exponential principle", *Advances in Mathematics*, vol. 129, pp. 188-221, 1997.
- [3] S. Eilenberg, *Automata, Languages and Machines - volume A*, Academic Press, 1974.
- [4] P. Flajolet, *Analytic Combinatorics*, Cambridge University Press, 2008.
- [5] I. P. Goulden and D. M. Jackson, *Combinatorial enumeration*, John Wiley & Sons, Inc., 1983.
- [6] G. H. E. Duchamp, K. A. Penson, A. I. Solomon, A. Horzela and P. Blasiak, "One-parameter groups and combinatorial physics", in *Proc. of the Third International Workshop on Contemporary Problems in Mathematic Physics (COPROMAPH3)*, Porto-Novo (Benin), 2003.
- [7] A. Joyal, "Une théorie combinatoire des séries formelles", *Advances in Mathematics*, vol. 42, pp. 1-82, 1981.
- [8] B. Harris and L. Shoenfeld, "The number of idempotent elements in symmetric semigroups", *Journal of Combinatorial Theory, Series A*, vol. 3, pp. 122-135, 1967.
- [9] C. Krattenthaler and T. W. Müller, "Equations in finite semigroups: explicit enumeration and asymptotics of solution numbers", *Journal of Combinatorial Theory, Series A*, vol. 105, pp. 291-334, 2004.
- [10] S. Lang, *Complex analysis*, Springer, 1999.
- [11] E. S. Ljapin and A. E. Evseev, *The Theory of Partial Algebraic Operations*, Kluwer Academic, 1997.
- [12] A. Pereira do Lago and I. Simon, "Free Burnside Semigroups", *Theoretical Informatics and Applications*, vol. 35, pp. 579-595, 2001.
- [13] R. J. Ridell and G. E. Uhlenbeck, "On the theory of the virial development of the equation of state of monatomic gases", *J. Chem. Phys.*, vol. 21, pp. 2056-2064, 1953.
- [14] G. Segal, "Configuration-spaces and iterated loop-spaces", *Inventiones Mathematicae*, vol. 21 (3), pp. 213-221, 1973.
- [15] R. Stanley, *Enumerative Combinatorics - Volume I*, in *Studies in Advanced Mathematics* vol. 49, Cambridge University Press, 1997.
- [16] J. Touchard, "Sur les cycles des substitutions", *Acta Math.*, vol. 70, pp. 243-297, 1939.

Exponential random graphs as models of overlay networks

Moez DRAIEF

In this talk, we describe a Metropolis-Hastings type algorithm to construct optimised overlay networks in the context of Peer-to-Peer systems to provide load balancing and minimise communication costs. We show that the graphs obtained are distributed according to a Gibbs like distribution $\mathcal{P}(G) = \frac{1}{Z} e^{-\beta \sum_i d_i^2}$ on a subset of the set of connected graphs with n nodes, d_i being the degree of node i . Using analogies with the configuration model for graph generation we derive a number of asymptotic results for the properties of such graphs. More precisely we show that the degrees are concentrated around their mean value, derive asymptotic results on the number of edges crossing a graph cut and use these results (i) to compute the graph expansion and conductance, and (ii) to analyse the graph resilience to random failures.

Coding rhombus tilings by multidimensional words: a first attempt

Damien JAMET

Since the discovery of quasicrystals, rhombus tilings have provided well adapted models for quasicrystalline alloys. In the present work, we focus on a particular class of rhombus tilings, namely the ones one can code with a 2D-sequence over the three-letter alphabet 1,2,3. Among these tilings, we deeply investigate the ones coding randomized rhombus tilings and the ones coding the trajectories of an element of the torus \mathbb{R}/\mathbb{Z} under the action of two irrational rotations. Firstly, we study the recognition problem of 2D-sequences coding a rhombus tiling: given a 2D-sequence, does it represent a rhombus tiling? We then show that the set of such sequences is of finite type, that is, a sequence coding a rhombus tiling is entirely characterized by a finite set of its configurations. In the second part of this talk, we investigate the combinatoric properties of such tilings such as, for instance, the enumeration of local configurations. We will end this talk by stating several perspectives and challenges of these researchs, in computer science as well as in mathematics.

Synchronization of countable cellular systems, localization of quasi-periodic solutions of autonomous differential systems

Laurent Gaubert

Abstract—We address the question of frequencies locking in coupled differential systems, related to the existence of quasi-periodic solutions of differential systems. Our tool is what we call “cellular systems”, quite general as it deals with countable number of coupled systems in some general Banach spaces. Moreover, the inner dynamics of each subsystem does not have to be specified. We reach some general results about how the frequencies locking phenomenon is related to the structure of the coupling map, and therefore about the localization of quasi-periodic solutions of some differential systems that may be seen as cellular systems. This paper gives some explanations about how and why synchronized behaviors naturally occur in a wide variety of complex systems.

Index Terms—coupled differential systems, synchronization, frequencies locking.

I. INTRODUCTION

SYNCHRONIZATION is an extremely important and interesting emergent property of complex systems. The first example found in literature goes back to the 17th century with Christiaan Huygens’ work [11], [2]. This kind of emergent behavior can be found in artificial systems as well as in natural ones and at many scales (from cell to whole ecological systems). Biology abounds with periodic and synchronized phenomena and the work of Ilya Prigogine shows that such behaviors arise within specific conditions: a dissipative structure generally associated to a nonlinear dynamics [20]. Biological systems are open, they evolve far from thermodynamic equilibrium and are subject to numerous regulating processes, leading to highly nonlinear dynamics. Therefore periodic behaviors appear (with or without synchronization) at any scale [21]. More generally, life itself is governed by circadian rhythms [9]. Those phenomena are as much attractive as they are often spectacular: from cicada populations that appear spontaneously every ten or thirteen years [10] or networks of heart cells that beat together [17] to huge swarms in which fireflies, gathered in a same tree, flash simultaneously [3]. This synchronization phenomenon occupies a privileged position among emergent collective phenomena because of its various applications in neuroscience, ecology, earth Science, for instance [27], [25], [16], as well as in the field of coupled dynamical systems, especially through the notion of chaotic systems’ synchronization [18], [7] and the study of coupled-oscillators [13]. This wide source of examples leads the field

L. Gaubert is with Centre Européen de Réalité Virtuelle, LISYC EA3883 UBO/ENIB, 25 rue Claude Chappe, 29280 Plouzané, France, e-mail: (gaubert@enib.fr).

Manuscript received April 19, 2005; revised January 11, 2007.

of research to be highly interdisciplinary, from pure theory to concrete applications and experimentations.

The classical concept of synchronization is related to the locking of the basic frequencies and instantaneous phases of regular oscillations. One of the most successful attempts to explore this emergent property is due to Kuramoto [14], [15]. As in Kuramoto’s work, those questions are usually addressed by studying specific kinds of coupled systems (see for instance [5], [22], [8]). Using all the classical methods available in the field of dynamical systems, researchers study specific trajectories of those systems in order to get information on possible attracting synchronized state [28], [13], [22], [19], [8], [12].

The starting point of this work was the following question : “Why synchronization is such a widely present phenomena ?” In order to give some mathematical answer to this question, the first step is to build a model of coupled systems that is biologically inspired. This is what is done in the second section, after having described some basic material, we define what we call cellular systems and cellular coupler. If one would summarize the specificities of cellular system, one could say that each cell (subsystem) of a cellular system receives information from the whole population (the coupled system) according to some constraints:

- a cell has access to linear transformations of all the others cell’s states
- the way this information is gathered depends (not linearly) on the cell’s state itself

In other words, a cell interprets its own environment via the states of the whole population and according to its own state.

It’s a bit surprising that despite this model arising very naturally, it gives a good framework to address the main question. Indeed, in the third section we expose a localization result concerning periodic trajectories of cellular systems, according to some sub-periods dependencies. It exhibits some links between the coupler’s properties and the structure of periodic trajectories.

The fourth section gives some example of general results that may be proved using the localization lemma. Moreover, it goes out of the scope of coupled systems as synchronization is strongly related to the more abstract field of dynamical systems. If one thinks about presence of regular attractors (in opposition with strange attractors) in a differential system, one may for example classify those as:

- point attractor

- limit cycle
- limit torus

Those attractors can be related to coupled systems in an obvious way: roughly speaking, a point attractor may be seen as a solution of coupled systems for which each of the subsystems has a constant behavior. Similarly, a limit cycle may be thought as the situation where every subsystem oscillate, all frequencies among the whole system being locked. A limit torus is a similar situation which differs from the previous one by the fact that the frequencies are not locked (non commensurable periods of a quasi-periodic solution of the whole coupled system). Hence, the three previous cases may be translated into the coupled dynamical systems context:

- point attractor \leftrightarrow constant trajectories
- limit cycle \leftrightarrow periodic trajectories, locked frequencies
- limit torus \leftrightarrow periodic trajectories, unlocked frequencies

Therefore, we deduce some results about the localization of solutions of the third type, quasi-periodic solutions, using the point of view of coupled dynamical systems. The results of this third section may help to understand why the second case is the most observed in natural systems, which may be seen as coupled dynamical systems, at many levels. Indeed, the section ends with a sketch of how the cellular systems point of view may be applied to a wide class of differential systems in order to systematically address those questions with algebraic tools.

II. BASIC MATERIAL AND NOTATIONS

As our model is inspired by cellular tissues, some terms clearly come from the vocabulary used to describe those kinds of complex systems.

A. Model of population behavior

Here are the basic compounds and notations of our model:

A *population* \mathcal{I} is a countable set, so we may consider it as a subset $\mathcal{I} \subset \mathbb{N}$. Moreover, as it's only the cardinality of \mathcal{I} that's important, \mathcal{I} may be chosen as an interval of integer. Elements of \mathcal{I} are called *cells*.

We suppose that the systems we want to study are valued in some Banach spaces. Thus, for any $i \in \mathcal{I}$, $(E_i, \|\cdot\|_i)$ is a Banach space, and the *state space* of \mathcal{I} is the vector space $\mathcal{S} = \prod_{i \in \mathcal{I}} E_i$.

We will sometimes identify E_i with

$$\prod_{j < i} \{0\} \times E_i \times \prod_{j > i} \{0\} \subset \mathcal{S}$$

and then consider it as a subspace of \mathcal{S} .

We denote \mathcal{S}_b the space of uniformly bounded states:

$$\mathcal{S}_b = \left\{ x \in \mathcal{S}, \sup_{i \in \mathcal{I}} \|x_i\|_i < \infty \right\}$$

This subspace will sometimes be useful as, embodied with the norm $\|x\|_\infty = \sup_{i \in \mathcal{I}} \|x_i\|_i$, it's a Banach space, allowing

the classic PicardLindelöf theorem to be valid.

Given an interval $\Omega \subset \mathbb{R}$, a *trajectory* x of \mathcal{I} is an element of $\mathcal{C}^\infty(\Omega, \mathcal{S})$. Such x is then described by a family of \mathcal{C}^∞ applications $(x_i)_{i \in \mathcal{I}}$ such that $\forall i \in \mathcal{I}$:

$$\begin{aligned} x_i : \Omega &\longrightarrow E_i \\ t &\longmapsto x_i(t) \end{aligned}$$

The space of trajectories on I is denoted \mathcal{T} .

A *period* on \mathcal{I} is a map $\tau : \mathcal{I} \rightarrow \mathbb{R}_+^*$. A trajectory $x \in \mathcal{T}$ is said to be τ -*periodic* if for any $i \in \mathcal{I}$, x_i is $\tau(i)$ -periodic and non constant. $\tau(i)$ is then said to be a *period of the cell* i . The space of such trajectories is written \mathcal{T}_τ .

Each cell i is supposed to behave according to an autonomous differential system given by a vector field $F_i : E_i \rightarrow E_i$. Thus, given a family of functions $\{F_i\}_{i \in \mathcal{I}}$ we define the vector field $F_{\mathcal{I}}$ on \mathcal{S} :

$$\begin{aligned} F_{\mathcal{I}} : \mathcal{S} &\longrightarrow \mathcal{S} \\ x &\longmapsto F_{\mathcal{I}}(x) \end{aligned}$$

with, for any $i \in \mathcal{I}$:

$$[F_{\mathcal{I}}(x)]_i = F_i(x_i)$$

Remark. The definition of periodic trajectory handle both classical concepts of periodic and quasi-periodic solutions of a differential system. From the point of view of coupled systems, it describes the situation in which each subsystem of the whole system oscillates. We stress the point that a period of a periodic trajectory needs not to be a minimal period ($\tau(i)$ isn't necessarily a generator of the group of periods of x_i). Nevertheless, our definition of \mathcal{T}_τ avoid any trajectory which contains some constant component (none of the x_i can be a constant map) as they may be seen as degenerate (localized into an "hyperplane" of \mathcal{S}).

We recall that a (finite) subset $\{\tau_1, \dots, \tau_k\}$ of \mathbb{R} is said to be **rationally dependent** if there exists some integers l_1, \dots, l_k non all zero and such that:

$$l_1\tau_1 + \dots + l_k\tau_k = 0$$

Then there exists a unique lowest common multiple (*lcm*) τ_0 for which there exists n_1, \dots, n_k such that:

$$n_1\tau_1 = \dots = n_k\tau_k = \tau_0$$

An infinite set of real numbers is said to be rationally dependent if any finite subset is rationally dependent.

Now, any period τ on \mathcal{I} (or, equivalently, any periodic trajectory) defines an equivalence relation on \mathcal{I} as:

$$i \sim j \Leftrightarrow \{\tau(i), \tau(j)\} \text{ is a dependent set}$$

Hence we may consider the (countable) partition $\mathcal{I}(\tau)$ of \mathcal{I} into equivalence classes (K countable):

$$\mathcal{I}(\tau) = \{\mathcal{I}_k\}_{k \in K}$$

B. Cellular coupler and cellular systems

In this section we build what we call *cellular systems* by way of *cellular coupler*. Most of the works always deal with a specific way of coupling dynamical systems: one adds a quantity (that models interactions between subsystems) to the derivative of the systems. This leads to equations with the following typical shape (here, there are only two coupled systems):

$$\begin{aligned}x'_1(t) &= F(x_1(t)) + G_1(x_1(t), x_2(t)) \\x'_2(t) &= F(x_2(t)) + G_2(x_1(t), x_2(t))\end{aligned}$$

The functions G_1 and G_2 are the coupling functions. The problem is then restated in terms of phase-shift variables and efforts are made to detect stable states and to prove their stability.

Our approach to coupling is different. We study exclusively a way of coupling where the exchanges are made on the current state of the system. This means that the coupling quantity applies inside the map F , which leads us to the following type of equation:

$$\begin{aligned}x'_1(t) &= F(x_1(t) + H_1(x_1(t), x_2(t))) \\x'_2(t) &= F(x_2(t) + H_2(x_1(t), x_2(t)))\end{aligned} \quad (1)$$

Remark. We stress the point that those two different ways of handling coupled systems are quite equivalent in most cases. Indeed, starting with the first two equations, as soon as G_1 and G_2 stay in the range of F (which is likely if the coupling functions are small), we can rewrite them in the second shape involving H_1 and H_2 .

The last kind of coupled systems is sometimes studied (for instance in [12]) but never broadly (indeed, if one wants some quantitative results about convergence of trajectories, one must work with specific equations and dynamical systems). Even in a few papers that are quite general (as the very interesting [24]) some strong assumptions are made (in [24] authors deal with symmetric periodic solutions). The kind of coupled systems we handle are a generalization of the one describe in equation (1). Its general shape is:

$$x'_i(t) = F_i \left(\sum_{j \in \mathcal{I}} c_{ij}(x_i(t)) x_j(t) \right)$$

Each cell $i \in \mathcal{I}$ holds it's own differential system represented by a map F_i (hence, all the dynamical systems are not forcibly identical nor have the same shape, nor that they are weakly coupled (as in the classical paper of Art Winfree [26])). A cell i "interprets" it's own environment by mean of the functions c_{ij} .

Before giving the exact definition of a cellular coupler, we recall that, as \mathcal{S} may be seen as a module on the ring $\prod_{i \in \mathcal{I}} \mathcal{L}(E_i)$, $\mathcal{L}(\mathcal{S})$ has to be understood as the space of linear operators on \mathcal{S} with coefficients in the $\mathcal{L}(E_i, E_j)$. Any

$M \in \mathcal{L}(\mathcal{S})$ may then be written as an infinite (if \mathcal{I} isn't finite) matrix:

$$M = [m_{ij}]_{(i,j) \in \mathcal{I}^2}, \quad m_{ij} \in \mathcal{L}(E_j, E_i)$$

In this context, here is the definition of a cellular coupler on \mathcal{I} :

Definition 1. A *cellular coupling map* on \mathcal{I} is a map c :

$$\begin{aligned}c : \mathcal{S} &\longrightarrow \mathcal{L}(\mathcal{S}) \\x &\longmapsto c(x)\end{aligned}$$

such that the matrix $[c_{ij}]_{(i,j) \in \mathcal{I}^2}$ satisfies:

- 1) $\forall (i, j) \in \mathcal{I}^2, \forall x \in \mathcal{S}, c_{ij}(x)$ depends only on x_i (so that we may consider it as a map $c_{ij} : E_i \rightarrow \mathcal{L}(E_j, E_i)$);
- 2) $\forall i \in \mathcal{I}, \forall x_i \in E_i, \sum_{j \in \mathcal{I}} \|c_{ij}(x_i)\| < +\infty$

Then, c defines a *cellular coupler* \tilde{c} on \mathcal{I} in the following way:

$$\begin{aligned}\tilde{c} : \mathcal{S} &\longrightarrow \mathcal{S} \\x &\longmapsto c(x).x\end{aligned}$$

In other words (for the sake of simplicity, we only take examples with a finite population), for any $x \in \mathcal{S}$, the matrix $c(x)$ has the following shape:

$$c(x) = \begin{bmatrix} c_{11}(x_1) & \cdots & c_{1k}(x_1) \\ \vdots & \ddots & \vdots \\ c_{k1}(x_k) & \cdots & c_{kk}(x_k) \end{bmatrix} \in \mathcal{L}(\mathcal{S})$$

And then :

$$\tilde{c}(x) = c(x).x = \begin{bmatrix} c_{11}(x_1).x_1 + \cdots + c_{1k}(x_1).x_k \\ \vdots \\ c_{k1}(x_k).x_1 + \cdots + c_{kk}(x_k).x_k \end{bmatrix} \in \mathcal{S}$$

Now we can define a cellular system:

Definition 2. Let $F_{\mathcal{I}}$ be a vector field on \mathcal{S} given by a family $\{F_i\}_{i \in \mathcal{I}}$ of vector fields on the E_i . Let \tilde{c} a cellular coupler on \mathcal{I} . $(\mathcal{I}, F_{\mathcal{I}}, \tilde{c})$ is called a *cellular system*. A trajectory of this system is a trajectory $x \in \mathcal{T}$ that satisfies:

$$x' = F_{\mathcal{I}} \circ \tilde{c}(x) = F_{\mathcal{I}}(c(x).x)$$

in other words:

$$\forall i \in \mathcal{I}, \forall t \in \Omega, x'_i(t) = F_i \left(\sum_{j \in \mathcal{I}} c_{ij}(x_i(t)).x_j(t) \right)$$

This equation may be naturally interpreted in biological terms: the cell i behaves according to a mean of the states of all other cells x_j , but only its state defines how this mean is computed (the cell interprets its own environment), and this link *state* \leftrightarrow *interpreting function* has no reason to be linear in x_i .

Remark. In order to avoid any confusion, we stress the differences between trajectory and solution regarding periodic behaviors. In this paper, periodic trajectory has a specific meaning related to the cells. A periodic trajectory of a cellular

system is a trajectory for which each cell has a periodic behavior. From the classic point of view of differential equations, periodic and quasi-periodic solutions of the cellular system are periodic trajectories. In the case of a periodic solution, $\tau(\mathcal{I})$ admit a *lcm*, which false in the case of a quasi-periodic solution.

In the next section we start of by exposing algebraic links between a cellular coupler and a periodic trajectory, and then we turn to our localization lemma.

III. LOCALISATION LEMMA

If M is a matrix indexed on \mathcal{I}^2 , and if $J \subset \mathcal{I}$, we write $\bar{J} = \mathcal{I} - J$ and we define M^J as the matrix:

$$M = (m_{ij})_{(i,j) \in J \times \bar{J}}$$

For $x \in \mathcal{S}$ (resp. $x \in \mathcal{T}$) we denote x^J the vector (resp. the map) $[x_i]_{i \in J}$ (see figure 1).

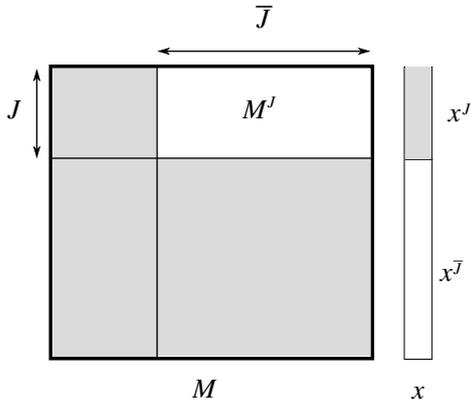


Fig. 1. Matrix and vectors associated to a subset J of \mathcal{I} .

If $\mathcal{I}(\tau) = \{I_1, \dots, I_K\}$ is a partition of \mathcal{I} , we define the matrix $M^{\mathcal{I}(\tau)}$ as (see figure 2):

$$m_{ij}^{\mathcal{I}(\tau)} = \begin{cases} 0 & \text{if } (i, j) \in I_1 \times I_1 \cup \dots \cup I_K \times I_K \\ m_{ij} & \text{if not} \end{cases}$$

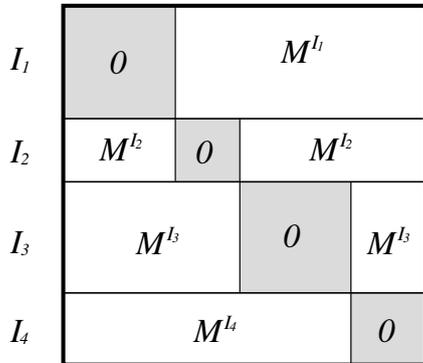


Fig. 2. Matrix associated to a partition of \mathcal{I} .

We can then go further and begin to work on the heart of our matter. The forthcoming result that can be used in many ways and generalized as, for the sake of simplicity, we did not use the weakest assumptions under which it holds (for example, the series convergence in the proof can be insured in many other contexts).

Lemma 1. *Let $(\mathcal{I}, F_{\mathcal{I}}, \tilde{c})$ be a cellular system and τ a period on \mathcal{I} . Let $U \subset \mathcal{S}$ on which $F_{\mathcal{I}}$ is injective. If $x \in \mathcal{T}^\tau$ is a periodic trajectory of cellular system that satisfies:*

- 1) $x(\Omega) \subset \mathcal{S}_b$;
- 2) $\tilde{c}(x)(\Omega) \subset U$

then there exists $b \in \mathcal{S}_b$ such that for any $t \in \Omega$:

$$x(t) - b \in \ker \left(c(x(t))^{\mathcal{I}(\tau)} \right)$$

Remark. Note that the first condition on x is useless if \mathcal{I} is finite.

The previous result is not very practical as it involves the trajectory x itself, which is unknown. As there is no ambiguity, we define the kernel of $c^{\mathcal{I}(\tau)}$ as:

$$\ker \left(c^{\mathcal{I}(\tau)} \right) = \bigcup_{x \in \mathcal{S}} \ker \left(c(x)^{\mathcal{I}(\tau)} \right)$$

Hence we may give a weaker version of the previous lemma

Corollary 1. *Under the conditions of lemma 1 there exists $b \in \mathcal{S}$ such that:*

$$x(\Omega) - b \in \ker \left(c^{\mathcal{I}(\tau)} \right)$$

Proof: (of lemma 1) First of all, let's check that $\tilde{c}(x)$ is τ -periodic.

For any $i \in \mathcal{I}$, x'_i is $\tau(i)$ -periodic and non constant for x_i is so. Let's $U_i = U \cap E_i$, F_i has to be injective on U_i . Hence, as x is a trajectory of the cellular system, $F_i(\tilde{c}(x)_i)$ must be periodic and then $\tilde{c}(x)_i$ is $\tau(i)$ -periodic. Therefore, $c(x)$ is τ -periodic.

Now, according to the partition $\mathcal{I}(\tau) = \{\mathcal{I}_k\}_{k \in K}$ defined by τ (see section II-A), let $k \in K$ and $i \in \mathcal{I}_k$. For any $M \in \mathbb{N}$ we define the following set:

$$\mathcal{I}_k^M = \mathcal{I}_k \cap \llbracket 0, M \rrbracket$$

The set $\tau(\mathcal{I}_k^M)$ is now a finite dependent set, so that we can consider its *lcm* τ_k^M . Now, for any $j \in \mathcal{I}_k^M$, x_j and $\tilde{c}(x)_j$ are

τ_j^M -periodic, so that, for any integer N :

$$\begin{aligned}
\tilde{c}(x)_i(t) &= \frac{1}{N+1} \sum_{l=0}^N \tilde{c}(x)_i(t + l\tau_k^M) \\
&= \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \mathcal{I}} c_{ij}(x_i(t + l\tau_k^M)) \cdot x_j(t + l\tau_k^M) \\
&= \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \mathcal{I}} c_{ij}(x_i(t)) \cdot x_j(t + l\tau_k^M) \\
&= \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \cdot x_j(t) \\
&\quad + \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \mathcal{I}_k - \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \cdot x_j(t + l\tau_k^M) \\
&\quad + \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \cdot x_j(t + l\tau_k^M) \\
&= \sum_{j \in \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \cdot x_j(t) \\
&\quad + \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \mathcal{I}_k - \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \cdot x_j(t + l\tau_k^M) \\
&\quad + \frac{1}{N+1} \sum_{l=0}^N \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \cdot x_j(t + l\tau_k^M)
\end{aligned}$$

As x is uniformly bounded, using the second property of a coupling map (def. 1), we may invert the summation orders in the previous equation and compute the limits when $M \rightarrow +\infty$ and $N \rightarrow +\infty$ in any order. Thus we have:

$$\begin{aligned}
\tilde{c}(x)_i(t) &= \sum_{j \in \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \cdot x_j(t) \\
&\quad + \sum_{j \in \mathcal{I}_k - \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \left[\frac{1}{N+1} \sum_{l=0}^N x_j(t + l\tau_k^M) \right] \\
&\quad + \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \left[\frac{1}{N+1} \sum_{l=0}^N x_j(t + l\tau_k^M) \right]
\end{aligned}$$

for the same reasons, it's easy to show that: and

$$\lim_{M, N \rightarrow +\infty} \sum_{j \in \mathcal{I}_k - \overline{\mathcal{I}_k^M}} c_{ij}(x_i(t)) \left[\frac{1}{N+1} \sum_{l=0}^N x_j(t + l\tau_k^M) \right] = 0$$

Now, as for all $j \in \overline{\mathcal{I}_k}$, τ_k^M and $\tau(j)$ are non commensurable, if we denote τ'_j the generator of x_j group of period, as $\tau(j) = n_j \tau'_j$ for a certain integer n_j , τ_k^M and τ'_j as well are non commensurable. Therefore, the sequence $\left(\frac{t + l\tau_k^M}{\tau'_j} \right)_{l \in \mathbb{N}}$ is equidistributed mod 1, and we may apply some classic ergodic

theorem (see for instance [23], [4]) and write:

$$\begin{aligned}
\lim_{N \rightarrow +\infty} \frac{1}{N+1} \sum_{l=0}^N x_j(t + l\tau_k^M) &= \frac{1}{\tau'_j} \int_0^{\tau(j)} x_j(s) ds \\
&= \frac{n_j}{\tau(j)} \int_0^{\tau(j)} x_j(s) ds
\end{aligned}$$

We can now define the state b as:

$$b = [b_j]_{j \in \mathcal{I}}, \quad b_j = \frac{n_j}{\tau(j)} \int_0^{\tau(j)} x_j(s) ds$$

so that:

$$\begin{aligned}
\lim_{N \rightarrow +\infty} \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \left[\frac{1}{N+1} \sum_{l=0}^N x_j(t + l\tau_k^M) \right] \\
= \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) b_j
\end{aligned}$$

hence, we have shown that:

$$\tilde{c}(x)_i(t) = \sum_{j \in \mathcal{I}_k} c_{ij}(x_i(t)) \cdot x_j(t) + \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \cdot b_j$$

But, obviously, from the beginning we had:

$$\tilde{c}(x)_i(t) = \sum_{j \in \mathcal{I}_k} c_{ij}(x_i(t)) \cdot x_j(t) + \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \cdot x_j(t)$$

So that:

$$\sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \cdot x_j(t) = \sum_{j \in \overline{\mathcal{I}_k}} c_{ij}(x_i(t)) \cdot b_j$$

The previous work can be done for any i which belongs to \mathcal{I}_k , thus we can summarize in the following way (see previously defined notations):

$$c(x(t))^{\overline{\mathcal{I}_k}} \cdot (x(t) - b)^{\overline{\mathcal{I}_k}} = 0$$

Again, the previous conclusion still holds for any $k \in K$, hence we may conclude using our notations:

$$c(x(t))^{\mathcal{I}(\tau)} \cdot (x(t) - b) = 0$$

■

In the next section we give some examples of results based upon this lemma. We will mainly show how lemma 1 may be applied to turn synchronization issues (and existence of quasi-periodic solutions to a differential system) into algebraic problems. One of the main argument is that one wants to avoid periodic trajectories for which one cell is inert (a constant map), as it may be discarded from the population (in the case of an infinite population, this may lead to recurrence reasoning).

IV. APPLICATIONS

A. Weakly injective coupler

In this example we just write down an elementary property of \tilde{c} which ensures that a periodic trajectory must have an inert cell.

Definition 3. Let \tilde{c} be a cellular coupler on \mathcal{I} . \tilde{c} is said to be weakly injective if for any non trivial partition $\mathcal{I}(\tau)$ of \mathcal{I} there exist $i \in \mathcal{I}$ such that:

$$\forall x \in \mathcal{S}, \ker \left(c(x)^{\mathcal{I}(\tau)} \right) \cap E_i = \{0\}$$

Now we can state a simple result:

Proposition 1. Under the conditions of lemma 1, if \tilde{c} is weakly injective and if x is a τ -periodic trajectory of the cellular system, then $\tau(\mathcal{I})$ is a dependent set.

Proof: Let suppose that $\mathcal{I}(\tau)$ is not trivial, applying lemma 1 we know that:

$$c(x(t))^{\mathcal{I}(\tau)} \cdot (x(t) - b) = 0$$

As \tilde{c} is weakly injective, there exists $i \in \mathcal{I}$ such that:

$$\forall t \in \Omega, x(t)_i = b_i$$

which contradicts the definition of a periodic trajectory. ■

This result may be restated in terms of quasi-periodic solution of the cellular system:

Proposition 2. Under the conditions of lemma 1, if \tilde{c} is weakly injective and if τ is bounded, the cellular system has no quasi-periodic solution.

The next example deals with some topological properties of a coupler (how it connects cells together).

B. Chained cellular system

In this section, for the sake of simplicity, all the vector spaces E_i have finite dimension.

We first study the case of differential systems for which the spaces E_i have same dimension and are coupled with k -nearest neighbors (the finite dimension condition isn't necessary, but it makes the exposure simpler). This case is formally described by a cellular system $(\mathcal{I}, F_{\mathcal{I}}, \tilde{c})$ where \mathcal{I} is countable, all $\dim(E_i) = n$ and \tilde{c} satisfies:

$$\forall i, j \in \mathcal{I}, |j - i| > k \Rightarrow c_{ij} = 0$$

This is what we call a *chained cellular system*. Adding the following condition on the coupler, we may reach a general result:

Definition 4. A cellular coupler \tilde{c} is said to have *full rank* if for any $i, j \in \mathcal{I}$ and $x \in \mathcal{S}$ the map $c_{ij}(x)$ has full rank

Proposition 3. Let $(\mathcal{I}, F_{\mathcal{I}}, \tilde{c})$ be a chained cellular system coupled with k -nearest neighbors (all E_i having same finite dimension). Let $F_{\mathcal{I}}$ be injective on $U \subset \mathcal{S}$ and x a τ -periodic trajectory that stays in U . If \tilde{c} has maximal rank and if there exists $I \in \mathcal{I}(\tau)$ which contains $2k$ consecutive cells, i.e. there exists $i \in \mathcal{I}$ such that:

$$\llbracket i, i + 2k - 1 \rrbracket \subset I$$

Then $\mathcal{I}(\tau) = \{I\}$ (equivalently, $\tau(\mathcal{I})$ is a dependent set).

Proof: Let suppose that $I \neq \mathcal{I}$. There must exist $\llbracket i, i + 2k \rrbracket \subset I$, such that $i - 1 \notin I$. Then, line $i + k - 1$ of the matrix

$c(x(t))^{\mathcal{I}(\tau)}$ contains only one non zero element $c_{i+k-1, i-1}$. As this linear map is injective for any $t \in \Omega$, we know that:

$$\ker(c(x(t))^{\mathcal{I}(\tau)}) \cap E_{i-1} = \{0\}$$

Applying lemma 1 we know that there exists $b_{i-1} \in E_{i-1}$ such that for any $t \in \Omega$:

$$x_{i-1}(t) - b_{i-1} \in \ker \left(c(x(t))^{\mathcal{I}(\tau)} \right) \cap E_{i-1}$$

i.e. $x_{i-1}(t) = b_{i-1}$ is a constant map, which contradicts the definition of a periodic trajectory. So we can conclude that $I = \mathcal{I}$. ■

If we assume that τ is bounded, this result may be restated as: "as soon as k consecutive cells are synchronized (locked frequencies), then all the population is synchronized".

Moreover, we may drop some assumptions made on the identical dimension of the E_i and reach an interesting connecting result concerning the case when $k = 1$.

Proposition 4. Let $(\mathcal{I}, F_{\mathcal{I}}, \tilde{c})$ be a chained cellular system coupled with 1-nearest neighbor. Let $F_{\mathcal{I}}$ be injective on $U \subset \mathcal{S}$ and x a τ -periodic trajectory that stays in U . If \tilde{c} has maximal rank and if there exists two sets I_1 and I_2 in $\mathcal{I}(\tau)$ such that:

$$\llbracket i, i + 1 \rrbracket \subset I_1 \quad \llbracket i + 2, i + 3 \rrbracket \subset I_2$$

Then $I_1 = I_2$.

Proof: Let suppose that the cells $i + 1$ have non commensurable periods with those of the cells $i + 2$ (i.e. $I_1 \neq I_2$). Following the previous proof, we know that the lines $i + 1$ and $i + 2$ of the matrix $c(x(t))^{\mathcal{I}(\tau)}$ contains only one non zero element, respectively $c_{i+1, i+2}$ and $c_{i+2, i+1}$. But, we recall that for any $t \in \Omega$:

$$c_{i+1, i+2}(x_{i+1}(t)) : E_{i+2} \rightarrow E_{i+1}$$

and

$$c_{i+2, i+1}(x_{i+2}(t)) : E_{i+1} \rightarrow E_{i+2}$$

As the coupler has maximal rank, one of the previous map must be injective for all $t \in \Omega$. Using the same argument we may conclude that either x_{i+1} is a constant map, either it's x_{i+2} , leading to a contradiction. ■

Moreover, one could restate those results in terms of quasi periodic solutions of differential systems, but it may sound less intuitive. We will do it in the next sections.

For the next example, we add some regularity conditions on the cellular system which lead to an interesting description of \mathcal{S} .

C. Localization results with bounded states

As $(\mathcal{S}_b, \|\cdot\|_{\infty})$ is a Banach space, the classic PicardLindelöf theorem is valid and we can give a version adapted to cellular systems (we stress the point that in this section, any vector field $F_{\mathcal{I}}$ has to be a vector field on \mathcal{S}_b , as well for any cellular coupler \tilde{c} , which brings some constraint on the families $(F_i)_{i \in \mathcal{I}}$ and $(c_{ij})_{(i,j) \in \mathcal{I}^2}$).

Proposition 5. If $F_{\mathcal{I}} : \mathcal{S}_b \rightarrow \mathcal{S}_b$ and \tilde{c} are locally lipschitz, which is the case if for any $x \in \mathcal{S}_b$ there exists a neighborhood $V = \prod_{i \in \mathcal{I}} V_i$, a positive number k and a sequence $(k_j)_{j \in \mathcal{I}}$ of positive numbers such that:

- 1) $\forall y, z \in V, \forall i \in \mathcal{I}, \|F_i(y_i) - F_i(z_i)\|_i \leq k \|y_i - z_i\|_i$
- 2) $\forall y, z \in V, \forall i \in \mathcal{I}, \|c_{ij}(y_i) - c_{ij}(z_i)\|_{(E_j, E_i)} \leq k_j \|y_i - z_i\|_i$
- 3) $\sum_{j \in \mathcal{I}} k_j < +\infty$

then, given any initial condition (t^0, x^0) in $\mathbb{R} \times \mathcal{S}_b$, the cellular coupling admits a unique maximal solution x that satisfies $x(t^0) = x^0$.

Before stating our localization result, we need to define the sets that any non synchronized periodic trajectory of the cellular system must avoid (or, with the classical point of view, any quasi periodic solution).

Definition 5. Let \tilde{c} be a cellular coupler on \mathcal{I} . The set of regular points for \tilde{c} is defined as:

$$R(\tilde{c}) = \left\{ x \in \mathcal{S}, \forall \mathcal{I}(\tau) \text{ partition of } \mathcal{I}, c(x)^{\mathcal{I}(\tau)} \text{ is injective} \right\}$$

We say that \tilde{c} is regular if $R(\tilde{c}) = \mathcal{S}$.

Now we can state a localization result:

Proposition 6. Under the conditions of lemma 1 and proposition 5, if there exists a infinite compact subset $V \subset \Omega$ such that:

$$\forall t \in V, x(t) \in R(\tilde{c})$$

then $\tau(\mathcal{I})$ is a dependent set.

One can rewrite this result in terms of differential systems:

Proposition 7. Under the conditions of lemma 1 and proposition 5, and if τ is bounded, a quasi-periodic trajectory must "avoid" $R(\tilde{c})$ (it can't cross this set on an infinite compact subset of Ω).

Proof: (of proposition 6) Let suppose that $\mathcal{I}(\tau)$ is not trivial, applying lemma 1 we know that:

$$c(x(t))^{\mathcal{I}(\tau)} \cdot (x(t) - b) = 0$$

the assumptions made on \tilde{c} ensure that:

$$\forall t \in V, x(t) = b$$

As V has an accumulation point, we may conclude that there exists $t_0 \in V$ such that:

$$x'(t_0) = 0$$

Proposition 5 may be applied, hence we know that $t \mapsto x(t)$ is a constant map, which contradicts the definition of a periodic trajectory. ■

The next example gives a more precise result in the case where the maps c_{ij} don't depend on the state of the system (homogeneous coupler).

D. Exact frequencies locking with homogeneous cellular coupler

If $x \in \mathcal{T}_\tau$, for any $i \in \mathcal{I}$ the map x_i equals its Fourier's series. We write:

$$e_{\tau(i)}^k(t) = \exp\left(\frac{2i\pi kt}{\tau(i)}\right)$$

and we define :

$$\hat{x}_i(k) = \frac{1}{\tau(p)} \int_0^{\tau(p)} x_i(t) e_{\tau(i)}^k(t) dt$$

so that we have :

$$x = \sum_{k \in \mathbb{Z}} \hat{x}(k) e^k$$

i.e. $\forall i \in \mathcal{I}$:

$$x_i(t) = \sum_{k \in \mathbb{Z}} \hat{x}_i(k) e_{\tau(i)}^k(t)$$

with normal convergence (note that $\hat{x}_i(k)$ is E_i -valued).

Theorem 1. Under the conditions of lemma 1, let \tilde{c} be homogeneous and regular. If τ is a bounded period on \mathcal{I} and x a τ -periodic trajectory of the cellular system then τ is constant on \mathcal{I} .

Remark. As this result is true as soon as τ is a period of x , it may be applied to the minimal periods of each x_i , then its conclusion is that all cells have exactly the same minimal period.

Proof: As \tilde{c} is homogeneous, we may identify it with c . Moreover, applying lemma 1 we know that $\tau(\mathcal{I})$ is a dependent set (unless at least one of the x_i would be a constant map). We now have to prove that τ is constant on \mathcal{I} .

Let's write a partition of \mathcal{I} according to τ 's values on \mathcal{I} (we must recall that τ is supposed bounded):

$$\{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_K\}$$

such that

$$\forall 1 \leq k \leq K, \tau(\mathcal{I}_k) = \tau_k$$

and $\tau_l \neq \tau_k$ if $l \neq k$.

We now suppose that $K > 1$.

As $\tau(\mathcal{I})$ is a finite dependent set, there exists n_1, \dots, n_K integers and τ_0 (the *lcm*) such that:

$$\tau_0 = n_1 \tau_1 = n_2 \tau_2 = \dots = n_K \tau_K$$

The trajectory x is τ_0 -periodic. We may therefore write its Fourier's series:

$$x(t) = \sum_{l \in \mathbb{Z}} \hat{x}(l) e_{\tau_0}^l(t)$$

and as well for $c.x$:

$$(c.x)(t) = \sum_{l \in \mathbb{Z}} \widehat{c.x}(l) e_{\tau_0}^l(t)$$

uniqueness of Fourier coefficients forces them to satisfy:

$$\widehat{c \cdot x}(l) = c \widehat{x}(l)$$

So that, for any $i \in \mathcal{I}$:

$$\widehat{c \cdot x}_i(l) = \sum_{j=1}^k c_{ij} \widehat{x}_j(l)$$

Now, let $i \in \mathcal{I}_k$, the properties of Fourier decomposition ensure that $\widehat{x}_i(l)$ and $c \cdot \widehat{x}_i(l)$ are zero as soon as n_k does not divide l (as $(c \cdot x)_i$ and x_i are τ_k -periodic and $\tau_0 = n_k \tau_k$).

So, if $l \in \mathbb{Z}$, let's define $I(l)$ as:

$$I(l) = \{k \in \{1, \dots, K\}, n_k \nmid l\}$$

For any integer l , if $k \in I(l)$ and $i \in \mathcal{I}_k$, then $\widehat{x}_i(l) = c \cdot \widehat{x}_i(l) = 0$, so that (with similar convergence arguments that in the proof of lemma 1):

$$\begin{aligned} c \cdot \widehat{x}_i(l) &= \sum_{j=1}^k c_{ij} \widehat{x}_j(l) \\ 0 &= \sum_{j \in I(l)} c_{ij} \widehat{x}_j(l) + \sum_{j \notin I(l)} c_{ij} \widehat{x}_j(l) \\ 0 &= \sum_{j \notin I(l)} c_{ij} \widehat{x}_j(l) \end{aligned}$$

This last property, (observable on figure 3), can be written as:

$$\forall l \in \mathbb{Z} \quad c^{I(l)} \widehat{x}(l)^{\overline{I(l)}} = 0$$

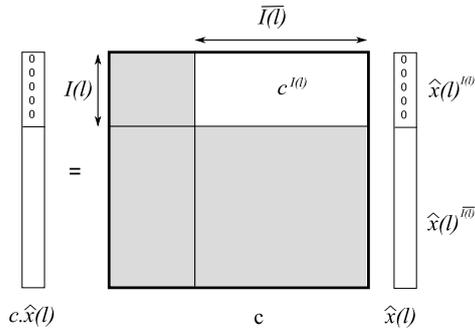


Fig. 3. Constraints on the Fourier's coefficients $\widehat{x}(l)$.

This property holds for any integer l , and is empty when l is a multiple of all the n_i . So that, if $\mathcal{I}(l)$ is the partition of \mathcal{I} defined as:

$$\mathcal{I}(l) = \{I(l), \overline{I(l)}\}$$

we can re-write it as:

$$\forall l \in \mathbb{Z} \quad c^{\mathcal{I}(l)} \widehat{x}(l) = 0$$

Let's now consider $\mathcal{I}_1 \neq \mathcal{I}_2$ (this is possible as $K > 1$). As those two classes are distinct, there exists l such that n_1 does

not divide l and n_2 divides l . As c is regular, $c^{\mathcal{I}(l)}$ is thereby injective. We deduce that:

$$\widehat{x}(l)^{\overline{I(l)}} = 0$$

This proves that for any l divisible by n_2 and not by n_1 , $\widehat{x}(l)^{\overline{I(l)}}$ is zero. Thus, for any coefficient of $\widehat{x}(l)^{\overline{I(l)}}$ to be non zero, n_1 must divide l , and consequently (as none of the x_i is a constant map) for all $i \in \mathcal{I}_2$, $x_i(t)$ is $n_1 \tau_0$ periodic. This is incompatible with the partition of \mathcal{I} . Thus, $K = 1$ and thereby τ is a constant map (in other words, \mathcal{I} is synchronized). ■

E. Perspectives of application to classical differential systems

In this last section, we show how the cellular systems point of view may be applied to classic differential systems and how dealing with different Banach spaces E_i may be useful. This discussion will be enlightened with a really simple example (finite population).

Let E be a Banach space and F a vector field on E . We want to see how this differential equation may be seen as a cellular system. For instance, one could consider a simple conservative system on $E = \mathbb{R}^4$ with an Hamilton's equation given by (see [1])

$$\begin{aligned} x'_1 &= y_1 \\ y'_1 &= \alpha x_1 - \beta x_1^3 + \varepsilon x_2 \\ x'_2 &= y_2 \\ y'_2 &= -\gamma x_2 + \varepsilon x_1 \end{aligned}$$

The first step is to identify the different cells of \mathcal{I} . The first step is to factorize each term in the equations according to the different variables. For example, the second equation may be seen as:

$$y'_1 = (\alpha - \beta x_1^2)x_1 + \varepsilon x_2$$

So that the term $(\alpha - \beta x_1^2)$ has to be a part of the coupler we are building. Moreover, as it's the equation giving y'_1 , and as the way a cell computes how it interprets the population's state depends only on its own state, x_1 and y_1 have to belong to the same cell. In this simple example it's the only case where to variables have to be gathered in the same cell. In the end, this leads to the following structure of cellular system:

$$\mathcal{I} = \{1, 2, 3\}$$

with the Banach spaces:

$$E_1 = \mathbb{R}^2, E_2 = E_3 = \mathbb{R}$$

As it should often be the case, the associated vector fields are just identity maps on E_i , and the coupler is then:

$$c = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

with

$$c_{11} : \begin{array}{ccc} E_1 & \longrightarrow & \mathcal{L}(E_1) \\ (x_1, y_1) & \longmapsto & \begin{bmatrix} 0 & 1 \\ \alpha - \beta x_1^2 & 0 \end{bmatrix} \end{array}$$

$$\begin{aligned}
c_{12} : E_2 &\longrightarrow \mathcal{L}(E_2, E_1) \\
x_2 &\longmapsto \begin{bmatrix} 0 \\ \varepsilon \end{bmatrix} \\
c_{13} : E_3 &\longrightarrow \mathcal{L}(E_3, E_1) \\
y_2 &\longmapsto \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
c_{21} : E_1 &\longrightarrow \mathcal{L}(E_1, E_2) \\
(x_1, y_1) &\longmapsto \begin{bmatrix} 0 & 0 \end{bmatrix} \\
c_{22} : E_2 &\longrightarrow \mathcal{L}(E_2) \\
x_2 &\longmapsto \begin{bmatrix} 0 \end{bmatrix} \\
c_{23} : E_3 &\longrightarrow \mathcal{L}(E_3, E_2) \\
y_2 &\longmapsto \begin{bmatrix} 1 \end{bmatrix} \\
c_{31} : E_1 &\longrightarrow \mathcal{L}(E_1, E_3) \\
(x_1, y_1) &\longmapsto \begin{bmatrix} \varepsilon & 0 \end{bmatrix} \\
c_{32} : E_2 &\longrightarrow \mathcal{L}(E_2, E_3) \\
x_2 &\longmapsto \begin{bmatrix} -\gamma \end{bmatrix} \\
c_{33} : E_3 &\longrightarrow \mathcal{L}(E_3) \\
y_2 &\longmapsto \begin{bmatrix} 0 \end{bmatrix}
\end{aligned}$$

Now, before applying some of the previous techniques, we may compute the different decomposition of c upon different non trivial partitions of \mathcal{I} . Those partitions are:

$$\begin{aligned}
P_1 &= \{\{1\}, \{2\}, \{3\}\}, P_2 = \{\{1, 2\}, \{3\}\} \\
P_3 &= \{\{1, 3\}, \{2\}\}, P_4 = \{\{1\}, \{2, 3\}\}
\end{aligned}$$

which gives:

$$\begin{aligned}
c^{P_1} &= \begin{bmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{bmatrix} & c^{P_2} &= \begin{bmatrix} 0 & 0 & c_{13} \\ 0 & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{bmatrix} \\
c^{P_3} &= \begin{bmatrix} 0 & c_{12} & 0 \\ c_{21} & 0 & c_{23} \\ 0 & c_{32} & 0 \end{bmatrix} & c^{P_4} &= \begin{bmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & 0 \\ c_{31} & 0 & 0 \end{bmatrix}
\end{aligned}$$

Now, in order to simplify, we replace the c_{ij} that are identically zero by 0, we obtain the following different matrices:

$$\begin{aligned}
c^{P_1} &= \begin{bmatrix} 0 & c_{12} & 0 \\ 0 & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{bmatrix} & c^{P_2} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{bmatrix} \\
c^{P_3} &= \begin{bmatrix} 0 & c_{12} & 0 \\ 0 & 0 & c_{23} \\ 0 & c_{32} & 0 \end{bmatrix} & c^{P_4} &= \begin{bmatrix} 0 & c_{12} & 0 \\ 0 & 0 & 0 \\ c_{31} & 0 & 0 \end{bmatrix}
\end{aligned}$$

In the end, writing the coupler as an application from \mathcal{S} to $\mathcal{L}(\mathcal{S})$, one finds those four matrices:

$$\begin{aligned}
\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 0 \\ 0 & 0 & 0 & 1 \\ \varepsilon & 0 & -\gamma & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \varepsilon & 0 & -\gamma & 0 \end{bmatrix} \\
\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\gamma & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 0 \\ 0 & 0 & 0 & 0 \\ \varepsilon & 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

At this point, we just have to check that the coupler is weakly injective:

$$\begin{aligned}
\ker(c^{P_1}) \cap E_2 &= \ker(c^{P_4}) \cap E_2 = \{0\} \\
\ker(c^{P_2}) \cap E_3 &= \ker(c^{P_3}) \cap E_3 = \{0\}
\end{aligned}$$

So, we may apply the proposition 1 and without any analytic calculus, state that this differential system may not admit any quasi-periodic solution. In other words, in case there exists periodic trajectories (which is well known to be true) they must be synchronized.

Moreover, those conclusions may hold in a more general case were the c_{ij} are less simple, and we can easily produce a result without any effort:

Proposition 8 (Generalized coupled pendulum). *Let's consider a differential system which is driven by the following equations:*

$$\begin{aligned}
x'_1 &= a_1(x_1, y_1)x_1 + a_2(x_1, y_1)y_1 + a_3(x_1, y_1)x_2 \\
&\quad + a_4(x_1, y_1)y_2 \\
y'_1 &= a_5(x_1, y_1)x_1 + a_6(x_1, y_1)y_1 \\
x'_2 &= a_7(x_2)x_2 + u(x_2)y_2 \\
y'_2 &= \varepsilon(y_2)x_1 + a_8(y_2)y_1 - \gamma(y_2)x_2 + a_9(y_2)y_2
\end{aligned}$$

If the maps u and ε never vanish, then the systems has no quasi-periodic solution.

This result doesn't have to be deep in itself, neither has it to be the most general one we could have deduced from the previous discussion. It's just a sketch of how one can handle some structure properties of a differential system, applying 1, without going into deep and specific calculus.

V. CONCLUSION

In this work we have built a general framework of cellular systems in order to handle a wide variety of coupled systems, and therefore a wide class of complex systems. We focused on an emergent property of those dynamical systems: the frequencies locking phenomenon. Usually one observes solutions of particular coupled systems and shows that within suitable conditions synchronization must occur. Those results are qualitatively dependent on the systems of interest and do not stand in the general cases. We tried to change our point of view and to bring out completing results. As we choose not to address the problem of the convergence to a periodic solution, we don't prove that synchronization ultimately happens. Instead, we consider the problem at its end: if one supposes that some coupled systems "truly" oscillates, then they must be synchronized, regardless to the individual dynamical systems as soon as the maps which define each of them are injective nearby the trajectories. The loss in time evolution information is compensated by very general results, almost independent from the individual differential systems to be coupled. For example, we proved results concerning countable coupled systems, each of them needing not to be finite dimensional. In most papers (see for instance [13]) this population of coupled systems is implicitly defined and has only two cells (sometimes a finite number N , and more

rarely an infinity). Moreover, on the contrary of what most studies about synchronization issues state, we do not assume anything concerning the cells dynamics (especially, we don't assume that they are oscillators). We only assume that they exhibit periodic behaviors under the coupling effects (the first assumption implies the second, but the opposite is clearly false).

We believe that this way of reaching general results about cellular systems gives some explanations about why the frequencies locking phenomenon emerges naturally in a large variety of coupled dynamical systems. Our results show that the following alternative is natural in many cases: either the whole population is synchronized, or its cells can't all have periodic behaviors.

Another interesting perspective is to apply this strategy to differential systems, as we outlined in the end of the fourth section. We think that it could prove useful to understand the stability or instability of limit torus when one adds some perturbations to a differential system. For example, if one already knows that there's no periodic solution, even with perturbations, and if one is able to prove, using our strategy, that a quasi-periodic solutions disappears under the effects of those perturbation, some conclusions about strange attractor emergence may be reachable.

Moreover, we have achieved some similar work on a natural generalization of this strategy to non countable population (in order to model natural systems, it's often necessary to handle continuous populations). We truly think that all those results are only a part of what can be done using cellular systems and that this work enlarges the possibilities of studying synchronization issues. But the scope of those kind of cellular systems may be beyond synchronization questions, as it's quite general and allows some theoretical studies. It could be a promising theoretical tool to model complex systems by the way of coupled differential systems.

REFERENCES

- [1] A. C. J. Luo, Predictions of quasi-periodic and chaotic motions in nonlinear Hamiltonian systems, *Chaos, Solitons & Fractals* 28 (3) (2006) 627–649.
- [2] M. Bennett, M. F. Schatz, H. Rockwood, K. Wiesenfeld, Huygens's clocks, *Proc. R. Soc. Lond.* 458 (2002) 563–579.
- [3] J. Buck, Synchronous rhythmic flashing of fireflies. ii, *Quarterly Review of Biology* 63 (3) (1988) 265–289.
- [4] L. Bunimovich, *Dynamical Systems, Ergodic Theory and Applications*, Encyclopaedia of Mathematical Sciences, 2nd ed., Springer, 2000.
- [5] G. B. Ermentrout, W. C. Troy, Phaselocking in a reaction-diffusion system with a linear frequency gradient., *SIAM-J.-Appl.-Math.* 46 (3) (1986) 359–367.
- [6] L. Gaubert, Auto-organisation et émergence dans les systèmes couplés, individuation de données issues de systèmes biologiques couplés., Ph.D. thesis, Université de Bretagne Occidentale (2007).
- [7] J.-P. Goedgebuer, P. Levy, L. Larger, Laser cryptography by optical chaos, in: J.-P. Goedgebuer, N. Rozanov, S. T. and A.S. Akhmanov, V. Panchenko (eds.), *Optical Information, Data Processing and Storage, and Laser Communication Technologies*, vol. 5135, 2003.
- [8] P. Goel, B. Ermentrout, Synchrony, stability, and firing patterns in pulse-coupled oscillators, *PHYSICA D* 163 (3-4) (2002) 191–216.
- [9] D. Gonze, J. Halloy, A. Goldbeter, Stochastic models for circadian oscillations: Emergence of a biological rhythm, *International Journal of Quantum Chemistry* 98 (2004) 228–238.
- [10] F. C. Hoppensteadt, J. B. Keller, Synchronization of periodical cicada emergences, *Science* 194 (4262) (1976) 335–337.
- [11] C. Huygens, Christiani Hugenii Zulichemii, Const. F. *Horologium oscillatorium sive De motu pendulorum ad horologia aptato demonstrationes geometricae*, Parisiis : Apud F. Muguet, Paris, France, 1673.
- [12] E. M. Izhikevich, F. C. Hoppensteadt, Slowly coupled oscillators: Phase dynamics and synchronization, *SIAM J. APPL. MATH.* 63 (6) (2003) 1935–1953.
- [13] N. Kopell, G. B. Ermentrout, *Handbook of Dynamical Systems II: Toward Applications*, chap. Mechanisms of phase-locking and frequency control in pairs of coupled neural oscillators, Elsevier, Amsterdam, 2002, pp. 3–54.
- [14] Y. Kuramoto, Self-entrainment of a population of coupled non-linear oscillators, in: *International Symposium on Mathematical Problems in Theoretical Physics*, vol. 39, Springer Berlin / Heidelberg, 1975, pp. 420–422.
- [15] Y. Kuramoto, *Chemical oscillations, waves, and turbulence*, Springer-Verlag, New York, 1984.
- [16] S. C. Manrubia, A. S. Mikhailov, D. H. Zanette, Emergence of Dynamical Order. Synchronization Phenomena in Complex Systems, World Scientific, Singapore, 2004.
- [17] D. C. Michaels, E. P. Matyas, J. Jalife, Mechanisms of sinoatrial pacemaker synchronization: A new hypothesis, *Circ. Res.* 61 (1987) 704–714.
- [18] L. M. Pecora, T. L. Carroll, G. A. Johnson, D. J. Mar, J. F. Heagy, Fundamentals of synchronization in chaotic systems, concepts, and applications, *Chaos* 7 (4) (1997) 520–543.
- [19] A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, 2001.
- [20] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, Wiley, New York, 1967.
- [21] I. Prigogine, *Theoretical Physics and Biology*, M Marois, North-Holland: Amsterdam, 1969.
- [22] L. Ren, B. G. Ermentrout, Phase locking in chains of multiple-coupled oscillators, *PHYSICA D* 143 (1-4) (2000) 56–73.
- [23] Y. G. Sinai, *Introduction to ergodic theory*, Princeton University Press, 1976.
- [24] I. Stewart, M. Golubitsky, Patterns of oscillation in coupled cell systems, in: P. P. Holmes, A. Weinstein (eds.), *Geometry, Dynamics and Mechanics: 60th Birthday Volume for J.E. Marsden*, Springer-Verlag, New York, 2002, pp. 243–286.
- [25] S. Strogatz, *Sync: The Emerging Science of Spontaneous Order*, New York: Hyperion, 2003.
- [26] A. Winfree, Biological rhythms and the behavior of populations of coupled oscillators, *Journal of Theoretical Biology* 16 (1) (1967) 15–42.
- [27] A. Winfree, *The geometry of biological time*, Springer-Verlag, New York, 1990.
- [28] C. Wu, L. O. Chua, A unified framework for synchronization and control of dynamical systems, Tech. Rep. UCB/ERL M94/28, EECS Department, University of California, Berkeley (1994).

Drawing solution curve of differential equation

Farida Benmakrouha, Christiane Hespel, and Edouard Monnier

Abstract—We develop a method for drawing approximated solution curves of differential equations. This method is based on the juxtaposition of local approximating curves on successive intervals $[t_i, t_{i+1}]_{0 \leq i \leq n-1}$.

The differential equation, considered as a dynamical system, is described by its state equations and its initial value at $t = t_0$. A generic expression of its generating series G_t truncated at any order k , of the output and its derivatives $y^{(j)}(t)$ expanded at any order k , can be calculated. These expressions are obtained from the vector fields, from the observation of the state at time t , in the state equations [3], [7].

We get an expansion of $y^{(j)}(t)$ as a linear combination of differential monomials indexed by some colored partitions.

At every initial point of the present interval, we specify the previous expressions of G_t and $y^{(j)}(t)$ for $t = t_i$. Then we obtain an approximated output $y(t)$ at order k in every interval $[t_i, t_{i+1}]_{0 \leq i \leq n-1}$. We present an example from physics: the Duffing equation.

By using Maple system, we have developed a package corresponding to the creation of the generic expression of G_t and $y^{(j)}(t)$ at order k and to the drawing of the local curves on every interval $[t_i, t_{i+1}]_{0 \leq i \leq n-1}$, by iterations on the initial points $t = (t_i)_{0 \leq i \leq n-1}$.

Index Terms—analysis of dynamical systems, symbolic algorithm, generating series, colored partitions, rational approximation

I. INTRODUCTION

The usual methods for drawing curves of differential equations consist in an iterative construction of isolated points, connected by straight lines (Runge-Kutta). Rather than calculate numerous successive approximate points $y(t_i)_{i \in I}$, it can be interesting to provide some few successive local curves $\{y(t)\}_{t \in [t_i, t_{i+1}]_{0 \leq i \leq n-1}}$.

Moreover, the computing of these local curves can be kept partly generic since a generic expression of the generating series G_{t_i} of the system can be provided in terms of t_i . The expression of the local curves $\{y(t)\}_{t \in [t_i, t_{i+1}]}$ is only a specification for $t = t_i$ at order k of the formula given in the proposition of section 3.

We consider a differential equation

$$y^{(N)}(t) = \phi(t, y(t), \dots, y^{(N-1)}(t), u(t)) \quad (1)$$

with initial conditions

$$y(0) = y_{0,0}, \dots, y^{(N)}(0) = y_{0,N}$$

We assume that $\phi(t, y(t), \dots, y^{(N-1)}(t), u(t))$ is polynomial in $y, \dots, y^{(N-1)}$.

Then this differential equation can be viewed as an affine input

F. Benmakrouha, C. Hespel, E. Monnier are with the Department of Computer Science, INSA-IRISA, 20 avenue des Buttes de Coesmes, 35043 Rennes cedex, France, e-mail: Benma@insa-rennes.fr, Hespel@insa-rennes.fr, Monnier@insa-rennes.fr.

Manuscript received March 27, 2009.

$(u(t) = (u_j(t))_{1 \leq j \leq m})$ dynamical system.

By derivating, in the Fliess's formula, the expression of $y(t)$ in a neighborhood of $t = t_0$, we get an expansion of $y^{(n)}(t_0)$. This expression can be written as a linear combination of differential monomials $\otimes_{1 \leq j \leq m} (u_j^{(i_1)})^{e_1} \dots (u_j^{(i_q)})^{e_q}$ indexed by some colored partitions $\mu = \otimes \mu_j$, for $u^\mu = u_1^{\mu_1} \dots u_m^{\mu_m}$. And then there exist some polynomials g_μ in noncommutative variables such that

$$y^{(n)}(t_0) = \sum_{\mu} \langle G_{t_0} | g_{\mu} \rangle \quad (2)$$

For the partition $(\mu_j) = (u_j^{(i_1)})^{e_1} \dots (u_j^{(i_q)})^{e_q}$, the weight $wt(\mu_j)$ and the length $lg(\mu_j)$ are

$$\begin{aligned} wt(\mu_j) &= \sum_{1 \leq k \leq m} e_k i_k \\ lg(\mu_j) &= \sum_{1 \leq k \leq m} e_k \end{aligned} \quad (3)$$

II. PRELIMINARIES

A. Affine system, Generating series

We consider the nonlinear analytical system affine in the input:

$$(\Sigma) \quad \begin{cases} \dot{q} &= f_0(q) + \sum_{j=1}^m f_j(q) u_j(t) \\ y(t) &= g(q(t)) \end{cases} \quad (4)$$

- $(f_j)_{0 \leq j \leq m}$ being some analytical vector fields in a neighborhood of $q(0)$
- g being the observation function analytical in a neighborhood of $q(0)$

Its initial state is $q(0)$ at $t = 0$. The generating series G_0 is built on the alphabet $Z = \{z_0, z_1, \dots, z_m\}$, z_0 coding the drift and z_j coding the input $u_j(t)$. Generally G_0 is expressed as a formal sum $G_0 = \sum_{w \in Z^*} \langle G_0 | w \rangle w$ where $\langle G_0 | z_{j_0} \dots z_{j_l} \rangle = f_{j_0} \dots f_{j_l} g(q)|_{q(0)}$ depends on $q(0)$.

B. Fliess's formula and iterated integrals

The output $y(t)$ is given by the Fliess's equation ([3]):

$$y(t) = \sum_{w \in Z^*} \langle G_0 | w \rangle \int_0^t \delta(w) \quad (5)$$

where G_0 is the generating series of (Σ) at $t = 0$:

$$\begin{aligned} G_0 &= \sum_{w \in Z^*} \langle G_0 | w \rangle w \\ &= \frac{g(q)|_{q(0)}}{\sum_{l \geq 0} \sum_{j_i=0}^m f_{j_0} \dots f_{j_l} g(q)|_{q(0)} z_{j_0} \dots z_{j_l}} \end{aligned} \quad (6)$$

and $\int_0^t \delta(w)$ is the iterated integral associated with the word $w \in Z^* = \{z_0, z_1, \dots, z_m\}^*$.

Remember that the iterated integral $\int_0^t \delta(w)$ of the word w for the input u is defined by

$$\begin{cases} \int_0^t \delta(\epsilon) &= 1 \\ \int_0^t \delta(vz_i) &= \int_0^t \left(\int_0^\tau \delta(v) \right) u_i(\tau) d\tau \\ &\quad \forall z_i \in Z \quad \forall v \in Z^* \end{cases} \quad (7)$$

where ϵ is the empty word, $u_0 \equiv 1$ is the drift and u_i , $1 \leq i \leq m$ is the i th input.

We define the Chen's series as follows ([2])

$$C_u(t) = \sum_{w \in Z^*} \int_0^t \delta(w) \quad (8)$$

From the previous definitions, we obtain the following expression of $y(t)$

$$y(t) = \sum_{w \in Z^*} \langle G_0 | w \rangle \langle C_u(t) | w \rangle \quad (9)$$

C. Iterated derivatives $y^{(n)}(0)$ of the output

G_0 being the generating series of the system, the i th derivative of $y(t)$ is

$$y^{(i)}(t) = \langle G_0 | C_u^{(i)}(t) \rangle \quad (10)$$

We prove the following lemma ([6]) based on the Picart-Vessiot theory ([4])

Lemma :

Let be $\sum_{0 \leq j \leq m} u_j \cdot z_j = A$. Then the derivative of the Chen's series is $\frac{d}{dt} C_u = C_u \cdot A$

From it, results the following recurrence relation:

$$C_u^{(i)} = C_u A_i, \quad A_1 = A, \quad A_{i+1} = AA_i + D_t A_i \quad (11)$$

D_t being the operator of time derivation.

Since $C_u(0) = 1$ and $C_u^{(i)}(0) = A_i(0)$ then

$$y^{(i)}(0) = \sum_{w \in Z^*} \langle G_0 | w \rangle \langle C_u^{(i)}(0) | w \rangle = \langle G_0 | A_i(0) \rangle \quad (12)$$

Let us remark that the successive derivatives $y(0), y^{(1)}(0), \dots, y^{(k)}(0)$ are obtained from the coefficients $\langle G_0 | w \rangle$ associated with the words whose length is $\leq k$.

It results that the Taylor expansion of $y(t)$ up to order k only depends on the coefficients of G_0 truncated at order k .

For instance, for a single input $u(t)$ with drift $u_0(t) \equiv 1$, the derivatives are the following

$$\begin{aligned} y(0) &= \langle G_0 | \epsilon \rangle \\ y^{(1)}(0) &= \langle G_0 | z_0 \rangle + \langle G_0 | z_1 \rangle u(0) \\ y^{(2)}(0) &= \langle G_0 | z_0^2 \rangle + (\langle G_0 | z_0 z_1 \rangle + \langle G_0 | z_1 z_0 \rangle) u(0) + \\ &\quad \langle G_0 | z_1^2 \rangle u(0)^2 + \langle G_0 | z_1 \rangle u^{(1)}(0) \\ \dots &= \dots \end{aligned} \quad (13)$$

This method allows us to compute recursively the successive derivatives of $y(t)$ at $t = 0$.

The derivation law D of the partitions, producing the effect of the time derivation D_t of the differential monomials satisfies

$$\begin{aligned} D(i_k) &= i_{k+1} \\ D(i_1^{e_1} \dots i_q^{e_q}) &= \sum_{k=1}^q e_k \times (i_1^{e_1} \dots i_k^{e_k-1} i_{k+1}^{e_{k+1}+1} \dots i_q^{e_q}) \end{aligned} \quad (14)$$

For a single input $u(t)$ with drift $u_0(t) \equiv 1$, the bicolored multiplicity is $\mu = \mu_0 \otimes \nu$ with

$$\begin{aligned} \mu_0 &= 1^p \\ wgt(\mu) &= p + wgt(\nu) \\ D(1^p \otimes \nu) &= 1^p \otimes D(\nu) \end{aligned} \quad (15)$$

III. APPROXIMATE VALUE OF $y^{(n)}(t)$

The Fliess's formula can be written

$$y(t) = \langle G_0 | \epsilon \rangle + \sum_{w \in Z^* - \{\epsilon\}} \langle G_0 | w \rangle \langle C_u(t) | w \rangle \quad (16)$$

An approximate function $y_k(t)$ de $y(t)$ up to order k in a neighborhood of $t = 0$ is obtained by expanding this expression up to the same order k . Then we have

$$|y(t) - y_k(t)| = O(t^{k+1}) \quad (17)$$

For instance, at order $k = 1$, $y(t)$ has the following approximate expression for a single input with drift

$$y_1(t) = \langle G_0 | \epsilon \rangle + \langle G_0 | z_0 \rangle t + \langle G_0 | z_1 \rangle \xi_1(t) \quad (18)$$

where $\xi_k(t)$ denotes the k th primitive of $u(t)$.

This computing can be generalized to the successive derivatives of $y(t)$.

Proposition

Given the expression of $y^{(n)}(0)$ in terms of the coefficients of G_0 and of the derivatives of order $\leq n-1$ of the input $u(t)_{t=0}$ obtained recursively according to the previous section, we can deduce the expression of $y^{(n)}(t)$ by executing in $y^{(n)}(0)$ the following transformations

- 1) **We substitute $u^{(i)}(t)$ to $u^{(i)}(0)$ for $0 \leq i \leq n-1$**
- 2) **For every occurrence of a coefficient $\langle G_0 | v \rangle$ where $v \in Z^*$, we add the following corrective term**

$$\sum_{w \neq \epsilon} \langle G_0 | wv \rangle \langle C_u(t) | w \rangle$$

The proof is based on the following properties

$$\begin{cases} \frac{d}{dt} \langle C_u(t) | vz_i \rangle &= \langle C_u(t) | v \rangle u_i(t) \\ \langle C_u(t) | \epsilon \rangle &= 1 \end{cases} \quad (19)$$

For instance, for a single input with drift, we compute from

$$y^{(1)}(0) = \langle G_0 | z_0 \rangle + \langle G_0 | z_1 \rangle u(0)$$

the expression of $y^{(1)}(t)$:

$$\begin{aligned} y^{(1)}(t) &= \langle G_0 | z_0 \rangle + \sum_{w \neq \epsilon} \langle G_0 | wz_0 \rangle \langle C_u(t) | w \rangle + \\ &\quad (\langle G_0 | z_1 \rangle + \sum_{w \neq \epsilon} \langle G_0 | wz_1 \rangle \langle C_u(t) | w \rangle) u(t) \end{aligned} \quad (20)$$

By restricting the sums to the words w whose length $|w|$ satisfies $1 \leq |w| \leq k$, we obtain a function $y_k^{(n)}(t)$ approximating $y^{(n)}(t)$ up to order k . And then

$$|y_k^{(n)}(t) - y^{(n)}(t)| = O(t^{k+1}) \quad (21)$$

A. Generalization at time $t = t_i$

For a single input with drift, the system (Σ) can be written at $t = t_i$:

$$\begin{cases} \dot{q}(t_i + h) = f_0(q(t_i + h)) + f_1(q(t_i + h))u(t_i + h) \\ y(t_i + h) = g(q(t_i + h)) \end{cases} \quad (22)$$

By setting

$$\begin{cases} U_i(h) = u(t_i + h) \\ Y_i(h) = y(t_i + h) \\ Q_i(h) = q(t_i + h) \end{cases} \quad (23)$$

we obtain the following system

$$(\Sigma_i) \quad \begin{cases} \dot{Q}_i(h) = f_0(Q_i(h)) + f_1(Q_i(h))U_i(h) \\ Y_i(h) = g(Q_i(h)) \end{cases} \quad (24)$$

And G_i is the generating series of (Σ_i) .

By setting $\psi_{i,k}(h) = \xi_k(t_i + h)$, then $\psi_{i,k}(h)$ is the k th primitive of $u(t_i + h)$ or the k th primitive of $U_i(h)$.

We have the equalities

$$\xi_1(t_i + h) = \int_{t_i}^{t_i+h} u(\tau) d\tau = \int_0^h U_i(t) dt = \psi_{i,1}(h) \quad (25)$$

And then, we can prove recursively that the Chen's integral $\int_{t_i}^{t_i+h} \delta(w)$ can be computed as an integral $\int_0^t \delta(W)$ by considering $U_i(t)$ instead of $u(t_i + t)$.

IV. APPLICATION TO CURVES DRAWING

We present an application to the curve drawing of the solution of differential equations. We consider a differential equation

$$y^{(N)}(t) = \phi(t, y(t), \dots, y^{(N-1)}(t), u(t)) \quad (26)$$

with initial conditions

$$y(0) = y_{0,0}, \dots, y^{(N)}(0) = y_{0,N}$$

It can be written for $y = q_1$:

$$\begin{cases} q_1^{(1)} = q_2 \\ q_2^{(1)} = q_3 \\ \dots = \dots \\ q_N^{(1)} = \phi(t, q_1, \dots, q_N) \end{cases} \quad (27)$$

We assume that

$$\phi(t, q_1, \dots, q_N) = P_0(q_1, \dots, q_N) + \sum_{j=1}^m P_j(q_1, \dots, q_N) u_j(t)$$

for P_0, P_1, \dots, P_m polynomials in commutative variables q_1, \dots, q_N .

For an analytical affine single input system (Σ) then $m = 1$ and the vector fields are f_0, f_1 , corresponding to P_0, P_1 .

We propose a curve drawing of the output $y(t)$ of this system in $[0, T] = \bigcup [t_i, t_{i+1}]_{0 \leq i \leq n-1}$ according to the

following algorithm:

Firstly, we compute a generic expression of the generating series G_t .

- Initial point $t_0 = 0$:
 $y(0) = q_1(0), \dots, y^{(N-1)}(0) = q_N(0)$ are given.
The vector fields f_0, f_1 applied to $g(q)$ evaluated in t_0 provide $\langle G_0 | w \rangle$ for $|w| \leq k$
- Step i :
Knowing $y(t_{i-1}) = q_1(t_{i-1}), \dots, y^{(N-1)}(t_{i-1}) = q_N(t_{i-1})$ and $\langle G_{i-1} | w \rangle$ (for $|w| \leq k$), we compute $y(t_i), \dots, y^{(N-1)}(t_i)$ according to section 3 and $\langle G_i | w \rangle$ (for $|w| \leq k$) by applying the vector fields f_0, f_1 to $g(q)$ at $q(t_i)$.
We draw the local curve of the function $t_{i-1} + dt \rightarrow y(t_{i-1} + dt)$ on the interval $[t_{i-1}, t_i]$.
- Final point $t = T = t_n$:
stop at $i = n$.

A. Genericity of the method

The computing of the coefficients

$$\langle G_i | z_{j_0} \dots z_{j_l} \rangle = f_{j_0} \dots f_{j_l} g(q) |_{q(t_i)}$$

is generic.

The computing of the expressions of

$$Y_i(h) = y(t_i + h) = y(t_i) + \sum_{|w| \leq k} \langle G_i | w \rangle \langle C_{U_i}(h) | w \rangle$$

and of

$$Y_i^{(1)}(h) = \langle G_i | z_0 \rangle + \sum_{1 \leq |w| \leq k} \langle G_i | w z_0 \rangle \langle C_{U_i}(h) | w \rangle + (\langle G_i | z_1 \rangle + \sum_{1 \leq |w| \leq k} \langle G_i | w z_1 \rangle \langle C_{U_i}(h) | w \rangle) U_i(h) \quad (28)$$

are generic too.

We use the previous algorithm by specifying t_i at every step in the previous expressions.

B. Example: Duffing equation

Its equation is the following:

$$\begin{aligned} y^{(2)}(t) + ay^{(1)}(t) + by(t) + cy^3(t) &= u(t) \\ y(0) &= y_0, \\ y^{(1)}(0) &= y_{1,0} \end{aligned} \quad (29)$$

It can be written as a first order differential system

$$\begin{cases} q_1^{(1)}(t) = q_2(t) \\ q_2^{(1)}(t) = -aq_2(t) - bq_1(t) - cq_1^3(t) + u(t) \\ y(t) = q_1(t) = g(q) \\ q_1(0) = y_0, \quad q_{2,0} = y_{1,0} \end{cases} \quad (30)$$

The vector fields are

$$\begin{aligned} f_0(q_1, q_2) &= q_2 \frac{\partial}{\partial q_1} - (aq_2 + bq_1 + cq_1^3) \frac{\partial}{\partial q_2} \\ &= q_2 \frac{\partial}{\partial q_1} + F(q) \frac{\partial}{\partial q_2} \end{aligned}$$

$$f_1(q_1, q_2) = \frac{\partial}{\partial q_2}$$

- 1) We write generic equations describing the generating series G_i at $t = t_i$:

$$\forall t_i \quad \langle G_i | z_{j_1} \cdots z_{j_i} \rangle = (f_{j_1} \cdots f_{j_i} g(q))|_{q(t_i)}$$

Let us remark that

$$\langle G_i | wz_1 \rangle = 0 \quad \forall w \in Z^*, \quad \langle G_i | wz_1 z_0 \rangle = 0 \quad \forall w \in Z^+$$

For instance, for order $k = 3$, we have only to compute 6 coefficients of G_i instead of 15 coefficients.

$$\begin{aligned} \langle G_i | \epsilon \rangle &= q_1(t_i) \\ \langle G_i | z_0 \rangle &= q_2(t_i) \\ \langle G_i | z_0^2 \rangle &= F(q(t_i)) \\ \langle G_i | z_1 z_0 \rangle &= 1 \\ \langle G_i | z_0^3 \rangle &= (q_2 \frac{\partial}{\partial q_1} F(q) + F(q) \frac{\partial}{\partial q_2} F(q))_{q(t_i)} \\ \langle G_i | z_1 z_0^2 \rangle &= -a \end{aligned} \quad (31)$$

- 2) We write generic approximate expression of the output $y(t_{i+1})$ and its derivative $y^{(1)}(t_{i+1})$ for every $t = t_{i+1} = t_i + h$ at order k :

$$\begin{aligned} y(t_{i+1}) &= \langle G_i | \epsilon \rangle + \sum_{1 \leq |w| \leq k} \langle G_i | w \rangle \langle C_{U_i}(h) | w \rangle \\ y^{(1)}(t_{i+1}) &= \langle G_i | z_0 \rangle + \\ &\sum_{1 \leq |w| \leq k} \langle G_i | wz_0 \rangle \langle C_{U_i}(h) | w \rangle + \\ &(\langle G_i | z_1 \rangle + \sum_{1 \leq |w| \leq k} \langle G_i | wz_1 \rangle \langle C_{U_i}(h) | w \rangle) U_i(h) \end{aligned} \quad (32)$$

For instance, for order $k = 3$

$$\begin{aligned} Y_i(h) &= y(t_i + h) \\ &= y(t_i) + \langle G_i | z_0 \rangle h + \langle G_i | z_0^2 \rangle h^2 / 2 + \\ &\langle G_i | z_1 z_0 \rangle \psi_{i,2}(h) + \langle G_i | z_0^3 \rangle h^3 / (3!) + \\ &\langle G_i | z_1 z_0^2 \rangle \psi_{i,3}(h) \end{aligned} \quad (33)$$

and

$$\begin{aligned} Y_i^{(1)}(h) &= y^{(1)}(t_i + h) \\ &= \langle G_i | z_0 \rangle + \langle G_i | z_0^2 \rangle h + \\ &\langle G_i | z_1 z_0 \rangle \psi_{i,1}(h) + \langle G_i | z_0^3 \rangle h^2 / 2 + \\ &\langle G_i | z_1 z_0^2 \rangle \psi_{i,2}(h) \end{aligned} \quad (34)$$

- 3) And we use the algorithm of section 4 by specifying t_i at every step. So we obtain the drawing of $y(t)$.

C. Contribution of symbolic computing

The symbolic computing allows us to profit from the genericity and from the precision.

- 1) Genericity : We propose that one uses the formal expression of the generating series G_i and of the output $y(t_i)$ and its derivative $y^{(1)}(t_i)$. Then we replace successively the expressions by their values at every step.
- 2) Precision : We can choose any order k for approximating the output and its derivative. The error is on the order of $k + 1$.

D. Comparison with other methods

The main interest of this method consists in choosing the precision, not only by the size of the time interval h but by the order of the approximation.

The quality of any approximation depends on the order, the

size of the interval but also depends on the roughness of the curve and the stability of the system [1]. When the system is stable, the drawing of the curve is suitable, by using our method, for a large period of time and a small order. In this case, our method is favourable. Otherwise we have to reduce the period of time in order to follow the true curve.

In comparison with Runge-Kutta methods, this method consists in selecting a much smaller number of steps, the local curve being acquired on every interval.

REFERENCES

- [1] Benmakrouha F., Hespel C., *Generating series for the study of stability of bilinear systems*, 23rd IFIP TC7 Conference on System Modelling and Optimization, Cracow, July 2007.
- [2] Chen K.T., *Iterated path integrals*, Bull. Amer. Math. Soc. 83, pp. 831-879, 1977.
- [3] Fliess M., *Fonctionnelles causales non linéaires et indéterminés non commutatives*, Bull. Soc. Math. France 109, pp. 3-40, 1981.
- [4] Fliess M., Reutenauer C., *Théorie de Picard-Vessiot des systèmes réguliers*, Colloque Nat.CNRS-RCP 567, Outils et Modèles Mathématiques pour l'Automatique, l'Analyse des Systèmes et le Traitement du Signal, Belle-Ile, sept. 1982.
- [5] Fliess M., Lamnabhi M., Lamnabhi-Lagarrigue F., *An algebraic approach to nonlinear functional expansions*, IEEE Trans. Circuits and Systems CAS-30, pp. 554-570, 1983.
- [6] Hespel C., *Iterated derivatives of a nonlinear dynamic system and Faà di Bruno formula*, Mathematics and Computers in Simulation, vol.42, pp. 641-657, 1996.
- [7] Hespel C., *Une étude des séries formelles non commutatives pour l'Approximation et l'Identification des systèmes dynamiques*, Thèse d'état, Université de Lille 1, 1998.

Generating series : a combinatorial computation

Benmakrouha Farida, Hespel Christiane

Computer Science Department INSA - IRISA

20 Avenue des Buttes de Coësmes CS 14315 35043 Rennes Cedex, France

I. INTRODUCTION

The purpose of this paper is to apply combinatorial techniques for computing coefficients of rational formal series (G_k) in two noncommutative variables and their differences at order k and $k-1$. This in turn may help one to study validation of a family (B_k) of bilinear systems, described by the series (G_k) and global modeling of an unknown dynamical system (Σ) .

The model validation is a central problem in system identification [2]. In almost cases, the model validation consists, in a test that falsifies or not falsifies the model, using a validation data set.

Computing and bounding these differences, we propose an estimation of the error due to approximations by (B_k) . This error computation is a sum of differential monomials in the input functions and behavior system. We identify each differential monomial with its colored multiplicity and analyse our computation in the light of the free differential calculus.

We propose also a combinatorial interpretation of coefficients of (G_k) , according to [12]. These coefficients are powers of an operator Θ which is in the monoid generated by two linear differential operators Δ and Γ .

The n -th power of Θ is equal to the sum of the labels of all forests of colored increasing trees.

This error computation allows one to better measure the impact of noisy inputs on the convergence of (B_k) . Indeed, one can determine the contribution of the inputs and of the system in the error computation.

II. A LOCAL MODELING OF THE UNKNOWN SYSTEM

The problem consists in modeling an unknown dynamic system (Σ) for $t \in [0, T] = \bigcup_{i \in I} [t_i, t_i + d]$, when knowing some correlated sets of input/output.

We construct a behavioral model, based on the identification of its input/output functional (the generating series), in a neighborhood of every t_i , up to a given order k [1], [4]. At once a local modeling by a bilinear system $(B_i)_k$ around every

t_i is provided. Then a family $((B_i)_{i \in I})_k$, global modeling of the unknown system is produced, such that the outputs of (Σ) and $((B_i)_{i \in I})_k$ coincide up to order k .

III. THE BILINEAR SYSTEM

We consider a certain class (\mathcal{GP}) enclosing the electric equation

$$y^{(1)}(t) = f(y(t)) + u(t) \quad (1)$$

where $u(t)$ is the input function Σ , the unknown system is an affine system. In this case, equation (1) can be written

$$(\Sigma) \quad \begin{cases} \dot{x} &= A_0(x) + A_1(x)u(t) \\ y(t) &= x(t) \end{cases}$$

- $u(t)$ is the real input
- $x(t)$ is the current state
- $A_0 = a^{(0)} \frac{\partial}{\partial x}$ where $a^{(0)} = f(x)|_{x(0)}$
- $A_1 = \frac{\partial}{\partial x}$

The class (\mathcal{GP}) encloses the nonlinear differential equation relating the current excitation $i(t)$ and the voltage $v(t)$ across a capacitor [9]

$$v^{(1)} + k_1 v + k_2 v^2 = i(t)$$

Let $a^{(i)} = f^{(i)}(x)|_{x(0)}$

We notice that the fundamental formula [9] provides the following bilinear system (B_k) , approximating at order k :

$$\begin{cases} \dot{x}_k(t) &= (M_0 + M_1 u(t)) x_k(t) \\ \bar{y}_k(t) &= \lambda x_k(t) \end{cases}$$

where $\lambda = (x(0) \quad 1 \quad 0 \dots 0)$

$$x_k(0) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$M_0 = (C_{z_0 z_1^k})$ (resp $M_1 = (C_{z_1^{k+1}})$) expressed in basis $(C_{z_i^k})$.

$$M_0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ a^{(0)} & a^{(1)} & a^{(2)} & \cdots & a^{(k)} \\ 0 & a^{(0)} & 2a^{(1)} & \cdots & 0 \\ 0 & 0 & a^{(0)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$M_1 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

So, at order k , we obtain the i th derivative of the state vector x as a function of the previous ones. Our solution consists of two steps : to compute $x_{(k-n)k}^{(k-n+i)}(0)$ and to compute the difference of the i th derivative $x_{2k}^{(i)}(0) - x_{2(k-1)}^{(i)}(0)$.

IV. FIRST STEP : COMPUTATION OF $x_{(k-n)k}^{(k-n+i)}(0)$

By derivating and term's regrouping, we can show that :

$$\begin{aligned} & x_{(k-n)k}^{(k-n+i)}(0) \\ = & \\ & \sum_{m=1}^{\min(i+1, k-1)} a^{(m)} \sum_{l=1}^{k-n-1} \binom{k-n-l+m-1}{m} \\ & (a^{(0)} + u^{(0)})^{l-1} x_{(k-n-l+m)k}^{(k-n-l+m+i-m)}(0) \\ + & \\ & \sum_{m=1}^{i+1} u^{(m)} \sum_{l=1}^{k-n-2} \binom{k-n-l+i}{m} \\ & (a^{(0)} + u^{(0)})^{l-1} x_{(k-n-l)k}^{(k-n-l+i-m)}(0) + 1 \\ & \text{(if } m = i + 1) \end{aligned}$$

We analyze now these equations in the light of the free differential calculus. Considering the derivative $a^{(i)}$ and $u^{(i)}$ specialized in time $t=0$ as differential letters, it is clear that our computation is a sum of differential monomials in a and u .

A. Colored partitions and multiplicities

A number partition or multiplicity is a sequence $\mu = (\mu_1, \mu_2, \mu_3, \dots)$ (often written as $1^{\mu_1} 2^{\mu_2} 3^{\mu_3} \dots$) of nonnegative integers. On a single letter a , the differential monomials become :

$$a^\mu = (a^{(i_1)})^{e_1} (a^{(i_2)})^{e_2} \dots (a^{(i_q)})^{e_q},$$

$$1 \leq i_1 < i_2 < \dots < i_q$$

Such a monomial is indexed by the following partition [10] :

$$\mu = (i_1^{\mu_{i_1}} i_2^{\mu_{i_2}} \dots i_q^{\mu_{i_q}})$$

Let $C = \{a, u\}$ be a set of two colors. We call colored partition on C an element of the free monoid generated by the cartesian product $N \times N$ i.e. any finite sequence of couples of nonnegative integers

$$\mu = ((\mu_1^a, \mu_1^u), (\mu_2^a, \mu_2^u), \dots)$$

So, a colored partition μ will denote the differential monomial

$$a^\mu = (a^{(i_1)})^{e_1} \dots (a^{(i_p)})^{e_p} (u^{(j_1)})^{f_1} \dots (u^{(j_q)})^{f_q}$$

$$1 \leq i_1 < i_2 < \dots < i_p, \quad 1 \leq j_1 < i_2 < \dots < j_q$$

where e_l (resp f_l) = $\mu_{i_l}^a$ (resp $\mu_{j_l}^u$). The weight and the size of μ are defined as follows :

$$wgt(\mu) = \sum_c \sum_k k \cdot \mu_k^c$$

$$size(\mu) = \sum_c \sum_k \mu_k^c$$

The empty partition is noted ϵ .

If L is the set of colored partitions, we define a partial order \ll on L :

$$\nu = \{(\nu_i^a, \nu_i^u)\} \ll \mu = \{(\mu_i^a, \mu_i^u)\}$$

if

$$\nu_i^a \leq \mu_i^a \quad \text{and} \quad \nu_i^u \leq \mu_i^u \quad \forall i$$

L , with this partial ordering forms a Young lattice. [11]

We consider now B_i a subset of L defined by :

$$\{\mu / wgt(\mu) = i\}$$

and we note $I(\mu_{max})$ the order ideal generated by μ_{max} , if

$$\mu_{max} = \max(\mu / \mu \in B_i)$$

B. Combinatorial analysis of our computation

Let us now interpret combinatorially our computation by identifying each differential monomial with its colored multiplicity. The recursive relation is captured by the operation :

$$\mu_{max} \odot c = \sum_{\substack{\nu \in I(\mu_{max}) \\ wgt(\nu) = j \leq i}} c^{(i-j+1)} \cdot \nu$$

By factorizing according to the colored partitions, we get :

$$x_{(k-n)k}^{(k-n+i)} = \sum_c \sum_{\substack{\nu \in I(\mu_{max}) \\ wgt(\nu) = j \leq i}} c^{(i-j+1)} \cdot \nu \cdot g_{(c^{(i-j+1)} \nu)}^1$$

where :

$$g_{a^{(m)} \nu}^l = (a^{(0)} + u^{(0)})^{m+1} \sum_{p=m}^{n_l+m} \binom{l}{m} g_p^v$$

and

$$g_{u^{(m)}\nu}^l = (a^{(0)} + u^{(0)})^{m-1} \sum_{p=1}^{n_l} \binom{l+i+1}{m} g_v^p$$

with $n_1 = k - n - 1$, $n_l = l \quad \forall l > 1$

$$g_\epsilon = 1$$

C. Computation of $x_{(k-n)k}^{(k-n+i)}(0)$

We consider now permutations of a colored partition μ on an alphabet $X = \bigcup_{c \in C} X_c$. A permutation [11] of μ is a word in which each letter belongs to X and for each $x_i \in X$, the total number of appearances of x_i in the word is μ_i^c , for some $c \in C$

Let us note $\pi = \xi_1 \xi_2 \dots \xi_{size(\mu)}$ a permutation of μ and σ_μ the set of permutations of μ .

Since, our alphabet

$$X_a = \{a^{(p)} | p = 1, \min(k-1, i+1)\}$$

and

$$X_u = \{u^{(p)} | p = 1, i+1\}$$

$$\xi_j = c^{(i_j)}$$

, for some (c, i_j) .

$x_{(k-n)k}^{(k-n+i)}$ is a linear combination of monomial $y_1^{\lambda_1} \dots y_n^{\lambda_n}$ ($y_i \in X_a \cup X_u$) and all distinct monomials obtained from it by a permutation of variables.

We get finally, if $s = (\sum_j j | \mu_j^u \neq 0)$ and $r = size(\mu)$

$$x_{(k-n)k}^{(k-n+i)} = \sum_{wgt(\mu)=i+1} \mu \cdot (a^{(0)} + u^{(0)})^{k-n+i-r-s} g_\mu^r$$

$$g_\mu^r = \sum_{\pi \in \sigma_\mu} A_1 \prod_{j=2}^r A_j + b \quad)$$

where:

$$A_j = \begin{cases} \sum_{m_j=i_j}^{m_{j-1}+i_j} \binom{m_j}{i_j} & \text{if } \xi_j = a^{(i_j)} \\ \sum_{m_j=1}^{m_{j-1}} \binom{m_j+i-j+2}{i_j} & \text{if } \xi_j = u^{(i_j)} \end{cases}$$

$$A_1 = \begin{cases} \sum_{m_1=m}^{k-n-2+m} \binom{m_1}{i_1} & \text{if } \xi_1 = a^{(i_1)} \\ \sum_{m_j=1}^{k-n-2} \binom{m_1+i+1}{i_1} & \text{if } \xi_1 = u^{(i_1)} \end{cases} \quad g_\mu^1 \text{ defined previously.}$$

and $b = 1$ if $\xi_1 = u^{(i+1)}$, 0 otherwise.

Remark : $x_{(k-n)k}^{(k-n+i)}$ is not a symmetric polynomial even if its structure is the same, because input and system contributions are different.

V. SECOND STEP: COMPUTATION OF

$$x_{2k}^{(k+i)}(0) - x_{2(k-1)}^{(k+i)}(0)$$

The first derivative coincide up to order $k-2$, but at order $k-1$, we have

$$x_{2k}^{(k-1)} - x_{2(k-1)}^{(k-1)} = 0 \text{ and } x_{jk}^{(k-1)} - x_{j(k-1)}^{(k-1)} \neq 0.$$

Let M (resp P) the set of partitions on the single letter a (resp u)

W_i a subset of M defined by

$$\{\nu | 1 \leq size(\nu) \leq i+2\}$$

V_i a subset of P defined by

$$\{\lambda | size(\lambda) = \lfloor \frac{i}{2} \rfloor, wgt(\lambda) \leq i-2 \text{ or } \lambda = u^{(i-2)} \text{ or } \lambda = u^{(i-1)}\}$$

and S_l a subset of L defined by

$$\{\mu | wgt(\mu) = l\}$$

We define now an operation $\nabla : M \times P \times L \mapsto L$

$$\nabla(\nu, \lambda, \mu) = ((\nu_i + \mu_i^a, \lambda_i + \mu_i^u))_i$$

and a subset P_t of $L \quad \forall 0 \leq t \leq i$

$$P_t = \{\tau = \nabla(\nu, \lambda, \mu) | \mu \in S_t, \lambda \in V_i, \nu \in W_i, wgt(\tau) = k+i-1\}$$

We obtain, by a straightforward computation :

$$x_{2k}^{(k+i)} - x_{2(k-1)}^{(k+i)} = \sum_{\substack{\nabla(\nu, \lambda, \mu) \in P_t \\ 0 \leq t \leq i}} \nabla(\nu, \lambda, \mu) \quad h_\nu \cdot f_\lambda \cdot g_\mu^1 \cdot (a^{(0)} + u^{(0)})^{k+i-2-r}$$

where

$$f_\lambda = \sum_{\pi \in \sigma_\lambda} \prod_{l=1}^{size(\lambda)} \binom{k+i-2l}{k+i-2l-i_j}$$

$$h_\nu = \begin{cases} \sum_{\pi \in \sigma_\nu} \prod_{j=1}^{r-2} \binom{i_j+i_{j+1}-1}{i_{j+1}} \binom{k-2}{i_r} & \text{if } size(\nu) \neq 1 \\ 1 & \text{if } size(\nu) = 1 \end{cases}$$

with

$$r = size(\nu)$$

$$r_1 = r + size(\mu)$$

$$s = (\sum_j j | \mu_j^u \neq 0)$$

$$\pi = \xi_1 \xi_2 \dots \xi_r$$

$$\xi_j = c^{(i_j)}$$

$$\sigma_\nu^1 = \{\pi \in \sigma_\nu | \pi \neq \nu_1 \cdot \mu, size(\nu_1) < size(\nu)\}$$

$$\text{and } \pi \neq (a^{(1)})^{r-1} \cdot \xi_r\}$$

Taking into account that $\bar{y}_k^{(i)}(0) = x_{2k}^{(i)}(0)$, we obtain a right computation of the output's difference at order k and $k-1$. By majorization of these output's differences, and when k tends towards infinity, we get an overestimation of the error due to approximation by the (B_k)

VI. COEFFICIENTS OF THE GENERATING SERIES

We give, in this section, a combinatorial interpretation of coefficients of generating series. In [12], the author define increasing trees, model used to describe powers of a differential linear operator. We extend this concept to multiple operators by introducing colored increasing trees.

A. Forest of colored increasing trees

A forest of increasing trees on $\{1, \dots, n\}$, according to [12], is a set of rooted increasing trees, the set of vertices of which is exactly $[n]$ and such each vertex is smaller than all its successors. To take into account the multiplicity of operators,, we use a notion of a ‘‘colored partition’’ ([11]). For each vertex i , we color i any one of c_i colors. Let C the set of colors. We define colored increasing trees on cartesian product $\{1, \dots, n\} \times C$.

B. Combinatorial interpretation

The author shows that the n -power of a linear differential operator is equal to the sum of the labels of all forests of increasing trees on $\{1, \dots, n\}$. So, in our case, the label of a forest on $\{1, \dots, n\} \times C$ is a noncommutative monomial and is defined as :

$$\prod_{(i,c) \in \{1, \dots, n\} \times C} P^{(\alpha(i,c))} \frac{\partial^k}{\partial q}$$

where

$$P(q) = 1 \text{ or } P(q) = a^{(0)}$$

$\alpha(i, c)$ is the number of sons of the node (i, c)

k is the number of trees of the forest.

C. Application

We consider the class $(\mathcal{G}P)$ given in the previous section. According to ([5]), the coefficients of the generating series are :

$$\langle G \mid z_{i_1} z_{i_2} \dots z_{i_k} \rangle = [A_{i_1} \circ A_{i_2} \circ \dots \circ A_{i_k} \circ h(q)]_0$$

where :

$$A_{i_j} = a^{(0)} \frac{\partial}{\partial q}$$

or

$$A_{i_j} = \frac{\partial}{\partial q}$$

Let us define two differential operators

$$\Delta = a^{(0)} \frac{\partial}{\partial q}$$

$$\Gamma = \frac{\partial}{\partial q}$$

These coefficients are powers of an operator Θ which is in the monoid generated by the two linear differential operators Δ and Γ . $C = \{c_1, c_2\}$

The 2-power of operator Θ is :

$$\Theta^2 = \langle G \mid z_1 z_0 \rangle + \langle G \mid z_0 z_1 \rangle + \langle G \mid z_0 z_0 \rangle + \langle G \mid z_1 z_1 \rangle$$

The colored increased trees are :



The labels of these trees are monomials $P^{(0)} P^{(1)} \frac{\partial}{\partial q}$, $P^{(0)^2} \frac{\partial^2}{\partial q^2}$

Each colored vertex is associated to $P(q) = 1$ or $P(q) = a^{(0)}$

We note that, since the observation function $h(q)$ is the identity function, all the powers of $\frac{\partial^n}{\partial q^n}$, $n \geq 2$ are zero.

VII. CONCLUSION

The validation which is presented in this paper is not statistical. It consists in valuing the convergence of a bilinear models family (B_k) on the unknown system (Σ) by an effective symbolic computation. It displays the respective contributions of the input and of the system itself.

More than a symbolic validation, these computing tools are parameterized by the input and the system's behavior. They can particularly provide a valuation process for rough and oscillating inputs as well as for smooth inputs.

REFERENCES

- [1] F. Benmakrouha, C. Hespel, G. Jacob, E. Monnier *Algebraic Identification algorithm and application to dynamical systems* CASC'2001, The 4th International Workshop on Computer Algebra in Scientific Computing
- [2] B.Ninness, G C.Goodwin *Estimation of Model Quality* 10th IFAC Symposium on System Identification, Copenhagen July 1994.
- [3] A.Juditsky, H.Hjalmarsson, A.Benveniste, B.Delyon, L.Ljung, J.Sjoberg, Q.Zhang, *Nonlinear black-box modeling in system identification:mathematical foundations*, Automatica, 31, 1995.
- [4] Benmakrouha F., Hespel C., Jacob G., Monnier E., *A formal validation of Algebraic Identification algorithm: example of Duffing equation*, IMACS ACA'2000, Saint Petersburg, june 25-28, 2000.
- [5] Fliess M., *Fonctionnelles causales non linaires et indetermines non commutatives*, Bull. Soc. Math. France 109, pp. 3-40, 1981.
- [6] Fliess M., *Sur certaines familles de series formelles*, These d'etat, Universite de Paris 7, 1972.
- [7] Hespel C., *Une etude des series formelles non commutatives pour l'Approximation et l'Identification des systemes dynamiques*, These d'etat, Universite de Lille 1, 1998.
- [8] Benmakrouha F., Hespel C., Monnier E. ‘‘Comparison of Identification Methods based on Fuzzy Systems and on Algebraic Model’’ 2001 WSES International Conference on Fuzzy sets and Fuzzy Systems, Tenerife, Feb. 2001.
- [9] M. Fliess, M.Lamnabhi, F. Lamnabhi-Lagarrigue *An Algebraic approach to nonlinear functional expansions* IEEE Trans. Circuits and Systems, vol. CAS-30, n^o 8,1983, . 554-570.

- [10] I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, 2d ed., Oxford Science Publications, 1995.
- [11] G.E. Andrews *The theory of Partitions* Encyclopedia of Mathematics and its applications, Addison-Wesleys, 1984
- [12] F. Bergeron, C Reutenauer, *Combinatorial interpretation of the powers of a linear differential operator* Rapport de recherche Université du Québec Montréal. Mars 1986.

On approximation of nonlinear generating series by rational series

Mikhail V. Foursov and Christiane Hespel

Abstract—In his article we propose an improvement of the method of identification and modeling of dynamical systems of Hespel–Jacob in the case of a single input. Our new method allows one to construct, whenever possible, the unique bilinear system of minimal rank satisfying all conditions obtained during the process of identification of the coefficients of the generating series of the dynamical system. Most importantly, we can unambiguously recognize a rational power series of rank r from the information obtained from the first $2r - 1$ derivatives of the output of the dynamical system.

Index Terms—Formal power series, Hankel matrices, dynamical systems, finite analysis of dynamical systems, generating series, identification of dynamical systems, modeling of dynamical systems.

I. INTRODUCTION

The causal input/output functionals can be described by a certain noncommutative formal power series: the generating (or Fliess) series. The generating series is a canonical representation of the causal functional, in the sense that different functionals have different generating series. The functional corresponding to a generating series is obtained as a product with another noncommutative power series depending on the input: the Chen series.

If the system of equations defining a causal functional is not known, we may consider it as a black box [12], [14] and identify the coefficients of the generating series from the input/output behavior. It was shown by Hespel and Jacob that it is possible to identify the coefficients of the generating series G using a sufficient number of appropriate correlated input/output sets and their derivatives, up to an arbitrary order k [10], [11]. The proof is of a combinatorial nature, the coefficient of the generating series being binomial coefficients.

Once a generating series is identified up to order k , it is possible to construct a rational series of minimal rank that coincides with it up to order k [7], [8], [9]. A rational series corresponds to a bilinear dynamical system that can be constructed using the dependencies between the columns of its Hankel matrix. As a result, the method of Hespel–Jacob allows one to construct a bilinear system that approximates an unknown system with an error of $O(t^k)$.

Since the combinatorial explosion makes it difficult to identify the coefficients of high order, it would be quite

Mikhail V. Foursov is with IRISA/Université de Rennes–1, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France, e-mail : mikhail.foursov@irisa.fr

Christiane Hespel is with IRISA/INSA de Rennes, 20, avenue des Buttes de Coësmes, 35043 Rennes Cedex, France, e-mail : christiane.hespel@insa-rennes.fr

Manuscript submitted April 2nd, 2009

interesting to better profit from all the information available to us and identify more generating series coefficients. Moreover, in general, there may be more than one rational series of minimal rank that coincides with a given nonlinear generating series up to order k .

In this article, we thus propose to reduce as much as possible the problem of choosing one of those rational series, for the case of systems with a single input. The main idea is to use the partial information about the coefficients of orders greater than k that was obtained during the identification. Indeed, during the modeling step, one uses only the values of the coefficients of orders up to k . However, some of linear combinations of the coefficients of higher order were also identified at the identification step. We propose thus an algorithm that uses this additional information in order to give the rational series that fits best to the known data. In the cases when the series is rational of rank r such that the output derivatives of orders up to $2r - 1$ were used during the identification, we show that this rational series can be uniquely determined.

II. PRELIMINARIES

By a dynamical system we will mean an affine system of ordinary differential equations of the form

$$(\Sigma) \quad \begin{cases} \dot{\mathbf{q}}(t) = \mathbf{v}_0(\mathbf{q}) + \sum_{j=1}^m \mathbf{v}_j(\mathbf{q})u_j(t), \\ y(t) = h(\mathbf{q}(t)), \end{cases} \quad (\text{II.1})$$

where

- 1) $\mathbf{u}(t) = (u_1(t), \dots, u_m(t))$ is the input vector,
- 2) $\mathbf{q}(t) \in \mathcal{M}$ is the current state, where \mathcal{M} is a real differential manifold,
- 3) $\{\mathbf{v}_0, \dots, \mathbf{v}_m\}$ is a family of smooth vector fields on \mathcal{M} ,
- 4) $h : \mathcal{M} \rightarrow \mathbb{R}$ is a smooth function called the observation map,
- 5) $y(t) \in \mathbb{R}$ is the output function.

We will be working with the *causal functional* that associates to the set of m input functions (commands) $\mathbf{u}(t)$ the corresponding output function $y(t)$. To the commands $u_1(t), u_2(t), \dots, u_m(t)$ we associate an *alphabet* $\mathcal{Z} = \{z_0, z_1, \dots, z_m\}$ of $(m+1)$ letters, z_0 being associated to the drift (which we will represent as an additional constant input function $u_0(t) \equiv 1$). To every multi-index $I = (i_1, i_2, \dots, i_k)$ we associate a word $w = z_I = z_{i_1}z_{i_2} \dots z_{i_k}$. These words form \mathcal{Z}^* , the free monoid over \mathcal{Z} . (The empty word is denoted by ε .)

The behavior of causal functionals is uniquely described by two noncommutative power series: the generating series and the Chen series.

The *generating series* $G = \sum_{w \in \mathcal{Z}^*} \langle G|z_I \rangle z_I$ [4] is the geometric contribution and it is independent of the input. Its coefficients $\langle G|z_I \rangle$ are obtained by iteratively applying Lie derivatives corresponding to the vector fields to the observation map and evaluating the resulting expression at the initial state \mathbf{q}_0 :

$$\langle G|z_I \rangle = \langle G|z_{i_1} z_{i_2} \cdots z_{i_k} \rangle = \mathbf{v}_{i_1} \circ \mathbf{v}_{i_2} \circ \cdots \circ \mathbf{v}_{i_k} \circ h \Big|_{\mathbf{q}_0}.$$

The generating series completely describes the causal functional. More precisely, two formal power series define the same functional if and only if they are equal [5], [16].

The *Chen series* $C_u(t) = \sum_{w \in \mathcal{Z}^*} \langle C_u(t)|z_I \rangle z_I$ measures the input contribution [1], [2], and is independent of the system. The coefficients of the Chen series are calculated recursively by integration using the following two relations:

- $\langle C_u(t)|\varepsilon \rangle = 1$,
- $\langle C_u(t)|w \rangle = \int_0^t \langle C_u(\tau)|v \rangle u_j(\tau) d\tau$ for a word $w = vz_j$.

The causal functional $y(t)$ is then obtained locally as the product of the generating series and the Chen series [6]:

$$y(t) = \langle G|C_u(t) \rangle = \sum_{w \in \mathcal{Z}^*} \langle G|w \rangle \langle C_u(t)|w \rangle \quad (\text{II.2})$$

This formula is known as the *Peano–Baker formula*, as well as the *Fliess’ fundamental formula*. Differentiating (II.2), we obtain

$$\frac{d^n y(t)}{dt^n} = \sum_{w \in \mathcal{Z}^*} \langle G|w \rangle \left\langle \frac{d^n}{dt^n} C_u(t) \Big| w \right\rangle. \quad (\text{II.3})$$

Note that only the time derivatives of the Chen series appear in this expression. Their exact or at least recursive formula is needed. It can be shown in a straightforward way that [15]

$$\frac{d^n}{dt^n} C_u(t) = C_u(t) A_n(t), \quad (\text{II.4})$$

where the noncommutative polynomials $A_n(t)$ are recursively defined by the following relations

$$A_0(t) = 1, \quad A_{n+1}(t) = \mathcal{L}_u(t) A_n(t) + \frac{d}{dt} A_n(t), \quad (\text{II.5})$$

where $\mathcal{L}_u(t) = \sum_{z_i \in \mathcal{Z}} u_i(t) z_i$. Thus we finally obtain that

$$\left. \frac{d^n}{dt^n} y(t) \right|_{t=0} = \sum_{w \in \mathcal{Z}^*} \langle G|w \rangle \langle A_n(0)|w \rangle \quad (\text{II.6})$$

The *Hankel matrix* of a formal power series G is an infinite matrix with columns and rows indexed by the monomials from \mathcal{Z}^* ordered lexicographically, such that the entry on the intersection of the row u and the column v is $\langle G|uw \rangle$.

Theorem II.1. *A (real-valued) formal power series is recognizable if and only if its Hankel matrix has finite rank [3].*

Theorem II.2. *A (real-valued) formal power series is rational if and only if it is recognizable [13].*

The following result due to Fliess is also important in this article.

Theorem II.3. *If all the rows (columns) corresponding to monomials of a certain fixed length are linear combinations of previous rows (columns), then all the following rows (columns) are also linear combinations thereof.*

Corollary II.4. *If a rational series in two variables is of rank n , then the upper left block of the Hankel matrix of size $(2^n - 1) \times (2^n - 1)$ is of rank n .*

Algorithm II.5. *The method of Hespel–Jacob consists in two steps: identification of the coefficients of the generating series and construction of a bilinear model. A short description follows. For a complete description, see [7], [8], [9], [10], [11].*

- **Identification.**

The derivatives of the output are linear in the generating series coefficients and polynomial in the inputs u_j and their derivatives (cf. (II.6)). Choosing appropriate input/output sets, certain linear combinations of the generating series coefficients can be identified (those are exactly the coefficients of the monomials in the inputs and their derivatives).

On the second stage of identification, the identified linear combinations are used to find the generating series coefficients themselves. Identification of coefficients of orders up to k can be done using the output derivatives of orders up to $m = k + \lfloor k/2 \rfloor (k - \lfloor k/2 \rfloor)$, in the case of a series in two letters. However, not all the linear combinations are used during this step, but only those involving the coefficients of orders $\leq k$. The remaining linear combinations give only partial information about the individual coefficients.

- **Modeling.**

During the identification, the coefficients of the generating series were identified up to order k . These values are inserted into the Hankel matrix whose column basis is then calculated. The bilinear model is of the form

$$\begin{cases} \dot{\mathbf{x}}(t) = (M_0 + \mathbf{u}(t)M_1)\mathbf{x}(t), \\ \mathbf{x}(0) = x_0, \end{cases}$$

where M_0 and M_1 are matrices that are computed by expressing, in terms of the basis vectors, the left–multiplicative action of the letters of \mathcal{Z} on the basis vectors. As the Hankel matrix is not completely determined, one obtains multi–parameter families of linear combinations of the basis vectors. The algorithm proposes to choose the linear combination that depends on the leftmost basis vectors.

We note that some partial information obtained during the identification is not used in the modeling. The main goal of this article is to fill this gap and to try to construct the unique model that fits best to all the available information. However, the unique model exists only if the generating series is rational of an appropriate rank. In the other cases, the modeling method we present here still has an advantage, as it allows one to construct a bilinear approximating system of rank r using the output derivatives of orders up to $2r - 1 \leq k + \lfloor k/2 \rfloor (k - \lfloor k/2 \rfloor)$. Thus the identification can be done using fewer input/output sets, which is quite important since

the necessary number of input/output sets grows exponentially with the increase of the order of differentiation of the output.

In order to simplify the explications, we will only deal with column operations. However, in practice, a more efficient strategy is to mix the column and row operations. In other words, our MAPLE package constructs both a row basis and a column basis of the Hankel matrix and uses both column and row dependencies during its “filling in”. Even though it is possible to rely exclusively on column operations, the algorithm is simpler for a mixed strategy.

III. IDENTIFICATION OF FORMAL POWER SERIES OF RANK 2

Corollary III.1. (of theorem (II.3)) *If the Hankel matrix is of rank 2, then either $\{C_\varepsilon, C_{z_0}\}$ or $\{C_\varepsilon, C_{z_1}\}$ span the space of column vectors.*

Theorem III.2. *A rational generating series of rank 2 can be uniquely identified from the output derivative conditions (II.6) of orders up to 3.*

Proof:

Without loss of generality, we can assume that $\{C_\varepsilon, C_{z_0}\}$ span the column space. If not, we can interchange z_0 and z_1 . Let $\lambda = (g_\varepsilon, g_0)$, $M_0 = \begin{pmatrix} 0 & a \\ 1 & b \end{pmatrix}$ and $M_1 = \begin{pmatrix} c & d \\ e & f \end{pmatrix}$. Let us consider the following system of equations, equivalent to the conditions obtained from the output derivative conditions of orders up to 3:

$$\begin{cases} \lambda M(z)\gamma = g; (i = \varepsilon, z_0, z_1, z_0^2, z_0z_1, z_1z_0, z_1^2, z_0^3, z_1^3), \\ \lambda(M(z_0^2z_1) + M(z_0z_1z_0) + M(z_1z_0^2))\gamma = g_{001} + g_{010} + g_{100}, \\ \lambda(M(z_0z_1^2) + M(z_1z_0z_1) + M(z_1^2z_0))\gamma = g_{011} + g_{101} + g_{110}, \end{cases} \quad (\text{III.1})$$

Here $M(z_{i_0} \cdots z_{i_k}) = M_{i_0} \cdots M_{i_k}$, the coefficient of the rational series corresponding to a bilinear system of the type (II.7).

Appending to it the conditions of vanishing of all 3×3 minors of the Hankel matrix involving the coefficients of orders up to 3, we obtain a system in the unknowns $\{a, b, c, d, e, f\}$. Using Gröbner basis techniques, it is possible to show (it takes a significant amount of time) that it admits a solution for any value of the coefficients g_i on the right-hand side of the equations (of course satisfying the minor vanishing conditions). Moreover, the solution is unique under an additional assumption that the rank of the Hankel matrix is strictly greater than 1. ■

Solving complicated systems of polynomial equations is a very useful tool from a theoretical point of view, but it becomes practically unfeasible as the number of equations and unknowns increases. However, we do not need to solve the most general system in every case. Firstly, one does not need to consider the most general case. Most of the generating series coefficients involved in the system are found during the identification. The system (III.1) can then be solved almost instantaneously. Secondly, the algorithm II.5 allows one to find the two-parameter family of rational series of rank 2 having the given coefficients of orders up to 2. Substituting it into the system (III.1) eliminates even more variables and any example of rank 2 can be easily solved this way. However, the systems

one has to solve become more difficult for higher-rank cases. We would like thus to propose a different method that involves solving mostly linear equations.

Algorithm III.3. *The algorithm consists in a loop that includes the following three main steps.*

- 1) *Identify dependencies between the columns and use them to fill in a part of the matrix.*
- 2) *Solve the system of linear equations obtained during the identification, and substitute the solution into the Hankel matrix (thus some entries will be linear expressions of other entries).*
- 3) *(necessary only at ranks 4 and higher) Find a parameter (or a linear combination of parameters) in such a way that the rank of the Hankel matrix is “too high” for all but one value of this parameter (or linear combination of parameters). “Too high” means that it is greater than $\lfloor (m+1)/2 \rfloor$ when m is the maximal order of output differentiation used during identification.*

For generating series of rank 2, there are 3 different possible scenarios :

- $\{C_\varepsilon, C_{z_0}\}$ and $\{C_\varepsilon, C_{z_1}\}$ both form bases of the column space and $\{R_\varepsilon, R_{z_0}\}$ and $\{R_\varepsilon, R_{z_1}\}$ both form bases of the row space. In this case, filling in the matrix using the column dependencies allows one to find both unknown parameters of the two-parameter family.
- only one of $\{C_\varepsilon, C_{z_0}\}$ and $\{C_\varepsilon, C_{z_1}\}$ is a base of the column space or only one of $\{R_\varepsilon, R_{z_0}\}$ and $\{R_\varepsilon, R_{z_1}\}$ is a base of the row space. In this case, filling in the matrix allows one to find one of the unknown parameters. The other one is found from the equations (II.6).
- only one of $\{C_\varepsilon, C_{z_0}\}$ and $\{C_\varepsilon, C_{z_1}\}$ is a base of the column space and only one of $\{R_\varepsilon, R_{z_0}\}$ and $\{R_\varepsilon, R_{z_1}\}$ is a base of the row space. In this case, filling in the matrix is not sufficient to find any unknown parameters. But it allows one to diminish the number of unknown coefficients of the generating series and to find the unknown parameters one by one.

The algorithm is very easy in each case. But since complete explanations will not be feasible in higher-rank cases, we will illustrate in detail our techniques here, for cases 1 and 3.

Example III.4. *Let us first consider the following (bilinear) dynamical system which is an example of case 1.*

$$\begin{cases} \dot{q}_1 = -q_2, \\ \dot{q}_2 = q_1 + q_2 + (2q_1 + q_2)u, \\ y(t) = q_1(t) + 2q_2(t). \end{cases} \quad q_1(0) = 1, \quad q_2(0) = 0. \quad (\text{III.2})$$

Using the algorithm II.5, we obtain the following information at order 3 of differentiation:

$$\begin{aligned} \langle G|\varepsilon \rangle &= 1, \langle G|z_0 \rangle = 2, \langle G|z_1 \rangle = 4, \langle G|z_0^2 \rangle = 1, \\ \langle G|z_0z_1 \rangle &= 2, \langle G|z_1z_0 \rangle = 2, \langle G|z_1^2 \rangle = 4, \langle G|z_0^3 \rangle = -1, \\ \langle G|z_0^2z_1 + z_0z_1z_0 + z_1z_0^2 \rangle &= -3, \\ \langle G|z_0z_1^2 + z_1z_0z_1 + z_1^2z_0 \rangle &= 0, \langle G|z_1^3 \rangle = 4. \end{aligned}$$

Its Hankel matrix is thus as follows (with monomials ordered lexicographically) :

$$\begin{pmatrix} 1 & 2 & 4 & 1 & 2 & 2 & 4 & \dots \\ 2 & 1 & 2 & -1 & -3-x_1-x_2 & x_1 & -y_1-y_2 & \\ 4 & 2 & 4 & x_2 & y_1 & y_2 & 4 & \\ 1 & -1 & -3-x_1-x_2 & & & & & \\ 2 & x_1 & -y_1-y_2 & & & & & \\ 2 & x_2 & y_1 & & & & & \\ 4 & y_2 & 4 & & & & & \\ \dots & & & & & & & \end{pmatrix}$$

(where x_1, x_2, y_1, y_2 are yet unknown values). The rank of this matrix should be 2. Using the algorithm II.5, we obtain $\{C_\varepsilon, C_{z_0}\}$ as the basis of the column space, as well as

$$\lambda = (1 \quad 2), \mu(z_0) = \begin{pmatrix} 0 & 1-2a \\ 1 & a \end{pmatrix}, \mu(z_1) = \begin{pmatrix} 0 & 2-2b \\ 2 & b \end{pmatrix},$$

where a and b are unknown parameters (recall that the algorithm does not use any coefficients of order 3 since most of them were not identified yet). The available information on the third-order terms allows us to conclude immediately that we also have $C_{z_0^2} = C_{z_1} - C_\varepsilon$ and $C_{z_1^2} = 2C_\varepsilon$.

Using the known values of the Hankel matrix, these two relations between the columns together with $C_{z_1} = 2C_{z_0}$ and their consequences, we obtain additional relations $-3 - x_1 - x_2 = 4$, $-y_1 - y_2 = 2x_1$, $y_1 = 2x_2$, $x_2 = -2$ and $8 = 2y_2$. Solving these equations, we obtain all the coefficients of order 3. The Hankel matrix is now

$$\begin{pmatrix} 1 & 2 & 4 & 1 & 2 & 2 & 4 & -1 & -2 & 1 & 2 & -2 & -4 & 2 & 4 \\ 2 & 1 & 2 & -1 & -2 & 1 & 2 & & & & & & & & \\ 4 & 2 & 4 & -2 & -4 & 2 & 4 & & & & & & & & \\ 1 & -1 & -2 & & & & & & & & & & & & \\ 2 & 1 & 2 & & & & & & & & & & & & \\ 2 & -2 & -4 & & & & & & & & & & & & \\ 4 & 2 & 4 & & & & & & & & & & & & \\ \dots & \dots & \dots & & & & & & & & & & & & \end{pmatrix}$$

giving us $a = 1$ and $b = 1$. The rank 2 rational series is thus completely determined. Constructing the bilinear system corresponding to it, we obtain the system (III.2). Let us remark that it was a different system that was found by the original method, using the same information. The rational generating series corresponding to (III.2) is

$$G = 1 + (2 - z_0)(z_0 + z_1 - (z_0 + 2z_1)z_0)^*(z_0 + z_1)$$

Remark. Of course, if the additional relations were contradictory, we would conclude that the rank of the series was greater than 2, and use only some additional conditions to find the values of a and b .

Example III.5. Let us now consider an example of case 3:

$$\begin{cases} \dot{q}_1 = q_1 + q_2, & q_1(0) = 1, \\ \dot{q}_2 = q_2 + (q_1 + 2q_2)u, & q_2(0) = 0. \\ y(t) = q_1(t) + 2q_2(t). \end{cases} \quad (\text{III.3})$$

At order 3 of differentiation of the output, we have identified

the following:

$$\begin{aligned} \langle G | \varepsilon \rangle &= 1, \langle G | z_0 \rangle = 1, \langle G | z_1 \rangle = 2, \langle G | z_0^2 \rangle = 1, \\ \langle G | z_0 z_1 \rangle &= 3, \langle G | z_1 z_0 \rangle = 2, \langle G | z_1^2 \rangle = 4, \langle G | z_0^3 \rangle = 1, \\ \langle G | z_0^2 z_1 + z_0 z_1 z_0 + z_1 z_0^2 \rangle &= 9, \\ \langle G | z_0 z_1^2 + z_1 z_0 z_1 + z_1^2 z_0 \rangle &= 16, \langle G | z_1^3 \rangle = 8. \end{aligned}$$

Its Hankel matrix is thus of the following form

$$\begin{pmatrix} 1 & 1 & 2 & 1 & 3 & 2 & 4 & \dots \\ 1 & 1 & 3 & 1 & 9-x_1-x_2 & x_1 & 16-y_1-y_2 & \\ 2 & 2 & 4 & x_2 & y_1 & y_2 & 8 & \\ 1 & 1 & 9-x_1-x_2 & & & & & \\ 3 & x_1 & 16-y_1-y_2 & & & & & \\ 2 & x_2 & y_1 & & & & & \\ 4 & y_2 & 8 & & & & & \\ \dots & \dots & \dots & & & & & \end{pmatrix}$$

By the algorithm II.5 we obtain, taking the basis $\{C_\varepsilon, C_{z_1}\}$

$$\lambda = (1 \quad 2), \mu(z_0) = \begin{pmatrix} 1 & 3-2a \\ 0 & a \end{pmatrix}, \mu(z_1) = \begin{pmatrix} 0 & 4-2b \\ 1 & b \end{pmatrix},$$

where a and b are again some unknown parameters. Using the only known relation $C_{z_0} = C_\varepsilon$ and its consequences, we obtain additional equations $x_1 = 3$, $x_2 = 2$ and $y_2 = 4$. This allows us to conclude that $C_{z_0 z_1} = C_\varepsilon + C_{z_1}$ and thus $a = 1$. This last relation between the columns implies in its turn that $y_1 = 6$ and thus $b = 2$. The rank 2 rational series is now completely determined:

$$G = \left(1 + 2(z_0 + 2z_1)^* z_1\right) \left(z_0 + z_0(z_0 + 2z_1)^* z_1\right)^*.$$

IV. IDENTIFICATION OF FORMAL POWER SERIES OF RANK 3

The rank 3 case would be rather similar to the rank 2 case, except for the fact that we may not know the complete basis of the column space at the beginning. However, the following easy proposition guarantees that at least 2 basis vectors are known.

Proposition IV.1. The part of the Hankel matrix of a rational series of rank 3, constructed using the coefficients obtained from derivatives of orders up to 5 of the output, cannot be of rank 1.

Theorem IV.2. All the coefficients of a rational power series of rank 3 can be uniquely determined from the information obtained from the output derivative conditions (II.6) of orders up to 5.

Proof:

This theorem, which is a counterpart to (III.2), cannot be realistically proven by solving a system of polynomial equations, since the corresponding system is quite complicated in this case. Solving this polynomial system is still feasible for a given series if we use the information obtained from the algorithm II.5.

However, the general proof can rely on the new techniques described in this paper and it can be done by considering separately different cases arising during the identification. We give its outline here. Further details can be easily filled in.

Now, considering separately all the different cases (for different possibilities of basis vectors and basis rows), we see that at most 1 minimization is necessary in order to complete the identification of the parameters. Indeed, the only coefficients of order 4 that were not identified at order 7 are those of the monomials involving 2 occurrences of z_0 and 2 of z_1 . There are 6 such coefficients in all, but only 5 equations to find them. However, the minimization technique allows us to find one of these coefficients. The remaining ones are then immediately found from the 5 identified linear combinations. Once this step is finished, the remaining steps are quite similar to the rank 3 case and are executed without difficulties. ■

Lemma V.2. *Let H be the Hankel matrix corresponding to a rational series of rank 4, whose coefficients were identified from the output derivative condition of orders up to 7. Then there exists a yet unidentified coefficient g_i with the property that the rank of H is greater or equal to 5, unless g_i is equal to a certain value k .*

Proof:

The lemma is again proven on a case-by-case basis. ■

Remark. The disadvantage of the minimization part is its nonlinearity, but it can still be efficiently implemented on computer, since it involves only one symbolic parameter.

Example V.3. *Consider the rational generating series for the bilinear system with $\lambda = (1 \ 1 \ 1 \ 1)$,*

$$\gamma = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, M_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, M_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

At the order 7 of differentiation of the output, all the known coefficients are equal to 1. The yet unknown conditions for the coefficients of the fourth-order terms give $g_{0011} = a$, $g_{0101} = 7 - 5a$, $g_{0110} = 4 - 3a$, $g_{1001} = 4 - 3a$, $g_{1010} = 6 - 5a$, $g_{1100} = a$, where a is an unknown parameter.

Since all the known coefficients are equal to 1, no dependency between the columns can be identified at this stage. As a consequence, we cannot obtain any other coefficient yet. Studying the rank of the Hankel matrix, we see that its rank is $r \geq 4$ for $a = 1$, $r \geq 8$ for $a \neq 1$ and $r \geq 9$ for $a = (9 \pm \sqrt{1105})/32$ (no matter what are the values of the coefficients of orders 5 and higher). Therefore, since we are trying to construct a rational series of rank ≤ 4 , $a = 1$ is the only value that could eventually allow us to obtain a matrix of rank 4. Once we choose $a = 1$, four independent columns are immediately found and the rest of the matrix is easily filled in.

VI. IDENTIFICATION OF SERIES OF RANK GREATER THAN

4

The identification algorithm works essentially in the same way as in the rank 4 case, except that there are no direct counterparts of the lemma V.2. However, it can be replaced with a following strategy, applied as many times as all the other strategies fail. It is based on the following conjecture based on strong experimental evidence :

Conjecture VI.1. *Let H be the Hankel matrix corresponding to a rational series of rank $r \geq 5$, whose coefficients were identified from the output derivative conditions of orders up to $2r - 1$. Then there exists a combination of unidentified coefficients $s = \sum_j g_{I_j}$ with the property that the rank of H is greater than r , unless s is equal to a certain value k .*

The rank minimization algorithm is as follows :

- 1) Let n be the length of the shortest word in the generating series (the smallest according to the lexicographical ordering) whose coefficient is unknown.
- 2) Let $m = \sum_i a_i g_i$ be the sum of the coefficients of order n that are not identified yet and a_i unknown constants.
- 3) We initialize the stack with $(\{m = 0\}, C_\varepsilon)$, the basis with C_ε (since it can be always considered as a part of the basis) and the current column with $\tilde{C} = C_{z_0}$.
- 4) During each iteration of the loop, one considers the column \tilde{C} .

- If the basis contains more than r columns, pop the stack and obtain the last-in element (s, C_K) . Solve the system s , replace the obtained values in the Hankel matrix and exit the loop.
- If \tilde{C} is the column corresponding to a word of length $> r$, we encountered a linear combination that does not work. Let (s, C_K) be the element on top of the stack. We pop the stack and reset \tilde{C} to the column that follows C_K .
- We write the (truncated) column C_J as a linear combination (with arbitrary coefficients) of the basis columns and solve the corresponding system of equations augmented with the system s , where (s, C_K) is the element on the top of the stack.
 - If the system is incompatible, we add this column to the basis.
 - If there is one solution m' that involves only linear combinations of g_i and such that not all a_i vanish, we append m' to s and push (s, C_J) onto the stack.
 - In all the other cases, no action needs to be done on this stage.

Finally, we reset \tilde{C} to the column that follows C_J .

Conjecture VI.2. *A rational series of rank r (or less) can be uniquely identified from the information obtained from the output derivative conditions (II.6) up to order $2r - 1$.*

Proof:

Due to a large number of different cases, it is not possible to clearly identify all the different possible scenarios and give a complete proof similar to the rank 2, 3 and 4 cases. Extensive experimental evidence shows however clearly that the system of rank k can be identified in all the cases. ■

VII. CONCLUSION

In this article, we propose an improvement of the method of Hespel–Jacob for modeling nonlinear dynamical systems in the case of a single input. Among its main advantages are: the use of all the information obtained during the identification,

which implies a better approximation and the possibility to precisely identify a rational series whenever there is sufficient data. Moreover, we have bounded by $2k - 1$ the order of differentiation of the output derivatives which is necessary for identifying a rational series of order k . This algorithm was successfully programmed and tested in MAPLE.

An interesting direction to pursue is to generalize this algorithm to the case of several inputs.

REFERENCES

- [1] K.-T. Chen, *Algebras of iterated path integrals and fundamental groups*, Trans. Am. Math. Soc. **156** (1971), 359–379.
- [2] K.-T. Chen, *Iterated path integrals*, Bull. Am. Math. Soc. **83** (1977), 831–879.
- [3] M. Fliess, *Matrices de Hankel*, J. Math. Pure Appl. **53** (1974), 197–222.
- [4] M. Fliess, *Fonctionnelles causales non linéaires et indéterminées non commutatives*, Bull. Soc. Math. France **109** (1981), 3–40.
- [5] M. Fliess, *On the concept of derivatives and Taylor expansions for nonlinear input/output systems*, in “IEEE Conference on Decision and Control” (San Antonio, Texas), 1983, 643–648.
- [6] M. Fliess, *Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices*, Invent. Math. **71** (1983), 521–537.
- [7] C. Hespel and G. Jacob, *Approximation of nonlinear systems by bilinear ones*, in “Algebraic and geometric methods in nonlinear control theory” (M. Fliess and M. Hazewinkel, eds.), Reidel, 1986, 511–520.
- [8] C. Hespel and G. Jacob, *Calculs des approximations locales bilinéaires de systèmes analytiques*, RAIRO APPII **23** (1989), 331–349.
- [9] C. Hespel and G. Jacob, *Approximation of nonlinear dynamic systems by rational series*, Theor. Comp. Sci. **79** (1991), 151–162.
- [10] C. Hespel and G. Jacob, *First steps towards exact algebraic identification*, Discrete Math. **180** (1998), 211–219.
- [11] C. Hespel and G. Jacob, *On algebraic identification of causal functionals*, Discrete Math. **225** (2000), 173–191.
- [12] A. Juditsky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, and Q. Zhang, *Nonlinear black-box models in system identification: mathematical foundations*, Automatica **31** (1995), 1725–1750.
- [13] M.P. Schützenberger, *On the definition of a family of automata*, Inform. and Control **4** (1961), 245–270.
- [14] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorenec, H. Hjalmarsson, and A. Juditsky, *Nonlinear black-box modeling in systems identification: a unified overview*, Automatica **31** (1995), 1691–1724.
- [15] H.J. Sussman, *A product expansion for the Chen series*, in “Theory and applications of nonlinear control systems” (C.I. Byrnes and A. Lindquist, eds.), North-Holland, 1986, 323–335.
- [16] Y. Wang and E.D. Sontag, *On two definitions of observation spaces*, Syst. Contr. Lett. **13** (1988), 279–289.
- [17] Y. Wang and E.D. Sontag, *Algebraic differential equations and rational control systems*, SIAM J. Contr. Optim. **30** (1992), 126–149.
- [18] Y. Wang and E.D. Sontag, *Generating series and nonlinear systems; analytic aspects, local realizability, and I/O representations*, Forum Math. **4** (1992), 299–322.

Counting Rooted and Unrooted Triangular Maps

Samuel Vidal and Michel Petitot

Abstract—In this paper, we describe a new way to count isomorphism classes of rooted triangular maps and unrooted triangular maps. We point out an explicit connection with the asymptotic expansion of the Airy function. The analysis presented here is used in a recent paper “Vidal (2007)” to present an algorithm that generates in optimal amortized time an exhaustive list of triangular maps of a given size.

Index Terms—rooted triangular maps, unrooted triangular maps, generating functions, Airy function, cycle index series

INTRODUCTION

Triangulations of surfaces constitute an important data structure in computer graphics as they provide a handy discrete model of surfaces. It has proven invaluable for instance to model the shape of objects in computed graphics. From the point of view of computer science, the applications of surface triangulations are well known and numerous, they touch both practical and theoretical aspects of the discipline and they range from computer graphics to discrete methods of solving partial differential equations. They also play a central role in many algorithms of computational geometry, a fast growing subject having an heavy industrial impact as it is used in computer aided design.

One particularly interesting treat of the subject, apart from its broad range of applications, is precisely its ubiquity both in computer science, mathematical physics and even pure mathematics, providing generous range of fruitful exchange between seemingly remote parts of science. From the point of view of mathematics, the theory of combinatorial maps is also a venerable subject dating back to Cayley and Hamilton. Since those times, it generated an impressive amount of results of all sorts concerning the particular enumeration problem of counting the *rooted* combinatorial maps. Those results came from various communities of researchers, each with its own methods and tradition. Among them, *enumerative combinatorists* of course played a significant role, starting with pioneering works by Tutte [17] on rooted planar maps. Those works where at first motivated by the four color problem. *Theoretical physicists* also played a significant role, starting with the work by t’Hooft [16] on integration on random matrix spaces and Feynman diagrams. Pure mathematicians like Harer and Zagier [5] also have contributed to the theory in connection with cutting edge algebraic geometry problems concerning moduli spaces of Riemann surfaces. Last but not least, one must mention in mathematical physics the Witten-Kontsevich model of quantum gravity [7] using in a central fashion the higher combinatorics of triangular maps and trivalent diagrams.

Although a lot is known concerning the theory of *rooted* combinatorial maps, *very little* is currently known about the outstanding problem of enumeration of *unrooted* combinatorial

maps up to isomorphism except for planar maps with the pioneering work of Liskovets [9]. It appears as a very difficult problem of combinatorics, which stayed barely untouched for almost 150 years. As a matter of fact, the only general result on those important objects were up to now contained in the recent paper by Mednykh and Nedela [14]. In section II-D of this paper, we give our first contribution to this problem, namely in the form of a generating series giving the number of *unrooted* triangular maps (c.f. series (20) on page 4).

In this paper, a *triangular map* is a triangulation of a (not necessarily connected) oriented surface without boundary, and its *size* is an integer divisible by 6, hence of the form $n = 6k$, such that the triangulation has $2k$ triangular faces and $3k$ edges. Apart from those unrooted map enumeration results the most interesting theorems of this article are the following.

Theorem 1: Let a_n be the number of *labelled* triangular maps of size n . Then, $a_n = 0$ if n isn’t a multiple of 6 and,

$$a_{6k} = \frac{(6k)!}{k!} \left(\frac{1}{6}\right)_k \left(\frac{5}{6}\right)_k 6^k \quad (1)$$

where $(x)_k = x(x+1)\dots(x+k-1)$ is the Pochhammer symbol. Therefore, the exponential generating series of the a_n is hypergeometric and divergent. We have,

$$\sum_{n \geq 0} \frac{a_n}{n!} z^n = {}_2F_0 \left(\begin{matrix} \frac{1}{6}, \frac{5}{6} \\ - \end{matrix} \middle| 6z^6 \right) \quad (2)$$

$$= \sum_{k \geq 0} \frac{(\frac{1}{6})_k (\frac{5}{6})_k}{k!} 6^k z^{6k} \quad (3)$$

Theorem 2: Let b_n be the number of *pointed connected unlabelled* triangular maps of size $6n$. Then b_n satisfies the following recurrence equation,

$$b_{n+1} = (6n+6)b_n + \sum_{k=1}^{n-1} b_k b_{n-k} \quad (4)$$

with $n \geq 1$ and $b_1 = 5$.

Those two theorems provide a connection with the asymptotic expansion of the Airy function as explained in section II-C.

I. COUNTING PRINCIPLE FOR TRIANGULAR MAPS

An *oriented* surface without boundary is described by a finite set of triangular faces. The orientation of the surface is given by a normal unitary vector on each point of the surface, the normal vector varying continuously with respect to the point. This normal vector induce a cyclic order on the tree edges belonging to a same triangle.

Any triangulation is transformed by a classical duality (due to Poincaré), to a triangular diagram (see figures 1 and 2). Faces of the triangulation become black vertices of

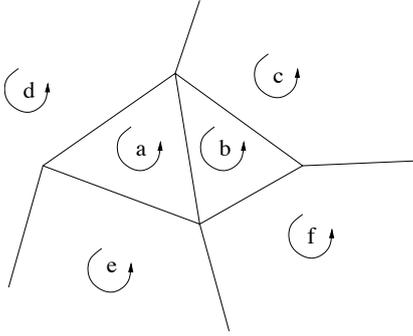


Fig. 1. A triangulation of the Riemann sphere made of $2k$ triangles (a, b, c, d, e, f). [case $k = 3$]

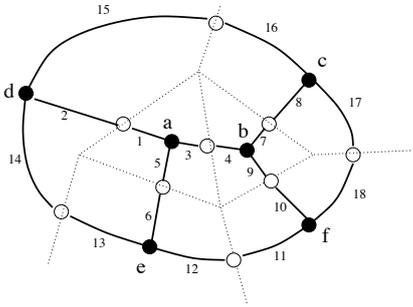


Fig. 2. The associated triangular diagram made of $n = 6k$ edges, $2k$ black vertices and $3k$ white vertices.

the diagram while edges of the triangulation become white vertices of the diagram. In such a diagram, every edges is adjacent to one black vertex and one white vertex. Black vertices have degree 3 (i.e. are adjacent to three edges). White vertices have degree 2 (i.e. are adjacent to 2 edges). Every three edges adjacent to a same black vertex are cyclically ordered. An isomorphism between two diagrams is given by a bijection which transforms edges and vertices of one of the diagrams into edges and vertices of the second, preserving color of the vertices and cyclic orientation of the edges.

In what follows, we shall describe a triangulations by a couple of two permutations (σ_0, σ_1) . The permutation σ_0 exchanges the two edges adjacent to a same white vertex while the σ_1 permutation cyclically permutes the three edges adjacent to a same black vertex. For example the two permutations corresponding to figure 2, are the following,

$$\begin{aligned} \sigma_0 &= (1, 2)(3, 4)(5, 6)(7, 8)(9, 10) \\ &\quad (11, 12)(13, 14)(15, 16)(17, 18), \\ \sigma_1 &= (1, 5, 3)(2, 15, 14)(4, 9, 7) \\ &\quad (6, 13, 12)(10, 11, 18)(8, 17, 16). \end{aligned}$$

Definition 1 (triangular map): A triangular map of size $n \in \mathbb{N}$ is given by a finite set of edges labelled from 1 to n and by a pair of permutations (σ_0, σ_1) belonging to \mathfrak{S}_n such that the cycles of σ_0 are only of length 2 and that the cycles of σ_1 are only of length 3.

Two maps (σ_0, σ_1) and $(\bar{\sigma}_0, \bar{\sigma}_1)$ both of size n are called *isomorphic* if there exists a permutation $\tau \in \mathfrak{S}_n$ such that, $\bar{\sigma}_0 = \tau \circ \sigma_0 \circ \tau^{-1}$ and $\bar{\sigma}_1 = \tau \circ \sigma_1 \circ \tau^{-1}$, this is an exact translation of the notion of diagram isomorphism defined above.

A. Species used in this paper and related generating series

In the sequel, we use the following species (see appendix A) :

- 1) The species of sets, denoted by E , which associates to any labelling set U the singleton $E[U] := \{U\}$. For every set U there is a unique structure which is U itself.
- 2) The species of permutations, denoted by S , which associates to any labelling set U the set $S[U]$ composed of permutations of U . For every relabelling $\sigma : U \rightarrow U$ and every permutation $\tau \in S[U]$, one put $S[\sigma](\tau) := \sigma\tau\sigma^{-1}$.
- 3) The species of cycles of length n , denoted by C_n , which associates to any labelling set U the set $C_n[U]$ composed of cyclic permutations of U of length exactly n . One put $C_n[U] = \emptyset$ when $\text{card } U \neq n$.
- 4) The species S_n of permutations having only cycles of length n .
- 5) The species of triangular maps, denoted by T^* , which associates to any labelling set U , the set $T^*[U]$ of triangular maps whose edges are labelled by U .
- 6) The species of connected triangular maps, denoted by T , which associates to any labelling set U the set $T[U]$ of connected triangular maps whose edges are labelled by U .
- 7) The species of pointed connected triangular maps, denoted by T^\bullet , having a distinguished edge.

The considered species are related by the following relations,

$$\begin{cases} S_2 = E(C_2) \\ S_3 = E(C_3) \\ T^* = S_2 \odot S_3 \\ T^\bullet = E(T) \end{cases} \quad (5)$$

In intuitive terms, this means,

- A permutation of S_2 decompose uniquely in a set of cycles of length 2.
- A permutation of S_3 decompose uniquely in a set of cycles of length 3.
- A triangular map is uniquely determined by a couple of permutation $(\sigma_0, \sigma_1) \in S_2 \times S_3$ acting simultaneously on a set of labels.
- A triangular map is uniquely decomposed in a set of connected triangular map.

The set of species permits to derivate in an automatic fashion from the equations (5), the following relations between

generating series,

$$\begin{cases} S_2(z) = \exp C_2(z) \\ S_3(z) = \exp C_3(z) \\ T^*(z) = S_2(z) \odot S_3(z) \\ T(z) = \log T^*(z) \\ T^\bullet(z) = z \frac{d}{dz} T(z) \end{cases} \quad (6)$$

II. COUNTING TRIANGULAR MAPS

A. Labelled triangular maps

In this section, we are counting the triangular maps having n edges labelled by numbers from 1 to n .

Proof of theorem 1 – There are $(n-1)!$ labelled cycles of length n then $C_n(z) = \frac{z^n}{n}$. This enables to compute the following generating series,

$$S_2(z) = \exp \frac{z^2}{2} = \sum_{k=0}^{\infty} \frac{z^{2k}}{2^k k!}, \quad (7)$$

$$S_3(z) = \exp \frac{z^3}{3} = \sum_{k=0}^{\infty} \frac{z^{3k}}{3^k k!}. \quad (8)$$

Let's study the coefficients of the series $T^*(z) = S_2(z) \odot S_3(z)$. In $T^*(z)$, the coefficient of z^n vanish when n is not a multiple of 6. One has,

$$[z^{6n}] T^* = \frac{1}{2^{3n} 3^{2n}} \frac{(6n)!}{(2n)! (3n)!}, \quad (n \geq 0).$$

Let's put $T^*(z) = f(6z^6)$. Let's show that the series $f(x) = \sum_{n \geq 0} f_n \frac{x^n}{n!}$ is hypergeometric. Using formula (21), the computation of f_n yields,

$$f_n = \frac{1}{6^n 2^{3n} 3^{2n}} \frac{n! (6n)!}{(2n)! (3n)!}, \quad (n \geq 0). \quad (9)$$

By a straightforward computation one deduces that the f_n coefficients satisfy the following linear recurrence,

$$\frac{f_{n+1}}{f_n} = (n+a_1)(n+a_2) \text{ where } a_1 = \frac{1}{6}, a_2 = \frac{5}{6}. \quad (10)$$

As $f(0) = 1$, one deduces that $f(x) = {}_2F_0 \left(\begin{matrix} a_1, a_2 \\ - \end{matrix} \middle| x \right)$ which ends the demonstration. \square

B. Pointed connected triangular maps

Proposition 1 ([23] lemma 1.4): The species T^\bullet of pointed connected triangular maps is rigid (see definition 3), or equivalently $\widetilde{T}^\bullet(z) = T^\bullet(z)$.

Proof: Let ϕ be an automorphism of a pointed connected diagram (Γ, a) preserving basepoint, i.e. such that $\phi(a) = a$. Let A be the set of edges of Γ left invariant under ϕ i.e. edges x such that $\phi(x) = x$. Our proposition is proved if we show that every edge of Γ belongs to A . As ϕ preserves adjacency and cyclic orientations around vertices, if $x \in A$ then every edges adjacent to x must be preserved by ϕ . Hence by induction the condition propagates to the full connected component of x . As $a \in A$ then A is nonempty, and as Γ is connected then every edges of Γ belongs to A , which ends the proof. \blacksquare

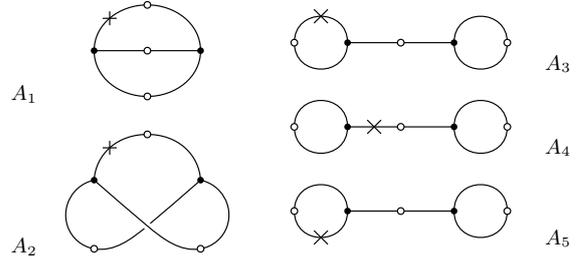


Fig. 3. As an example, this figure gives the 5 pointed connected triangular maps having 6 edges. This corresponds to $b_1 = 5$.

Proof of theorem 2 – From the linear recurrence 10, the hyper geometric series $f(x)$ satisfies the differential equation $Lf(x) = 0$ where,

$$L := \theta - x(\theta + a_1)(\theta + a_2), \quad \theta = x \frac{d}{dx}. \quad (11)$$

Using the change of variables $x = 6z^6$, the equations (6) implies,

$$\begin{aligned} T^\bullet(z) &= z \frac{d}{dz} T(z) = z \frac{d}{dz} \log T^*(z) \\ &= 6x \frac{d}{dx} \log f(x) = 6v(x). \end{aligned} \quad (12)$$

the function $v(x) := xf'(x)/f(x)$ is solution to the Riccati equation associated to the linear equation (11),

$$v - x[\theta v + v^2 + (a_1 + a_2)v + a_1 a_2] = 0. \quad (13)$$

From this equation, one deduce that the coefficients v_n of the series $v(x) = \sum_{n \geq 0} v_n x^n$ satisfies the following quadratic recurrence (where δ is the Kronecker symbol),

$$v_{n+1} = (a_1 + a_2 + n) v_n + \sum_{k=0}^n v_k v_{n-k} + a_1 a_2 \delta_{n,0}$$

with $n \geq 0$. Replacing a_1 et a_2 by their values yields,

$$v_{n+1} = (n+1) v_n + \sum_{k=0}^n v_k v_{n-k} + \frac{5}{36} \delta_{n,0}, \quad (14)$$

with $n \geq 0$ and initial condition $v_0 = 0$. One deduce $v_1 = 5/36$.

The sequence b_n one gets from the sequence v_n while putting One then have $b_n = 6^{n+1} v_n$. The recurrence (14) defining the sequence v_n is equivalent to the recurrence,

$$b_{n+1} = 6(n+1) b_n + \sum_{k=0}^n b_k b_{n-k} + 5 \delta_{n,0} \quad (15)$$

with $n \geq 0$ and initial condition $b_0 = 0$ which defines the sequence b_n of theorem 2. \square

That recurrence (15) gives,

$$\begin{aligned} \widetilde{T}^\bullet(z) = T^\bullet(z) &= 5z^6 + 60z^{12} + 1105z^{18} + 27120z^{24} \\ &+ 828250z^{30} + 30220800z^{36} + \dots \end{aligned}$$

C. Connection with Airy asymptotics

The following special function,

$$\text{Ai}(x) := \frac{1}{\pi} \int_0^{+\infty} \cos\left(\frac{1}{3}t^3 + xt\right) dt \quad (16)$$

due to the british astronomer Airy, find applications in optics. It solves the linear differential equation $y''(x) = xy(x)$. In ordre to numerically compute the zeros of this function, Stokes in 1857, used the following asymptotic expansion

$$\text{Ai}(x) \underset{x \rightarrow +\infty}{\sim} \frac{\exp(-2/3 x^{3/2})}{2\sqrt{\pi} x^{1/4}} {}_2F_0\left(\frac{1}{6}, \frac{5}{6} \mid -\frac{3}{4}x^{-3/2}\right), \quad (17)$$

wich is divergent. It establishes the connection with the generating series (2) of $T^*(z)$.

Reworking the computation of the proof of theorem 2, one proves the following asymptotic expansion,

$$z^{2/3} \frac{\text{Ai}'(1/4 z^{2/3})}{\text{Ai}(1/4 z^{2/3})} + \frac{1}{2}z + 1 \underset{z \rightarrow +\infty}{\sim} \frac{b_1}{z} - \frac{b_2}{z^2} + \frac{b_3}{z^3} + \dots = -T^*(-1/z), \quad (18)$$

where Ai' is the derivative of the Airy function.

D. Connected triangular maps

In this section, we are computing the generating series $\tilde{T}(z)$ which counts connected triangular maps having n edges. Two isomorphic maps are counted once. The species T isn't rigid ; we shall thus use the cycle index series of T through the formula $\tilde{T}(z) = \mathcal{Z}_T(z, z^2, z^3, \dots)$. To that purpose we shall first compute the cycle index series of the specie C_2, C_3 and E which are,

$$\begin{aligned} \mathcal{Z}_{C_2}(z_1, z_2, \dots) &= \frac{1}{2}z_1^2 + \frac{1}{2}z_2 \\ \mathcal{Z}_{C_3}(z_1, z_2, \dots) &= \frac{1}{3}z_1^3 + \frac{2}{3}z_3 \\ \mathcal{Z}_E(z_1, z_2, \dots) &= \exp\left(z_1 + \frac{z_2}{2} + \frac{z_3}{3} + \dots\right) \end{aligned}$$

For the species $S_2 := E(C_2)$, one gets using (23),

$$\begin{aligned} \mathcal{Z}_{S_2}(z_1, z_2, \dots) &= \mathcal{Z}_E\left(\frac{1}{2}(z_1^2 + z_2), \frac{1}{2}(z_2^2 + z_4), \right. \\ &\quad \left. \frac{1}{2}(z_3^2 + z_6), \dots\right) \\ &= \exp\left(\frac{1}{2}(z_1^2 + z_2) + \frac{1}{4}(z_2^2 + z_4) \right. \\ &\quad \left. + \frac{1}{6}(z_3^2 + z_6) + \dots\right) \\ &= A_1(z_1) A_2(z_2) A_3(z_3) \dots \end{aligned}$$

where for all integer $n \geq 1$,

$$A_n(z_n) = \begin{cases} \exp\left(\frac{z_n^2}{2n} + \frac{z_n}{n}\right) & n \equiv 0 \pmod{2} \\ \exp\left(\frac{z_n^2}{2n}\right) & n \equiv 1 \pmod{2} \end{cases}$$

A similar computation gives for the species $S_3 := E(C_3)$,

$$\mathcal{Z}_{S_3}(z_1, z_2, \dots) = B_1(z_1) B_2(z_2) B_3(z_3) \dots$$

where for all integer $n \geq 1$,

$$B_n(z_n) = \begin{cases} \exp\left(\frac{z_n^3}{3n} + \frac{2z_n}{n}\right) & n \equiv 0 \pmod{3} \\ \exp\left(\frac{z_n^3}{3n}\right) & n \not\equiv 0 \pmod{3} \end{cases}$$

The cycle index series $\mathcal{Z}_{S_2}(z_1, z_2, \dots)$ and $\mathcal{Z}_{S_3}(z_1, z_2, \dots)$ are separated (see definition 4). The cycle index series of the species $T^* := S_2 \odot S_3$ is computed using lemma 1,

$$\mathcal{Z}_{T^*}(z_1, z_2, \dots) = (A_1 \odot B_1)(z_1) (A_2 \odot B_2)(z_2) (A_3 \odot B_3)(z_3) \dots$$

The cycle index series of the connected species T is computed using formula (24),

$$\mathcal{Z}_T(z_1, z_2, \dots) = \sum_{k \geq 1} \frac{\mu(k)}{k} \sum_{n \geq 1} \log(A_n \odot B_n)(z_{nk}). \quad (19)$$

The series $\tilde{T}(z) := \mathcal{Z}_T(z, z^2, z^3, \dots)$ is eventually computed using a computer algebra package. One gets,

$$\begin{aligned} \tilde{T}(z) &= 3z^6 + 11z^{12} + 81z^{18} + 1228z^{24} \\ &\quad + 28174z^{30} + 843186z^{36} + 30551755z^{42} \\ &\quad + 1291861997z^{48} + 62352938720z^{54} \\ &\quad + 3381736322813z^{60} + \dots \end{aligned} \quad (20)$$

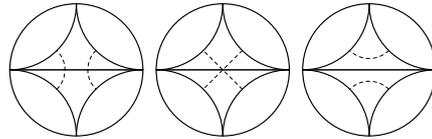


Fig. 4. The three triangular maps with two faces. Using the notion of Poincaré duality described in section I, the diagrams A_1 and A_2 of figure 3 corresponds to the leftmost map and to the middle map respectively while the three diagrams A_3, A_4 and A_5 being conjugated, altogether correspond to the same triangular map on the right.

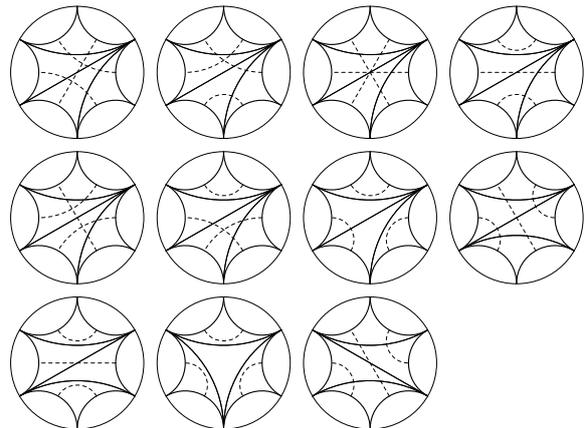


Fig. 5. The eleven triangular maps with four faces. Among them one finds the tetrahedron which can be seen as a triangulation of the Riemann sphere.

A. Definitions

The goal of species theory is to provide a simple way to describe *labelled* structures that are met in computer science. That description style is broader than the pure syntactic style for it explicitly takes account of the relabelling process. It enables to count structures up to isomorphism, which means that two structures are counted as one if they are equal modulo a permutation of their labels.

Definition 2 (species): A species of structure [6], [1] is a functor F which produces

- (i) for all finite label set U , a finite set of structures denoted by $F[U]$,
- (ii) for all bijections $\sigma : U \rightarrow V$, a bijection

$$F[\sigma] : F[U] \longrightarrow F[V].$$

$F[U]$ denotes the set of F -structures labelled by U , σ is any permutation of the labels and $F[\sigma]$ is the related permutation of labelled structures. One assumes that F is a functor, which means that for all permutation $\sigma : U \rightarrow V$ and $\sigma' : V \rightarrow W$, one has $F[\sigma' \circ \sigma] = F[\sigma'] \circ F[\sigma]$. Moreover, for all set U , one has $F[\text{Id}_U] = \text{Id}_{F[U]}$.

B. Generating series

Two F -structures $s_1 \in F[U_1]$ and $s_2 \in F[U_2]$ are called *isomorphic* if there exists a bijection $\sigma : U_1 \rightarrow U_2$ such that $s_2 = F[\sigma](s_1)$.

One usually consider the following series:

- 1) the *exponential* generating series which counts the number $a_n(F)$ of *labelled* F -structures of size n :

$$F(z) = \sum_{n \geq 0} a_n(F) \frac{z^n}{n!}$$

- 2) the *ordinary* generating series which counts the number $\tilde{a}_n(F)$ of *unlabelled* F -structures of size n *i.e.* counted up to isomorphism:

$$\tilde{F}(z) = \sum_{n \geq 0} \tilde{a}_n(F) z^n$$

- 3) the *cycle index* series which counts the number $f_{n_1, n_2, \dots, n_k}(F)$ of F -structures left invariant under a permutation $F[\sigma]$ when the cycle type of the permutation σ is the partition $\lambda := (n_1, n_2, \dots, n_k)$. Equivalently, σ decomposes in n_1 cycles of length 1, n_2 cycles of length 2 etc. One sets for $n = n_1 + 2n_2 + \dots + kn_k$:

$$\begin{aligned} \mathcal{Z}_F(z_1, z_2, \dots) &= \sum_{\sigma} f_{n_1, \dots, n_k}(F) \frac{z_1^{n_1} \dots z_k^{n_k}}{n!} \\ &= \sum_{\lambda} f_{n_1, \dots, n_k}(F) \frac{z_1^{n_1} \dots z_k^{n_k}}{1^{n_1} n_1! \dots k^{n_k} n_k!} \end{aligned}$$

One gets the second formula from the first collecting in the sum all the permutations σ having the same cycle type λ .

One can recover the two other generating series from the cycle index series using,

$$\begin{aligned} F(z) &= \mathcal{Z}_F(z, 0, 0, \dots) \\ \tilde{F}(z) &= \mathcal{Z}_F(z, z^2, z^3, \dots) \end{aligned}$$

1) *Rigidity:* The automorphism group (*i.e.* symmetry group) of a F -structure $s \in F[U]$ is the set of permutation σ of U such that $F[\sigma](s) = s$.

Definition 3 (rigid species): A species F is called *rigid* whenever the group of automorphism of any F -structure is reduced to the identity transformation.

A F -structure is rigid if and only if $a_n(F) = n! \cdot \tilde{a}_n(F)$ for all integer n , which is equivalent to the equality of the two generating series $F(z)$ and $\tilde{F}(z)$.

C. Two basic operations on structure species

1) *Cartesian product (superposition):* The cartesian product $F \odot G$ of two species F and G corresponds to a superposition of two structures labelled over a same alphabet. More precisely, to any set of labels U , one sets $(F \odot G)[U] := F[U] \times G[U]$ and $(F \odot G)[\sigma] := F[\sigma] \times G[\sigma]$ for all permutation $\sigma : U \rightarrow U$.

By definition of the superposition and following the notations above, for all natural number n one has $a_n(F \odot G) = a_n(F) \cdot a_n(G)$ and for all partition $\lambda := (n_1, n_2, \dots, n_k)$ one has $f_{\lambda}(F \odot G) = f_{\lambda}(F) \cdot f_{\lambda}(G)$. At the opposite, for a non-rigid species $\tilde{a}_n(F \odot G) \neq \tilde{a}_n(F) \cdot \tilde{a}_n(G)$, which justifies the use of cycle index series. Translation in the language of generating series is straightforward ; this yields a variation of the Hadamard product,

$$(F \odot G)(z) = \sum_{n \geq 0} a_n(F) a_n(G) \frac{z^n}{n!} \quad (21)$$

$$\begin{aligned} \mathcal{Z}_{F \odot G}(z_1, z_2, \dots) &= \\ &= \sum_{\lambda} f_{\lambda}(F) f_{\lambda}(G) \frac{z_1^{n_1} \dots z_k^{n_k}}{1^{n_1} n_1! \dots k^{n_k} n_k!} \quad (22) \end{aligned}$$

2) *Composition of two species:* Composition of two species F et G will be denoted by $F \circ G$ or else $F(G)$. A $(F \circ G)$ -structure labelled by a set U is given by (see book [2, page 41])

- (i) a partition $\pi = \{U_1, U_2, \dots, U_k\}$ with $U = U_1 \sqcup U_2 \sqcup \dots \sqcup U_k$,
- (ii) a F -structure labelled by the set $\{U_1, U_2, \dots, U_k\}$,
- (iii) a list (s_1, s_2, \dots, s_k) of G -structures labelled by U_1, U_2, \dots, U_k respectively,

which can be summed up by,

$$(F \circ G)[U] = \sum_{\pi \text{ partition of } U} F[\pi] \times \prod_{p \in \pi} G[p].$$

This construction translates to the following relations on generating series,

$$\begin{cases} (F \circ G)(z) = F(G(z)), \\ (\widetilde{F \circ G})(z) = \mathcal{Z}_F(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3, \dots), \\ \mathcal{Z}_{F \circ G}(z_1, z_2, \dots) = \mathcal{Z}_F(g_1, g_2, \dots), \end{cases} \quad (23)$$

using the following conventions $\tilde{g}_k := \tilde{G}(z^k)$ and $g_k := \mathcal{Z}_G(z_k, z_{2k}, z_{3k}, \dots)$ for all integers $k \geq 1$.

The application we have in mind here is the decomposition of a structure in its connected components. One may write the following relation between species $F = E(F^c)$, this is a particular case of the composition of species : E is the species of sets F is a given species and F^c is the species of connected F -structures. In this particular situation, one may use the following formulae (see book [2, page 54]) where μ is the Mobius function,

$$\left\{ \begin{array}{l} F^c(z) = \log F(z) \\ \tilde{F}^c(z) = \sum_{k \geq 1} \frac{\mu(k)}{k} \cdot \log \tilde{F}(z^k) \\ \mathcal{Z}_{F^c}(z_1, z_2, \dots) = \sum_{k \geq 1} \frac{\mu(k)}{k} \cdot \log \mathcal{Z}_F(z_k, z_{2k}, \dots) \end{array} \right. \quad (24)$$

3) Separated series:

Definition 4 (separated series): A cycle index series is said to be separable if

- (i) it can be decomposed in a product of series each one on a single indeterminate, equivalently, if it is of the following form,

$$\mathcal{Z}_F(z_1, z_2, \dots) = F_1(z_1) \cdot F_2(z_2) \cdot F_3(z_3) \cdots$$

- (ii) for all partitions $\lambda := (n_1, n_2, \dots, n_\ell)$, one has

$$f_\lambda(F) = f_{\lambda_1}(F) \cdot f_{\lambda_2}(F) \cdots f_{\lambda_\ell}(F),$$

using the following partitions $\lambda_k = (0, \dots, 0, n_k, 0, \dots, 0)$ for $1 \leq k \leq \ell$.

Lemma 1: Let

$$\begin{aligned} \mathcal{Z}_F(z_1, z_2, \dots) &= F_1(z_1) \cdot F_2(z_2) \cdot F_3(z_3) \cdots, \\ \mathcal{Z}_G(z_1, z_2, \dots) &= G_1(z_1) \cdot G_2(z_2) \cdot G_3(z_3) \cdots, \end{aligned}$$

be two separated cycle index series. Then the series $\mathcal{Z}_{F \odot G}(z_1, z_2, \dots)$ is also separated and we have,

$$\mathcal{Z}_{F \odot G}(z_1, z_2, \dots) = (F_1 \odot G_1)(z_1) \cdot (F_2 \odot G_2)(z_2) \cdots \quad (25)$$

REFERENCES

- [1] F. Bergeron, G. Labelle, and P. Leroux. *Théorie des espèces et combinatoire des structures arborescentes*. LACIM Montréal, 1994.
- [2] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial Species and Tree-like Structures*. Cambridge University Press, 1998. English edition of [1].
- [3] P. Flajolet, B. Salvy, and G. Schaeffer. Airy Phenomena and Analytic Combinatorics of Connected Graphs. *The electronic journal of combinatorics*, 11(34), 2004.
- [4] F. Harary and E.M. Palmer. *Graphical Enumeration*. Academic Press New York/London, 1973.
- [5] J. Harer and D. Zagier. The euler characteristic of the moduli space of curves. *Invent. Math.*, 85:457–486, 1986.
- [6] A. Joyal. Une théorie combinatoire des séries formelles. *Adv. Math.*, 42:1–82, 1981.
- [7] M. Kontsevich. Intersection theory on the moduli space of curves and the matrix airy function. *Commun. Math. Phys.*, 147:1–23, 1992.
- [8] S.K. Lando and A.K. Zvonkin. *Graphs on Surfaces and Their Applications*. Springer-Verlag, 2004.
- [9] V. A. Liskovets. A census of nonisomorphic planar maps. In L. Lovasz and V.T. Sos, editors, *Algebraic methods in graph theory*, volume II, 1981. North-Holland, Amsterdam.

- [10] V. A. Liskovets. Enumeration of nonisomorphic planar maps. *Selecta Math. Sovietica*, 4:303–323, 1985.
- [11] V. A. Liskovets. A reductive technique for enumerating non-isomorphic planar maps. *Discrete Math.*, 156:197–217, 1996.
- [12] V. A. Liskovets. Reductive enumeration under mutually orthogonal group actions. *Acta Appl. Math.*, 52(7):91–120, 1998.
- [13] V. A. Liskovets and T. R. Walsh. Counting unrooted loopless planar maps. *European Journal of Combinatorics*, 26(5):651–663, 2005.
- [14] A. Mednykh and R. Nedela. Enumeration of unrooted maps of a given genus. *J. Comb. Theory, Ser. B*, 96(5):706–729, 2006.
- [15] N. J. A. Sloane. The on-line encyclopedia of integer sequences. Available on the net at : <http://www.research.att.com/~njas/sequences/>, 2005.
- [16] G. t'Hooft. A planer diagram theory for strong interactions. *Nuclear Physics B*, 72:461–473, 1974.
- [17] W. T. Tutte. A census of planar triangulations. *Canad. J. Math.*, 14:21–38, 1962.
- [18] W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963.
- [19] W. T. Tutte. On the enumeration of planar maps. *Bull. Amer. Math. Soc.*, 74:64–74, 1968.
- [20] W. T. Tutte. The enumerative theory of planar maps. In J. N. Srivastava, editor, *A Survey of Combinatorial Theory*, pages 437–448. North-Holland, 1973.
- [21] S. Vidal. Asymptotics of the Airy Function and Triangular Maps Decomposition. In preparation.
- [22] S. Vidal. Multiparametric Enumeration of Unrooted Combinatorial Maps. In preparation.
- [23] S. Vidal. Sur la classification et le dénombrement des sous-groupes du groupe modulaire et de leurs classes de conjugaison. *Pub. IRMA, Lille*, 66(II):1–35, 2006. Preprint : <http://arxiv.org/abs/math.CO/0702223>.
- [24] S. Vidal. An Optimal Algorithm to Generate Pointed Trivalent Diagrams an Pointed Triangular Maps. Preprint : <http://arxiv.org/abs/0706.0969>, 2007.

About a group of Drinfel'd associators

Hoang Ngoc MINH

The biological activity of DNA molecule depends mainly on the way it is arranged in space and the way in which it is twisted things which fall within the province of the mathematical theory of knots [5]. There is a real parallel between mathematical transformations of knots (see the Conway's ip and incrossing elementary operations) and enzymatic mechanisms (the topo-isomerases) [9].

Combinatorics & Schelling's model

Cyril BANDERIER , Hanane TAFAT

Thomas Schelling got the Nobel prize in economic sciences in 2005 for having enhanced our understanding of conflict and cooperation through game-theory analysis. In 1971, in the Journal of Mathematical Sociology, he published "Dynamic Models of Segregation." This seminal article exhibited a very simple model (a variant of cellular automata) which clearly showed that a small racial preference for each individual leads to a society with a high level of segregation. Thought being modelizable via a Markov chain, the model is however not that simple to analyze, So they have being a lot of experimental studies on this subject ; we adopt here an automata theory approach, coupled with generating functions (analytic combinatorics), in order to establish enumerative and asymptotic rigorous results. How many rounds before the periodic regime ? What is the ratio of each species ? What is the length of the period ? What is the migration rate (global displacement) ?

Systemic modelling

Daniel KROB

We will propose an overview of systems architecture & engineering, focusing on the systems modelling aspects. The talk will adress the main aspects of these topics: what is a system ? what are the main issues with respect to systems design ? how to describe a system ? how to model a system ?

Risk and Complexity

Organizing Committee

- Emmanuel ELIOT
University of Le Havre, France
`emmanuel.eliot@univ-lehavre.fr`
- Damienne PROVITOLLO
Sophia-Antipolis University, Nice, France
`Damienne.Provitolo@geoazur.unice.fr`

Description

The main goal of this special session is to join together different disciplines concerned by the analysis of risk in terms of methodologies and management.

The proposed researches contribute to showing that risk and catastrophe are often the outcome of structural complexities related to the number of components and their interactions, of complexities of spatial scales and levels of organization, and of complexity of non-linear systems (A. Dauphiné, 2003). The various research results show that, in a more or less explicit way, these complexities do not exclude one another but may combine in the course of one and the same event. Through its applied character, these works also contribute to territorial risk management.

These researches relate to the themes of natural risk, health risk and technological risk. Two orientations are proposed. First the theme of risk is addressed in terms of hazard, population vulnerability and complex decision-making about a risk or catastrophe (flooding, car travel, crowd movement, panic, population evacuation, pedestrian behaviour in urban transport, population vulnerability to a cholera epidemic, decision-making aids in crisis situations). The simulation models (simulation or GIS) are often proposed as part of a risk-management approach and meet the needs of private or public actors such as the RATP, CODAH or the Haute-Normandie regional council. Secondly, the theme of risk is addressed from a more conceptual angle. One paper looks at the methods of map algebra to create spatial information and new themes in the form of a GRID; another proposes a factual ontology of risk and catastrophe.

The session is organized on a non-thematic basis. In this way the various contributions can be brought together. Three headings are proposed: 1) risk management 2) risk diffusion 3) conceptual models of risk.

Contents

Exploring crowd dynamics based on individual behaviors and local interactions	
Haïfa ABDELHAK , Éric DAUDÉ , Damien OLIVIER	102
Cholera in the 19th century: constructing epidemiological risk with complexity methodologies	
Éric DAUDÉ , Emmanuel ELIOT , Emmanuel BONNET	110
A multiagent urban traffic simulation Part II: dealing with the extraordinary	
Éric DAUDÉ , Pierrick TRANOUEZ , Patrice LANGLOIS	116
Use of geosimulations and the complex system theory to better assess flash floods risks in the Paris Basin watersheds (France)	
Johnny DOUVINET , Daniel DELAHAYE , Patrice LANGLOIS	121
Compiled risks of spatial complexity: the map algebra contribution	
Emmanuel BONNET , Thierry SAINT-GÉRARD , Eliane PROPECK , David GAILLARD	125
Modeling and simulation of pedestrian behaviors in transport areas: the specific case of platform/train exchanges	
Jérémy FIEGEL , Arnaud BANOS , Cyrille BERTELLE	129
Agent-oriented approach for detecting and managing risks in emergency situations	
Fahem KEBAIR , Frédéric SERIN	134
Validation of an ontology of risk and disaster through a case study of the 1923 Great Kanto Earthquake	
Damienne PROVITOLLO , Jean Pierre MÜLLER , Edwige DUBOS-PAILLARD	140
Exploitation of a displacement survey to detect road network use vulnerability	
Michel NABAA , Cyrille BERTELLE , Antoine DUTOT , Damien OLIVIER , Patrick LIONS	151

Exploring crowd dynamics based on individual behaviors and local interactions

Haïfa Abdelhak¹, Eric Daudé², Damien Olivier¹

¹Laboratoire d'Informatique du Havre

University of Le Havre- France

25 rue Philippe Lebon

76600 Le Havre

haifa_molka@hotmail.fr

damien.olivier@univ-lehavre.fr

²UMR 6266 IDEES, University of Rouen

7 rue Thomas Becket, 76821 Mont Saint-Aignan

eric.daude@univ-rouen.fr

Abstract—This paper aims to present an approach to model crowd motion in normal and panic situations in order to study the influence of the individual's decision on a crowd situation. Different aspects of the human behavior are modelled for example the individual and group strategy for evacuation, the negotiation strategy in a blocked situation and the spreading of panic into the crowd. This approach is based on the Individual Based Model (IBM) to study and simulate the complex system such as the social system of crowd motion, and presents another way to study the special spreading of panic into the environment.

The prototypes produced by our approach for crowd simulation in a virtual environment illustrate the importance of the individual's decisions within the crowd and show how disruption affects the efficiency of evacuation.

Keywords—Crowd motion simulation, human behavior, panic, Individual Based Model, Cellular Automata, individual's strategy.

I. INTRODUCTION

Pedestrian dynamics studies have caught the attention of scientists in recent years. In this domain some models study and simulate pedestrian evacuation strategies in panic and normal situations to test their effectiveness and then help decision-makers to reduce human and material losses. In addition, this better understanding of crowd behavior is needed to improve safety procedures in a variety of buildings and areas and should optimize pedestrian motion.

Crowd dynamic is a complex systems where global behavior emerge from local interactions between individuals and between individuals and their environment. In this system of interactions, decisions of one individual can modify environment and in this way affect human and non human entities.. One can schematically distinguish two ways to model pedestrian crowd: the macroscopic model and the microscopic

model which mainly reflect a differentiation in term of geographical scale.

The microscopic class model (e.g Helbing's force social model [1], CA model [2], *etc.*) focuses on individuals and their characteristics as well as their capacities to take decisions in different situations. The global behavior of crowding is then the result of individual's interactions and the resulting emergent behavior.

The macroscopic class model (e.g D. Provitolo's dynamic model for panic propagation [3], Henderson's approach "Fluid approach" [4]) studies the social system as a whole to predict a possible evolution of human motion. Detailed interactions are then looked at and the model's description is shifted toward a description of the dynamic population variation.

In this paper, we present our approach to model pedestrian motion based on the Individual Based Model (IBM). This approach is a microscopic approach according to IBM particularities, so take into account the individual's details and decisions to give a coherent evolution of the system. To simplify the representation of the spatial environment we use a discretized representation of the space based on a grid of cells, where cell state includes the representation of the presence of individuals or other environmental obstacles.

The paper is organized in the following manner. In section II, we present an overview of our model in normal and panic situations with a brief view of the IBM approach. In section III, we analyze the model based on the simulation results. Finally, we provide discussion for the evolution of the approach.

II. THE IBM APPROACH FOR PEDESTRIAN MOTION

Pedestrian dynamic models are used to control and understand crowd behavior. Different types of approaches are developed, some focus on the prediction of a possible crowd evolution, others focus on the observation of the emergent behavior produced by pedestrians' interactions. The latter are called microscopic models. Two types of these models are described in literature: continuous in space like the Helbing's

social force model, and discrete on space like the cellular automata models.

Helbing's studies are based on pedestrian observations in the real world in normal and panic situations. He demonstrated that "The transition between the "rational" normal behavior and the apparently "irrational" panic behavior is controlled by a single parameter, the "nervousness", which influences fluctuation strengths, desired speed..." [1]. Helbing's social force model is based on force terms which influence the evolution of pedestrian motion and where the variation of one parameter may cause the appearance of panic motion.

"Cellular automata (CA) microsimulation is used to model complex behavior and is named after the principle of *automata* (entities) occupying *cells* according to localized neighborhood rules of occupancy" [5]. In fact, the CA models are based on cells changing state what is determine the emergent results, they approximate the more complex models with a minimal set of simple rules. Pedestrians' intelligence is not really modeled in CA approach, other alternative are introduced to model interactions [6] which provide collective effects and self organization.

According to Helbing's studies, the evolution of pedestrian behavior from rational to irrational is caused by the variation of certain pedestrian characteristics. These variations, brought about by a sudden disruption of the environment, influence the reasoning process of the pedestrian and causes him to panic. In fact, pedestrian motion in panic situations is the result of the variations of some distinct individual's characteristics that occur in normal pedestrian motion. Our approach is a model for pedestrian motion in general; we suppose that the disruptions that occur within the environment affect pedestrian characteristics and cause their transition to panic behavior. This approach to model panic is quite realistic because pedestrian panic motion is the result of the feeling fear that obliges pedestrians to modify their priorities and behavior to escape the potential danger.

A. The IBM approach

In our approach we use Individual based models (IBM) which are a part of a multi agent system which includes a computational system composed of entities in interaction to produce a solution. The particularity of IBM is the autonomy of the entities which interact with each other and with their environment to produce a global behavior. Individual based models are used in many social sciences and physical sciences, in fact in all complex systems to investigate the properties that emerge from entities in interaction.[7]

This microscopic approach offers tools to take into account the heterogeneity of the entities, the characteristics of each individual are tracked through time, in contrast with the macroscopic approach where "the characteristics of the population are averaged together and the model attempts to simulate changes in these averaged characteristics for the whole population" [8]. (Graig Reynolds).

The IBM approach provides agent notion to model system's entities (e.g an individual, a wall, a moving obstacle within the environment etc.). Agent can be a mobile or a static entity and

can have reactive or cognitive behavior, it depends on the characteristics of the entity it models.

1) Reactive agents

Reactive agents are autonomous agents with a partial representation of the environment. They choose actions by using their current state and the external stimuli as a reference in the base rule to get the appropriate reaction. The architecture of these agents is based on the perception to produce behaviors called "stimuli-response" behavior. The simplicity of the architecture and the behavior of these agents allow them to model and study complex systems where behaviors emerge from interactions of a large number of entities.

The purely reactive agent's drawback is its lack of adaptability. This type of agent cannot generate a plan to reach one's goal because of its incapacity of taking into account past actions in later decisions.

2) Cognitive agents

Cognitive agents also called Belief Desire Intention (BDI) agents[9], have a global representation of the environment. The architecture of those agents contains three classes; beliefs, desires and intentions:

-Beliefs: describe the internal representation of the agent's state and properties, the agent's beliefs and habits and the agent's environment and the neighbors' characteristics.

-Desires: define the goals of agent. First of all the agent has to select a goal according to the environmental interactions.

-Intentions: define the goal selected to reach first

The specificity of these agents is their architecture which allow reasoning. Their ability to generate a plan according to the particularity of their environment, even in unforeseen situations catch up with specially human ability.

The main problem with a purely cognitive agent is the reasoning process, because, for some simple well known situations, reasoning may not be required at all. So controlling the agent's reasoning process can save time in simulations.

B. The IBM model for pedestrian motion in normal and panic situations

The agent in our approach is a hybrid agent which use "reactive" and "cognitive" approaches to get the best properties. In fact, a graduation exists between purely reactive agents and purely cognitive agents which represents the extremity of a line and not a categoric opposition [10].

Our hybrid agent adapts its reasoning process to specific situations and reacts in a deterministic way to different situations. It is able to act as a reactive agent and gets a quick response for external stimuli those which do not need a plan, and acts as a cognitive agent to generate or use plans in some complicated situations.

The architecture of the agent comprises three classes or containers: Beliefs, Desires, Intentions. We introduce three essential functions: Perception, Deliberation and Action upon which the agent behavior is based:

-Perception: the agent's perception updates the agent's knowledge base "Belief" according to the interactions with the environmental entities. These interactions require a

“communication” between the entities which are normally assured by sensory capacities (i.e vision capacity, hearing capacity, oral capacity).

–Deliberation: deliberation is a cognitive capacity of the agent, it assures the reasoning process to find the desire to firstly achieve. To deliberate, the agent takes into account his beliefs, his action base (i.e the base is an agent base rules which contain a set of actions that the agent can carry out, in response to stimulations) and his perceptions.

–Action: after deliberation the agent's goal is to realize the desire chosen, a set of actions (i.e plan) is then established, the agent is able to update its plan according to its perceptions.

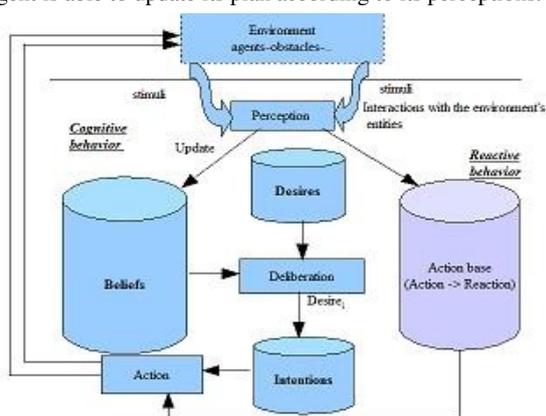


Fig. 1: Hybrid agent's architecture

Space is defined as a grid of cells where every cell can contain one agent (e.g an individual, a wall, an obstacle, etc.) per time step. Pedestrians move into this discrete space from one cell to another according to their velocity characteristics. They have to refer to their perceptions to recognize other pedestrians and objects in their vision field, and have to communicate with their environment to exchange information and so coordinate their actions.

Perceptions are controlled by the sensory capacities of each agent, namely the vision capacity and the hearing capacity. In our approach vision is defined by two parameters the *vision angle* and the *vision radius* controlling the visual perception of the agent. The agent's hearing capacity is defined by the parameter *perturbation*, the agent is continually listening to its environment, thus the hearing capacity is always active and is ready to detect any perturbation in sounds such as an alarm.

Communication is a basic agents' interaction process, it provides the interchange of decisions, thoughts, or information by speech, signs, etc. Three major parts in human face to face communication are distinguished in literature; body language, voice tonality and words [11]. In order to simplify the modelisation of the agents' communication, we firstly consider that the agent can communicate only with its Moore neighborhoods (8 neighbors) [12], and then we propose a simple protocol of communication based on the sending and the receiving of messages. When communication is needed between two agents, each agent asks its neighbor for

information. The reception of these information allows the update of the agent's knowledge base, so that it can decide the appropriate actions to perform.

The agent's perceptions are the base of the interaction process. The exchange of information between agents, and the reception of the vision or the sonorous information, influence the decisions taken and therefore the evolution of the system modeled.

The agents' motion is based on realistic pedestrian motion rules. When we look at Helbing's observations [1], we can note that pedestrian motion is controlled by certain rules which influence its orientation choice. Helbing's observations demonstrated that:

–Pedestrians keep a certain distance from other pedestrians and borders (walls, obstacles, etc.).

–Pedestrians prefer to walk at their own speed which depends on their characteristics and desires.

–Pedestrians choose the fastest route to their next destination rather than the shortest one.

To model such behavior we introduce a few known properties in our agent base rules like repulsion and attraction properties, and collision avoidance strategies. Agents in our approach are attracted by agents who are walking on the same side as them and avoid agents walking in the opposite direction [13]. These repulsive and attractive behaviors are controlled by the agents' interactions and allow the description of a lot of quite realistic observed phenomena such as *lane formation* which refers to a self-organization of pedestrians into separate flows according to their walking direction, and *oscillations at bottlenecks* which refer to oscillations of the passing direction at bottlenecks [14].

However, despite the collision avoidance strategies' the agent sometimes may be in a conflictual situation when its desired next position is occupied by another agent, and without a reliable strategy, agents can remain still for a long time which may cause bottlenecks. To solve this situation we propose a conflict resolution strategy inspired by the “Game theory”, which is a reliable theory to study and model social science problems and situations where the individuals' decisions affect each other [15].

In a blocked situation where two pedestrians aim to reach the same position, a negotiation must be done between the individuals to decide who does what. In the “Game theory” this situation is represented in an hybrid game which contains cooperative and non cooperative player behavior, negotiation and competition strategies are then defined to guide individuals' decision and to control the evolution of the system. The conflict resolution strategy that we propose in our approach is based on this hybrid strategy where the agent is free to decide whether to cooperate or not.

In a blocked situation the agent's action depends on its strategy (i.e a cooperative strategy or a competitive strategy). In fact, in that situation deliberation function provides the action to perform according to agent's conflict resolution strategy and agent's current state. After deliberation each agent informs the other about its decision and moves to target cell. To accelerate simulations, agents move to the target cell providing that they

do not create another blocked situation, otherwise agents have to negotiate again. In fact, blocked situations will be resolved in all cases, the only variable which changes is time. The cooperative behavior may reduce conflict time between agents, and so save time for pedestrians to achieve their destinations.

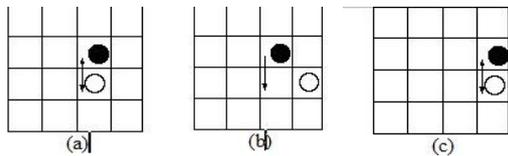


Fig. 2: Solving conflict situations. These figures present a solving conflict situations of the initial situation (a). The figure (b) illustrate a cooperative behavior of the agent in white. The figure (c) present a conflict situation due a conflict behavior of the two agents , agents still blocked again.

In panic situations people are overwhelmed by a feeling of fear and anxiety. The sudden disruption which accrued affect their behaviors and makes them loose their reasoning capacities and behave “irrationally”[16]. In our model, agent’s perceptions detect disruptions which affect its characteristics and cause variations in its behaviors. Panicking agent boost then its velocity in order to escape as fast as possible the danger. Agent’s interactions increase then, an imitation behavior appears which affects the evacuation process and may produces the emergence of collective behavior like “boids”[13].

Hybrid agent increases its walking speed when it starts to panic. Because of the lack of information about the panic propagation process, we concentrate on interactions as a unique means for panic propagation. When a non panicking agent interacts with panicking neighbors, panic ensues if the number of those panicking reaches a fixed limit. The new panicking agent then increases its speed which corresponds to the maximum speed that it can reach, and may follows agents with same speed. Boids will be formed then and the emergent behavior will appear as a response to the environmental variations.

III. SIMULATION RESULTS

We describe in this section the results of simulations of two situations, namely normal situations and panic situations. The simulations provide critical informations to study the variation of pedestrians’ behavior according to environment stimuli and the pedestrians’ conflict resolution strategy.

A. The influence of the conflict resolution strategy on evacuations

We investigate in this part the behavior of pedestrians leaving the environment when strategies to resolve potential conflict situations vary. The simulations are performed in a 41*41 NetLogo[17] lattice, with 112 heterogeneous agents located in a random cell (i.e 56 agents in the up part of the environment and 56 agents in the down part, the distribution is uniform). In fact, in the simulations agents do not have the same Beliefs (i.e speed, conflict resolution strategy, faith neigh

boors, and alarm detection), because these beliefs depend on individual’s culture, age, sex, experience, etc. We take into account these human characteristics in the model to make it as much realistic as possible.

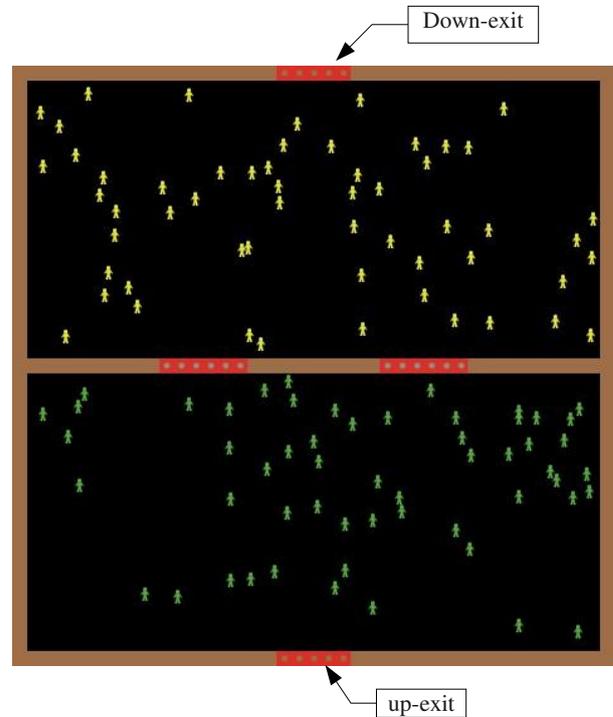


Fig. 3: The figure shows the screen shots of the simulation with a population of 112 heterogeneous agents. Agents in yellow represent pedestrian on the up part of the environment who try to reach the up-exit. Agents in green represent pedestrians on the down part of the environment who try to reach the down-exit.

The exits *up-exit*, *down-exit*, and *middle-exit* are represented by a still agents that occupy red patches. To exit environment agents on the up-part (i.e yellow agents) have to reach first the *middle-exit* and then the *up-exit* to exit the environment. At the same time agents on the down-part (i.e green peoples) have to reach first the *middle-exit* but the *down-exit* to exit the environment. When the agents reach their exit the *down-exit* or the *up-exit*, they exit the simulations.

The middle exits in the simulations act like doors walkable in both direction, they are performed in order to reduce space for agents and to provoke bottlenecks with agents moving on opposite direction. This is environment is artificial but it allows the study of conflict phenomena and the observation of the emergent one.

Figure 4 illustrates the variation of the pedestrians’ flows in normal situations. We see that in cooperative strategy pedestrians evacuate quickly than in conflict strategy, this is because in a conflict strategy the face-to-face conflict can make last blocked situations. The figure 5 illustrates evacuation flows in panic situations with varying conflict resolution strategy, the

result is similar than the one in normal situations. With the same initial conditions, in cooperative strategy pedestrians leave the environment in less than 45 time step whereas evacuation time pass the 60 time steps with a conflict resolution strategy.

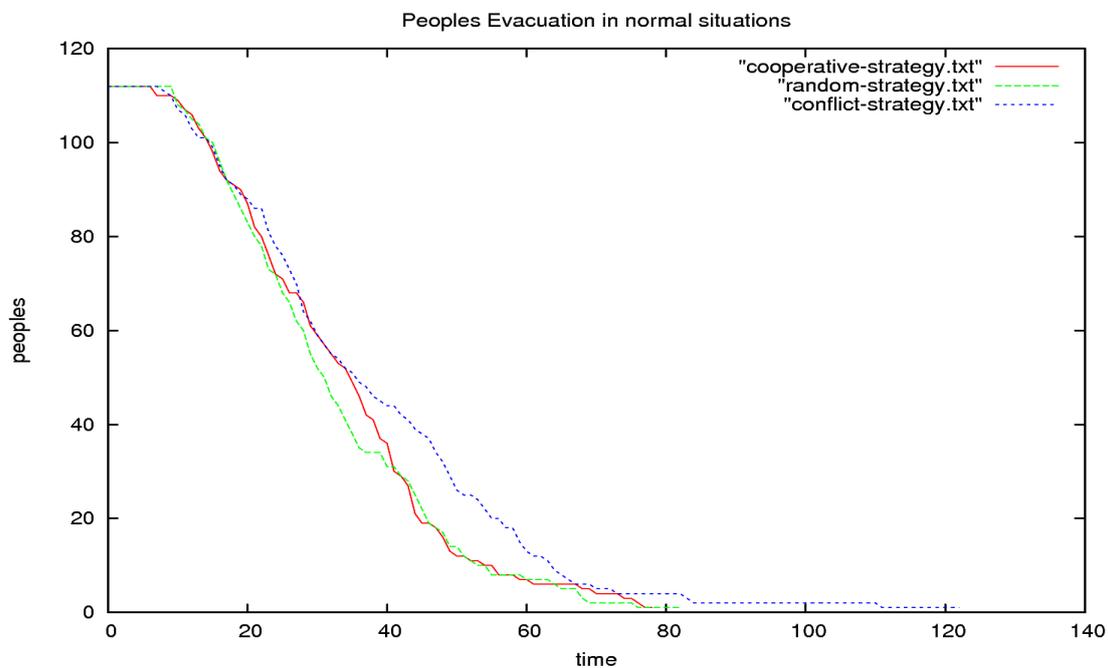


Fig. 4: Variation of the evacuation flows in normal situations

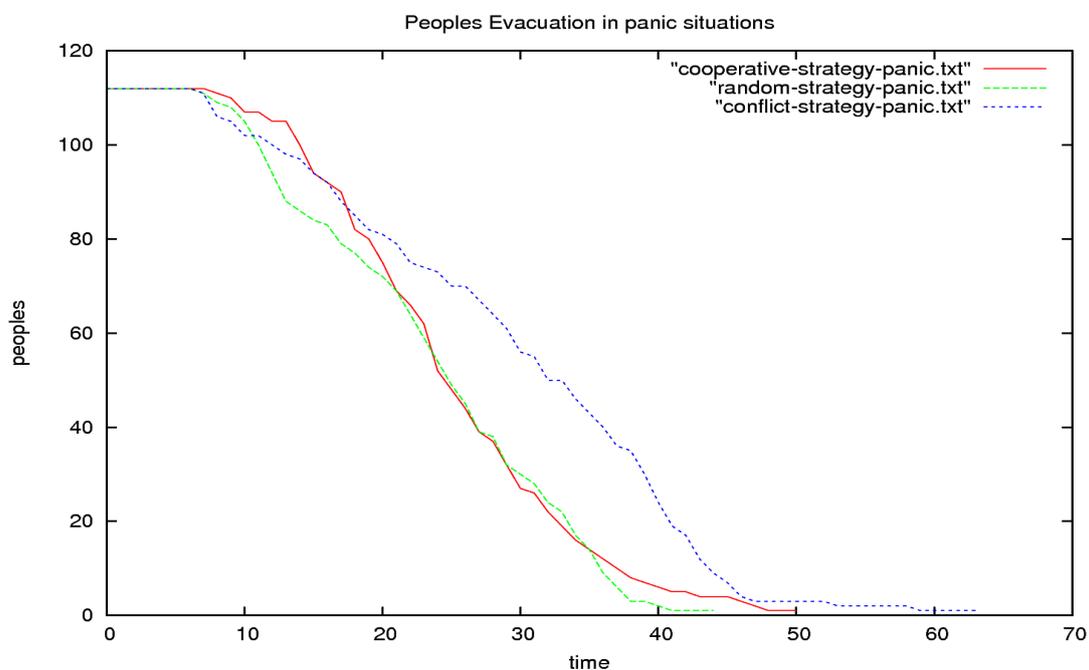


Fig. 5: Variation of the evacuation flows in panic situations

From these two figures we note that conflict resolution strategies influence the pedestrians evacuations flows. The cooperation is the best behavior that individuals have to take when they are in face-to-face situations to avoid bottleneck situations. This result is confirmed in figure 6 which illustrates the variation of bottlenecks in normal situation with the two conflict resolution strategy, namely cooperative and conflict resolution strategy.

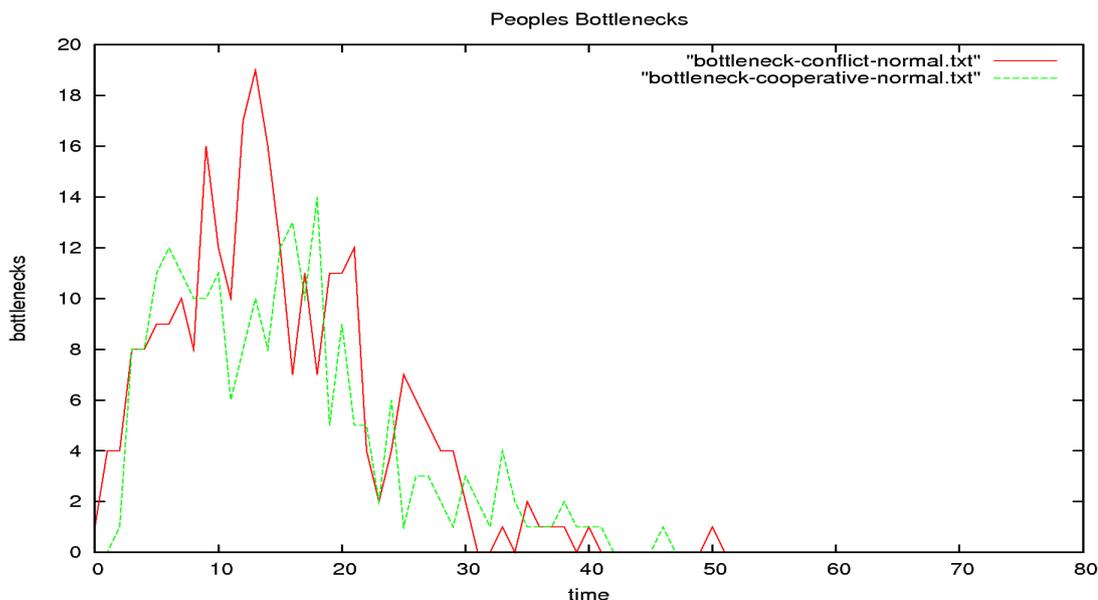


Fig. 6: Bottlenecks variation according to the conflict resolution strategy

B. The influence of panic behavior on evacuations

We investigate in this part the effect of panic on evacuations process. The figure 7 presents the evolution of evacuation on normal and panic situations with pedestrians' cooperative strategy. In fact, for 150 time steps about 121 pedestrians are evacuated in normal situations whereas only 71 pedestrians are evacuated in panic situations. We note that evacuation time in panic situations is greater than the one in normal situation, thus trying to move fast can slow down evacuation and then may cause human disasters by reducing chances to survival in catastrophic situations.

Trying to move fast increase pedestrians bottlenecks (see figure 8) which has a bad effect on the pedestrians' evacuation flows. This phenomena has been observed in Helbing's social force model [1] "*Faster-is-slower-effect*" due to impatience where he demonstrated that trying to move fast can cause a smaller average speed of leaving.

Getting nervous and panic make pedestrians behave "irrationally" and make them take decisions that worsen the catastrophic situations. Alternative solution can be developed for panic solution to avoid such behavior and to improve the evacuations flow.

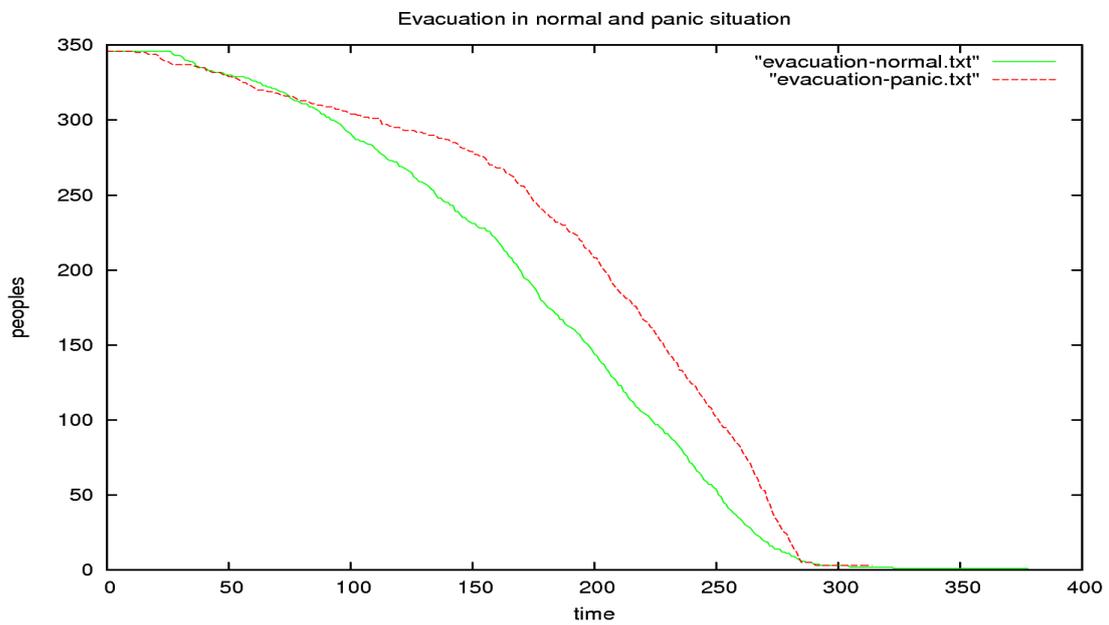


Fig. 7: Influence of the conflict resolution strategy in peoples' evacuations. 346 heterogeneous pedestrians (173 pedestrians in up part, 173 pedestrians in down part). The Beliefs' parameter "*detect-alarm? = true*" and "*speed = 1*" for all pedestrians.

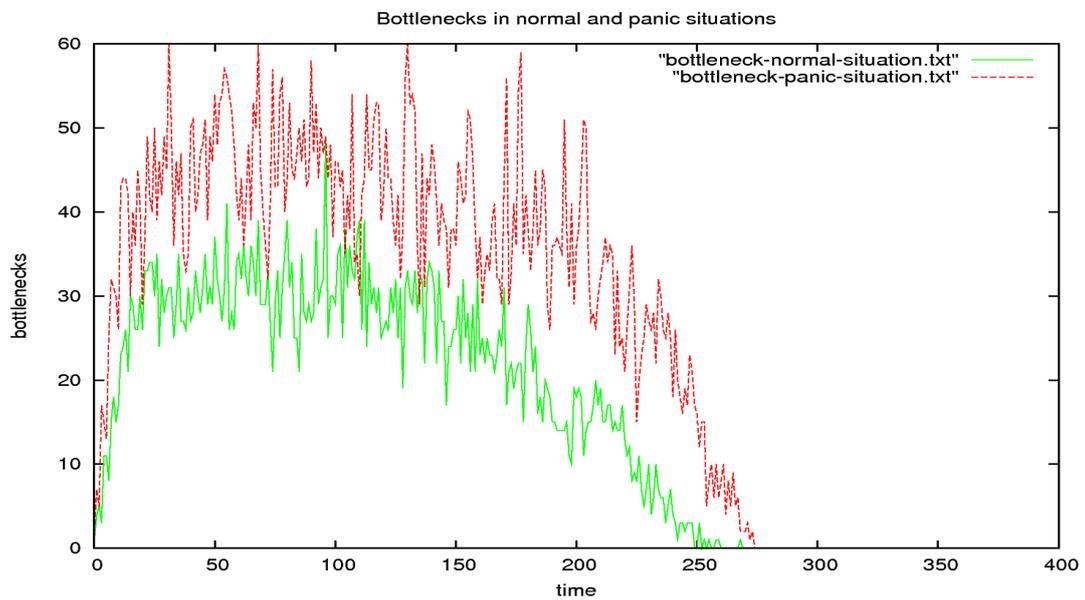


Fig. 8: Bottlenecks' variations on panic and normal situations. The beliefs' parameter are the same a Fig. 7.

IV. CONCLUSION AND PERSPECTIVES

We have presented a model of crowd motion in normal and panic situations. Pedestrians' characteristics and behaviors are tracked through time to investigate the properties that emerge from entities' interactions. Results obtained by this IBM model shows that cooperative strategy reduce blocked situation between pedestrians while conflict strategy makes them stand still for a long time. In panic situations pedestrians try to move fast to escape danger. The increase of the velocities increase on the same time bottlenecks which can produce deadlocks.

Since human life is connected with delays the best strategies have to be taken and the best behaviors have to be adopted by pedestrians in order to reduce human loses.

These are a preliminary results and we aim to improve our model by detailing and including more pedestrians' characteristics, and by introducing more detailed conflict situations. Besides, we will concentrate on emergent phenomena that have been observed like *lane formation*, or *freezing by heating*.

ACKNOWLEDGMENTS

This work is supported by *La region Haute-Normandie*.

REFERENCES

- [1] D. Helbing, I. J. Farkas, P. Molnár, and T. Vicsek (2002), "Simulation of Pedestrian Crowds in Normal and Evacuation Situations" . Pages 21-58 in: M. Schreckenberg and S. D. Sharma (Eds.) "Pedestrian and Evacuation Dynamics", Springer, Berlin.
- [2] A. Kessel, H. Klüpfel, J. Wahle, and M. Schreckenberg (2002), "Microscopic simulation of pedestrian crowd motion". Pages 193-200 in: M. Schreckenberg and S. Sharma (Eds.) "Pedestrian and Evacuation Dynamics", Springer.
- [3] D. Provitolo (2007), "A proposition for a classification of the catastrophe systems based on complexity criteria", European Conference on Complex System, Dresden-Germany.
- [4] L. F. Henderson (1974), "On the fluid mechanics of human crowd motion". Pages 509-515 Transportation Research.
- [5] V. J. Blue, J. L. Adler (2001), "Cellular automata microsimulation for modeling bi-directional pedestrian walkways", in Transportation Research Part B: Methodological. Pages 293-312.
- [6] C. Burstedde, K. Klauck, A. Schadschneider, J. Zittartz (2001), "Simulation of pedestrian dynamics using a two-dimensional cellular automaton", in Physica A Statistical Mechanics and its Applications. Pages 507-525.
- [7] V. Grimm, S. F. Railsback (2005), "Individual-based Modeling and Ecology". Princeton University press.
- [8] <http://www.red3d.com/cwr/ibm.html>
- [9] Greg M. P. O'Hare, Nick Jennings (1996), "Foundations of distributed artificial intelligence" Chap. 5. Published by Willey-IEEE.
- [10] J. Ferber (1995), "Les Systèmes Multi Agents: vers une intelligence collective". InterEditions.
- [11] Mehrabian and Ferris (1967), "Inference of Attitude from Nonverbal Communication in Two Channels". In: The Journal of Counseling Psychology Vol.31, 1967., Pages 248-52.
- [12] G. Weishuch (1982), "Dynamique des systèmes complexes: Une introduction aux réseaux d'automates". Savoirs Actuels, InterEditions/Editions du CNRS
- [13] Graig W. Reynolds (1987), "Flocks, herds and schools: A distributed behavioral model", in Computer Graphics, 21(4), SIGGRAPH '87 Conference Proceeding). Pages 25-34
- [14] D. Helbing, P. Molnár, I. Farkas, and K. Bolay (2001), "Self-organization pedestrian movement. Pages -383 in: Environment and Planning B: Planning and design.
- [15] E. Rasmusen (2001), "Games & Information 3rd Edition: An Introduction to game theory". Blackwell Publishers.
- [16] N. R. Johnson, "Panic at " The Who Concert Stampede": An empirical assessment (1987), Social Problems 34(4), 362-373 .
- [17] <http://ccl.northwestern.edu/netlogo/>

Cholera in the 19th century: Constructing epidemiological risk with complexity methodologies

Éric Daudé, Emmanuel Eliot, Emmanuel Bonnet

Abstract— Risk epidemic and complexity are linked by space and interactions. First, space matters in risk situations because of its ability to hold concurrently and simultaneously favorable conditions for a future emergence or re-emergence of epidemics. Second, space matters as a mediator of interactions, social as environmental, and at different levels. Risk is dynamic and its spatio-temporal dimension increases difficulties to catch it. Empirical data lack off precision to follow epidemiological outbreak. Complex system theory and connected methodologies can help us to enlighten this empirical failure.

First, we present some knowledge about risk, health and complexity. Second we present social and spatial data based on the first epidemic of cholera in the city of Rouen, in 1832. Third we propose two models to explore the diffusion of this epidemic.

Index Terms— Risk, epidemic, dynamical system, cellular automata, modeling, simulation.

I. INTRODUCTION

THE spatial analysis of risk may be defined as the investigation of probability of being affected by a hazard in space and in time. This type of analysis requires a deep focus on the multilayered and complex combinations of indicators that are located in space. The analysis is ever difficult for certain risks but it reaches a peak when it concerns epidemiological ones. Because mobility and the ways people move in space and time is a major factor in the dynamic, especially in the case of an epidemic, the investigation of epidemiological risk faces three major problems:

- First, the need of understanding the ways people move and interact with space. Mobility is socially constructed and its patterns vary in history and according to social and cultural characteristics.
- Second, the question of the emergence of risk and

its location. Because the intensity of the epidemic depends upon the level of interactions between people, we need to consider the accessibility of spaces in regards with others.

- Third, the evaluation of the temporality of the epidemic.

We propose to explore these problems on the basis of the analysis of the second cholera pandemic that affected most parts of France at the beginning of the 19th century. We examine it on the basis of an ecological approach of the epidemic but also by adding complexity theories analysis. The study will take advantage of these frameworks in order to avoid and overcome data bias.

In the analysis, space and risk are strongly related. Risk appears in space and can be created by space: not only considered as a support but also as an ‘incubator’ of risk situations. Risk is thus space related: presence and density of the *vibrio cholera* are dependent upon both aquatic reservoir and upon the more or less high concentration of population in the environment. In addition, risk occurs at different scale (world, nations and cities) and involves many actors: disease control, doctors and inhabitants. Risks are dynamics: going from the emergence of the virus to the pandemic may reflect this.

Self-organization theory [1] is adapted to explain emergence of risk for which local disruptions may product global and unpredicted events [2]. The self-organization theory identifies processes which allow describing behaviour at a global level, persisting in time and space, from numerous interacting entities located at one or several lower levels. Most of these interactions are local one and such systems are characterized by an absence of planning: no global control which would pilot such structure, such behaviour, or such form. Activity of

The authors would like to thank the IRSHS and the region Haute-Normandie for the funding of the program from which this work stems from.

Éric Daudé is with Rouen University (email: Eric.Daude@univ-rouen.fr), Emmanuel Eliot is with Le Havre University (email: Emmanuel.eliot@univ-lehavre.fr) and Emmanuel Bonnet is with Caen University (email: Emmanuel.bonnet@univ-caen.fr). Authors are with UMR-CNRS-IDEES.

such system, dynamic and open to their environment, is in evolution. Evolution between attractors can be cyclic. Such systems are characterized by phases of intense activities: evolution of societies is tagged by wave of huge pandemics. Otherwise, system can evolve towards a stationary state, converged at an attraction point and absorbing progressively its activity. Activity of a system, in an epidemic perspective, can lead it through different states through the time. This switch from a state to another is situated close to a bifurcation point that may lead towards chaos. In an earlier work, we explored the different phases of activity of the logistic function often linked with diffusion processes [2]. When the system evolves from a bifurcation threshold, the transition from one state to another qualitatively similar refers us to the concept of resilience. The stability of self-organized systems refers to the possibility of change which explains that all living systems go through distinct phases during their activities. These phases are theorised by the criticality [1], which shows that all self-organized systems evolve towards a critical state and that a small and local disruption is enough to produce huge alterations. This event is characterized by a system which goes into a phase of mutual and global interaction during which level of connections and interdependences is maximal: this is the case of pandemics.

If they are useful in a heuristic context, such concepts are however difficult to use when one wants to apply them or to spot them in an empirical way. For example, how evaluating the intensity of relationships between elements at the same level and between elements at different levels? These uncertainties lead us to propose simple models of diffusion based at the same time on empirical evidence and theoretical knowledge.

II. THE GLOBAL AND NATIONAL CONTEXTS OF THE EPIDEMIC AT THE BEGINNING OF THE 19TH CENTURY

After decades, routes of the epidemic have been recomposed on the basis of archives, reports and medical information. The second epidemic seems to have started in the British colony of Bengal in 1826. In 1837 the west coast of Mexico, the Anglo Egyptian protectorate and the French colonies of the

North West of Africa seemed to have reported the last cases in the known world at the beginning of the 19th century (figure 1).

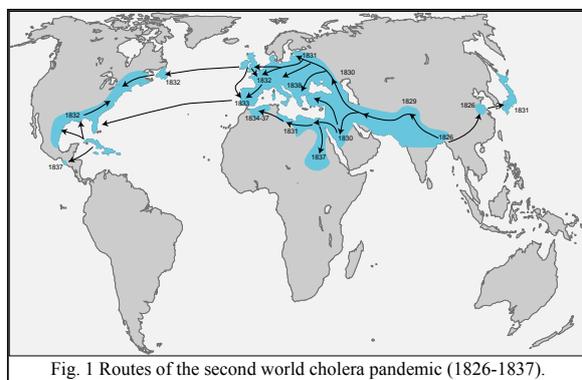


Fig. 1 Routes of the second world cholera pandemic (1826-1837).

Although the etiologic of *vibrio cholerae* was unknown yet, details about the transmission disease were reported by doctors of the British Raj since the beginning of the 19th century. As many other unknown diseases, cholera produced social reactions. In the French context of the 1830's, the epidemic broke out in a period of political troubles, which contributed to reinforce both the political conflicts and the social representations. Officially, the epidemic broke out in Paris in April 1832 and spread until the month of November. However, high rates of mortality due to diarrhoea were already reported in the northern parts of France by the end of 1831 [4].

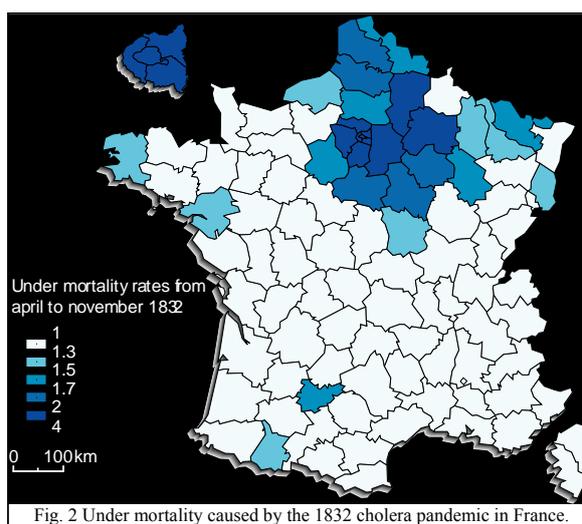


Fig. 2 Under mortality caused by the 1832 cholera pandemic in France.

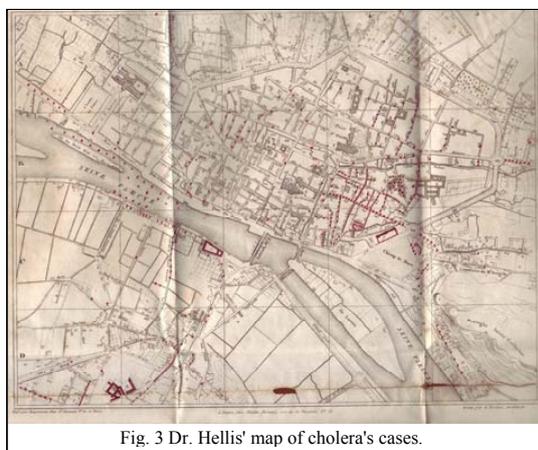
The 'département' of the 'Seine-Inférieure' was not the most severely affected by the epidemic

according to the official sources, but located between Paris and the English coastline - two major epicentres of the epidemic - this region is interesting for analysing the course of the epidemic at several levels (figure 2).

III. THE DATA

A. *The Cholera data*

All the cholera data were collected in the archive services of the region between 2006 and 2007. The collection covers the period 1832-1893, i.e. from the second to the fifth pandemic. The sources of information are mixed: medical reports, medical topographies, municipal and administrative reports. The present paper focuses on the second pandemic (« the cholera morbus epidemic ») and in the main regional centre of the department: Rouen. The epidemic reached this city in April 1832 and left it in October 1832. In Rouen, the analysis of the epidemic is based on two complementary materials: a medical topography done by the chief doctor of the hospital, Dr. Hellis (figure 3) and the report of the municipal officers. Obviously, the collected data suffer from bias.



First, in 1832, the etiology of the disease was unknown yet. The second pandemic was the first that reached France in the beginning of the 19th century and the causes and the ways of treatments were totally unknown, although information circulated all over the colonial empires. The main debate was based on the explanation of the causes of the disease: either contagious or not. This debate fed all the policies and fuelled all the theoretical

conflicts until the end of the 19th century. The discovery of the cholera organism by F. Pacini in 1854 and after all by R. Koch in 1883 however improved the knowledge about the disease.

Second, the data were produced by a health care system that was centralized at the national and departmental levels. It produces statistics and topographies that gave an overview of the epidemic but it does not have the possibility to evaluate the under-reporting. Moreover, sources of report were very mixed: doctors, sanitary and municipal officers. In addition, the topographies (map) done on cholera aimed at proving the interpretation of the disease. In other words, all the cartographic methods were used in order to show that the epidemic comes from outside and was imported by seamen. However, comparisons between local data based on hospital and municipal reports with the Hellis' map seem to converge. As a conclusion, the map and the associated data are in fact the only available source able to trace the first cholera epidemic in Rouen.

B. *From data to visualization and interpretation*

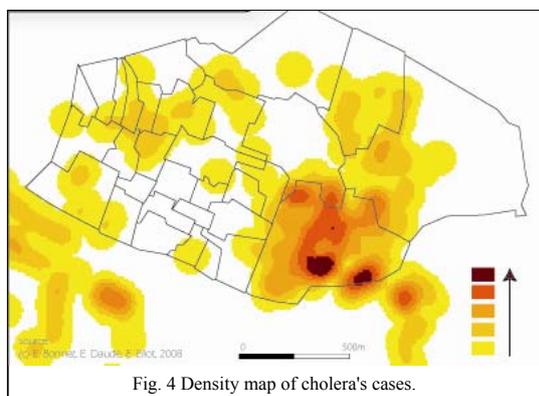
We used different methodologies to integrate historical data from the archives services. The most important difficulties using archives data is the lack of statistical and geographical information. The first step was to consider if this collect was representative about the disease. The second step aimed at validating the data by checking their localization and their translation between manuscript and database. The third step aimed at calibrating the data with geographical information. Thus, we needed to reconstruct the geography of the 19th century by recreating the different administrative levels, by modifying the city's names and by identifying some city groupings.

The interpretation of the epidemic necessitates a better understanding of the topography and of the social geography of Rouen. Different surveyor's maps were available about the city. This step of modelling allowed us to construct environmental factors which were necessary to understand the context of the epidemic. Therefore, the hydrology, the topography, the open spaces, the fountains, and some public places (market) were integrated in a Geographical Information System (GIS). Based on

the available parishes, charity expenditures, rental values of housing, population density were integrated and compiled into indicators in the GIS. These information were based on academic and archives information that were compiled by historians [5].

For the geocoding, we have used the current street names for the first treatment, and for the unavailable data, we have compared current and past street names. After the first and the second treatment, 90% of the archive data were located.

The analysis of each cholera case is difficult because the information on the map were insufficient. So, by using some spatial analysis treatment we have produced a density map of the cholera cases in the city (figure 4).



There were many interpolation methods that provided this type of representation. We chose a Smooth Surface Reconstruction because this method uses natural neighbor interpolation, works in any dimension and allows dealing with non uniform samples. All these treatments allow identifying the most affected places of the epidemic.

A cluster of cases is reported in the South-eastern parts of the city. By using animations based on the weekly available data, we also identify the diffusion of cholera in the western and north western parts of Rouen. An analysis of the mapping shows a relation between the cholera cases and socio-economic indicators, the aquatic environment and the density of population. Based on this interpretation of the epidemic, we aim at describing and explaining the dynamic of this epidemic at an infra-urban level. In the next section of this paper, we describe two models that capture these social and spatial aspects.

IV. MATHEMATICAL AND COMPUTATIONAL EXPLORATION OF THE CHOLERA DIFFUSION

First, we define a model based on ordinary differential equations. This macro model aims at understanding the general mechanisms of the spread of cholera. A second model is then presented, based on cellular automata. It takes into account both spatial and social heterogeneity as well as local interactions.

A. A basic model of contagion

The basic classical SIR model of Kermack-McKendrick [6] has to be modified to take into account the indirect process of contagion [7], i.e. the ingestion of contaminated water. In this case propagation is contingent on the existence of a mediator which is the *vibrio cholera*. In the model, population is divided in four groups: Susceptible, Infected, Removal and Death. During an epidemic, a fraction of population is contaminated by the virus, mostly by ingestion of infected water. Once infected, people becomes actors of the propagation because they produce and reject *vibrio cholera* in the environment. After few days, evolution of the infection may lead to death or recovering, depending mostly on the health state of the individual and of the care conditions. We capture all this elements in the following model.

$$S(t + dt) = s(t) - r.S(t).dt \quad \text{avec } r = \beta.f(C) \quad (1)$$

$$I(t + dt) = I(t) + (r.S(t) - \gamma.I(t)).dt \quad (2)$$

$$R(t + dt) = R(t) + (\gamma.\alpha.I(t)).dt \quad (3)$$

$$D(t + dt) = D(t) + (\gamma(1 - \alpha).I(t)).dt \quad (4)$$

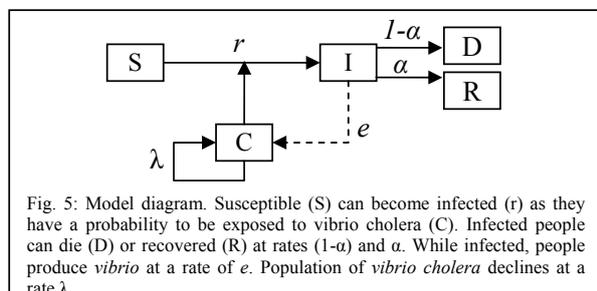
$$C(t + dt) = C(t) + (e.I(t) + \lambda.C(t)).dt \quad (5)$$

In equation (1), parameter r represents the proportion of susceptible which is infected by the *vibrio*. Parameter β is the probability to interact with an infected aquatic environment and $f(C)$ is the probability to be contaminated by the vibrio, which is a function of the quantity of *vibrio cholera* (C) in water.

In equation (2), γ represents the proportion of infected case which get out from infection. Part of them will remove from infection (α in equation (3)) and other will die ($(1 - \alpha)$ in equation (4)).

In equation (5), λ represents the loss rate of *vibrio cholera* in the aquatic environment and e the growth

rate of C due to the excretions of each infected case. The flowchart of this model is presented in figure 5.



We aim now at modifying this macro model in order to take into account the spatial dimension and the heterogeneity of the population. Implicit hypothesis of this model is a perfect mixing of the population with environment, but aquatic environment is not present everywhere in the city. The second implicit hypothesis is the homogeneity of the distribution and of the type of inhabitant, which is not the case.

B. A cellular automata model of contagion

A conceptual framework has been developed to capture the structures and dynamics which occur in the propagation processes [1, 8]. This framework is applied in the context of a cellular automaton. Structure of this model is defined by three elements:

- *Elementary entities*: a cell represents a square of one hundred meters. A cell can be an environmental cell {river, green space, public building...} or a 'social' one. In this last case, it has variable states which represent social attributes, such as number of inhabitants and the level of income. The domain is then the cellular grid (47x29) which shapes the city of Rouen.

- *Propagation channel*: it is the local spatial interaction structure of the cellular automata, i.e. the Moore neighborhood. Each cell can then interact with its 8 neighboring cells.

- *Virus*: It is the driver of the diffusion. This particle is generated by the fraction of infected people and is transported both by the environment and the infected cases. Each cell has a variable state which stocks an amount of the vibrio.

The diffusion dynamic is related to this structure and is composed of three processes:

- *Emission*: it represents the propagation of the

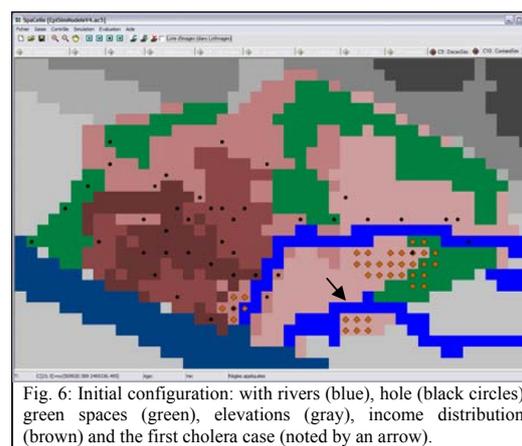
vibrio, from the people to the environment. In the following simulation, each aquatic cell computes the stock of infected people (I) in its surrounding and receives a fraction e of vibrio related to this stock.

- *Circulation*: this process defines the 'life' of the virus in the environment. The circulation of the stock of vibrio cholera is mainly aquatic dependent. Three processes define this life: a water cell receives a fraction of virus from upper cells (1) and gives a fraction of virus to lower cells (2) - based on the elevation ground - and the stock declines at a λ rhythm (3);

- *Infection*: the rule of infection is relevant to this process. People interact with their environment and can be in contact with infected water, and be contaminated. In the model, each inhabited cell compute the volume of vibrio present in its vicinity - order 4 - and the stock of susceptible population has a risk r to be infected by this amount of virus.

Finally, once infected, the number of people which die or recover depend upon two parameters, d and $(1-d)$. These parameters, which are proportions, are the same for all cells. This latter process is linked with the circulation process in the sense that it represents an indirect rule life for the vibrio cholera.

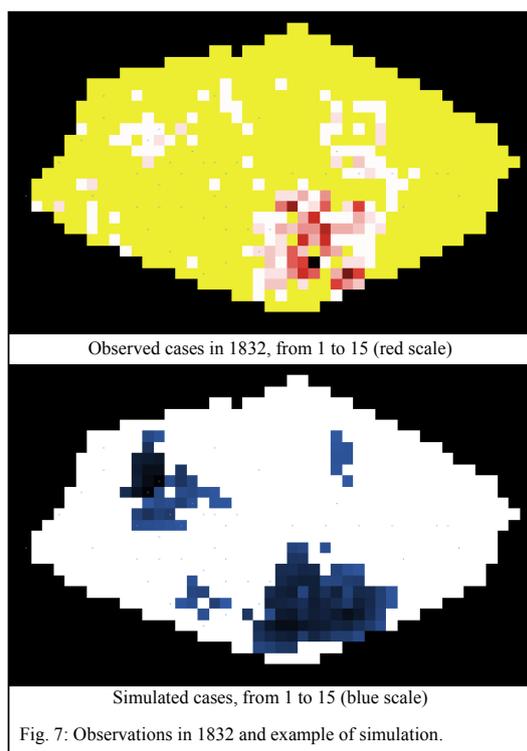
Figure 6 presents the initial configuration of the model and the localization of the first cholera case in the south-eastern parts of the city.



In the next section we present some simulations and first results.

V. SIMULATION OF THE CHOLERA DIFFUSION

In this cellular automata model, the two main factors which are responsible of the propagation are the presence of aquatic environment (river and wells) and the density and the level of income of the population, which is measured by the charity expenditures. Firstly, the proximity to a wet environment increases the probability to get the vibrio. There are two rivers in Rouen, the *Robec* and the *Aubette*, and a high number of cases are reported along and between these two rivers. The main reason of this spatial correlation is the location of mills, spinning and paper mills where many workers were concentrated. Secondly, and correlated to the first, both the high density of population in these areas and the low level of income of these populations increase the probability to be contaminated. Health risk based on the analysis of social and spatial determinants is thus very high in this model, as probably in this past reality.



These first results presented in figure 7 hold attention because there are good qualitative and quantitative correlations between the simulations

and the observations.

Thus, the model has to be tested in different ways: What is the sensibility of the results to the parameters? Are the results significantly different if the first case is located in other sites in the city? Have the parameters significantly closed values to the observed ones?

REFERENCES

- [1] P. Bak, *Quand la nature s'organise*, Flammarion, Paris, 1999.
- [2] E. Daudé, D. Provitolo, E. Dubos-Paillard, E. Gaillard, E. Eliot, P. Langlois, E. Propeck-Zimmermann and T. Saint-Gérard, "Spatial risks and complex systems: methodological perspectives" in *From System Complexity to Emergent Properties*, Springer series understanding complex systems, in press.
- [3] P. Langlois, E. Daudé, "Concepts et modélisations de la diffusion géographique", *Cybergéo: revue Européenne de géographie*, <http://193.55.107.45/articles/364res.htm>, n°364, 23 p.
- [4] P. Bourdelais P., "La marche du choléra en France, 1832 et 1854", *Annales E.S.C.*, n° 1, pp. 125-142, 1978.
- [5] J.-P. Bardet, *Rouen aux XVII^e et XVIII^e siècles. Les mutations d'un espace social*, SEDES, Paris, 197 p., 1983.
- [6] W. Kermack and A. McKendrick, "A contribution to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society of London*, 115, pp. 700–721, 1927.
- [7] V. Capasso, S.L. Paveri-Fontana, "A mathematical model for the 1973 cholera epidemic in the european mediterranean region," *Revue Epidémie et Santé Publique*, n°27, pp. 121-132, 1979.
- [8] E. Daudé, P. Langlois, "les formes de la diffusion", actes du colloque GéoPoint'04, pp. 171-175, Avignon, 2006.

A multiagent urban traffic simulation Part II: dealing with the extraordinary

Éric Daudé, Pierrick Tranouez, Patrice Langlois

Abstract— In Probabilistic Risk Management, risk is characterized by two quantities: the magnitude (or severity) of the adverse consequences that can potentially result from the given activity or action, and by the likelihood of occurrence of the given adverse consequences.

But a risk seldom exists in isolation: chain of consequences must be examined, as the outcome of one risk can increase the likelihood of other risks. Systemic theory must complement classic PRM. Indeed these chains are composed of many different elements, all of which may have a critical importance at many different levels.

Furthermore, when urban catastrophes are envisioned, space and time constraints are key determinants of the workings and dynamics of these chains of catastrophes: models must include a correct spatial topology of the studied risk.

Finally, literature insists on the importance small events can have on the risk on a greater scale: urban risks management models belong to self-organized criticality theory. We chose multiagent systems to incorporate this property in our model: the behavior of an agent can transform the dynamics of important groups of them.

Index Terms— Risk management, self-organized criticality, multiagent systems, modeling, simulation.

I. INTRODUCTION

SPACE is an important factor of risks situations, not only as a support, but also as an actor in itself of the situation. Risk is space related. In epidemic contexts such as cholera, presence and density of the *vibrio cholera* are dependent both on aquatic reservoir and on the density of population in the environment. Risk has spatial impacts. In environmental context, *flash floods* caused high damages because of their torrential nature and of their high spatial concentrations. Risk management makes tracks in space, and risk sometimes stands to management. In technological context, urban land

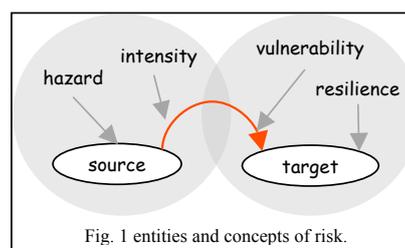
The authors would like to thank the GRR SER and the region Haute-Normandie for the funding of the MOSAIIC program from which this work stems from.

P. Tranouez is with Litis, Rouen University. (e-mail: Pierrick.Tranouez@univ-rouen.fr)

Patrice Langlois and Éric Daudé are with UMR IDEES, Rouen University (email: Patrice.Langlois@univ-rouen.fr and Eric.Daude@univ-rouen.fr)

use and planning reveals some tensions between industrial and residential areas. Risks are multi-layered (world, nations, cities) and imply different kinds of actors, human and non-human. Fight against a possible A *flu* pandemic implies many actors at different levels (World Health Organization, national centers for disease control such as INSERM, local government and doctors) and control measures to reduce risks are both global (air traffic limitation) and local (public services closure). Furthermore, risks are dynamic. In industrial context, one can observe *Domino effect* as an explosion in one site produces secondary accidents in the neighborhood, due to the high concentration of activities.

Risk is defined as a probability of space-time interaction between a source and a target [1]. Four concepts are relevant to this definition and are linked to capture the risk: hazard, intensity, vulnerability and resilience (figure 1).



- Hazard represents the occurrence probability of an alteration into the source that could have effects on target: the probability of emergence or re-emergence of a virus for example;
- Intensity is viewed as an output of the source, it depends on the power and duration of the phenomenon and of the involved surface area: the volume and extension of a toxic cloud for example;
- Vulnerability is an input of the target, it measures at the same time the sensitivity of the target to alterations in its environment, and the related

damages, in term of population or equipments: the probability for a group to panic and to avoid confinement in a technological accident for example;

- Resilience is the capacity for an organization to return gradually an equilibrium state without modifying its final goal: the return time to a normal behavior in a transportation network after a crisis for example.

The main difficulty to characterize risk is the huge amount of interactions that links entities: risk is complex because targets often become sources. And this is all the more true in an urban context characterized by a large number and a great diversity of entities susceptible to be target and source. So space and interactions matter in risk, and they are the two main entrances of our MOSAIC project.

II. DEALING WITH HUMAN BEHAVIORS IN RISK CONTEXT

The MOSAIC project aims to observe and understand local and global effects of individual behaviors in the dynamic of a transportation network system after an industrial accident. Few researches take into account the behaviors of group or individual when studying the risk at the scale of a city. Physical aspects override the measure of risk and population damage is just a result of these major forces. In this way, intensity of toxic cloud or of earthquake defines buffers that are used to estimate the number of inhabitants and equipment concerned by the event, and then give an estimation of the vulnerability. When human behaviors are considered in risk situation, it is mostly at a very fine scale, for example rooms or building [2], and with the same kind of behavior: panic and escaping [3]. At a global scale, deterministic model are mostly used as they are supposed to be more efficient to describe the mean behavior of individuals, particularly if there is a large number of people concerned. We argue that it is possible, and necessary in some sense, to go beyond this approach.

In risk management, many studies have shown that early stages of the phenomenon are critical on the level of the global damage. It is true with epidemic

outbreak when the very few infectious people present in the city can affect, by their individual actions, the course of the epidemic [4]. The same situation can appear when mimetic of panic between some individuals can produce a snow-bowl effect on the entire population. But individual behaviors in risk situations are not limited to panic and escaping behaviors. If one considered Bhopal (1984) or Toulouse (2001) accidents, the number of victims or the resilience of the system have largely increased due to the wondering behavior, curiosity behavior: in some circumstances, people want to see the damage. The aim is then to detect where and in which conditions these bifurcations have a high probability of occurrence in order to prevent them. We then develop a model of simulation in which first, space, as a mediator of interactions, matters. Space is a traffic-oriented network [5]. And second, in which individual behaviors are predominant to explain the dynamic of the vulnerability.

III. MODELS OF BEHAVIORS IN EXTRAORDINARY SITUATIONS

As described in depth in [5], our model and its resulting simulation builds a transportation graph from GIS data, upon which it create mobile agents modeling vehicles.

These agents enter the network at a controlled random place (their insertion is based on scenarios, they are not necessarily uniformly randomized on the whole network), and try to reach a controlled random destination. To each edge of the network is attributed a *weight*, which combines numerous characteristics of this edge, such as its length, speed limits, number of lanes etc. in a quantification of its attractiveness. This lets our agents the possibility to compute an efficient path from where they are to their destination through Dijkstra's algorithm. The planned trajectory of an agent is then a succession of edges. Once in an edge the agent tries to drive to its end, the next connection, where it will be able to choose the next planned edge.

Our mobile agents then drive to their destination, interacting one with another, as their speed, length, driving brashness etc. are considered at each step. Furthermore, these agents can adapt their goals to what they perceive of the traffic, using different

methods to choose other paths to reach their destination.

MOSAIC is concerned by situations where contextual mobility can occur and can diffuse or have large consequences on the global circulation. We call *contextual mobility* a mobility which is associated to short-range goals (to avoid a crowd) and whose result differs from the initial planning (to change a destination). We will now consider an urban industrial accident. This accident has a finite extension area and well-determined intensity, represented by a buffer. Inside this buffer, a proportion of people, related to intensity, want to escape. Outside this buffer, behaviors are less reactive. Some want to escape, others want to see and for others "show must go on", and they want to follow their way. We have then defined different kinds of behaviors and methods related to these different goals:

- *Chicken behavior*: the goal is to find the opposite direction of the source (the buffer), and to drive following this way;
- *Bystander behavior*: the goal is to find the source of danger and to go there. If agent is already in the place, then he stays here;
- *Pragmatic behavior*: here the agent selects a new destination in the network and tries to reach it. This behavior simulates the fact that some people will want to reach their children at school or husband or wife at their working place;
- *Wandering behavior*: there is no goal, this behavior is the sign of distress. At each time step, just select a road and go there.
- *Roadrunner behavior*: this method consists in always selecting the less congested road and to go there. This method can be connected to the Chicken or Bystander behavior;
- *Sheep behavior*: here agent follows the crowd whatever the direction.

We will now present implementations of these behaviors.

IV. SIMULATION OF BEHAVIORS

We will discuss here how the behaviors themselves can be implemented, not why or when one or the other will be chosen.

A. Behaviors classification

In order to implement them, we will distinguish three *categories* of behavior: global, planar and local. These categories are based on the actual behavior, and not on the motivations behind it.

A *global* behavior is one that makes a reasoning about the road network. *Pragmatic* behavior will probably fall in this category: the agent will try to find a good path to his newly decided destination using his knowledge of the network. *Bystander* can also fall here.

A *planar* decision also chooses a destination but tries to reach it using orientation as if no roads existed, as if the vehicle was on an open plan. Of course the network will offer constraints, but a general cardinal like direction will guide the agent. *Chicken* and possibly *Bystander* will fall in this category. This means there are two sub-behaviors in by standing.

A *local* decision is one based on local-only data: *Wandering*, *Roadrunner* and *Sheep* will fall there.

B. Class implementations

Global behaviors are implemented in the agents to allow them to reach their initial destination.

Local require little complexity. *Wandering* is trivial, *Roadrunner* and *Sheep* differ only by the sign of their optimization. We also implemented a simple anti-loop measure: *Roadrunners* for example will choose the less congested road unless they already went recently through this one.

Planar require the ability to choose an edge out of a node based on a global direction. Depending on what the modeler desires, he can choose a distance from the current road intersection, and the agent will choose the intersection at less than the selected distance (expressed in Euclidean distance or number of edges in a path leading to it) that is the closest to the desired direction. An anti-loop measure can be added.

C. Examples of simulation

The behaviors previously described are ways of coping with an extraordinary situation. Most urban important accidents will have their consequences felt locally at the beginning, before it spreads. The evolution of the perturbation will be like waves spreading from the initial locus outward. If the extraordinary behaviors are the waves, the

metaphorical medium of this propagation is the ordinary traffic flow. We therefore need a sophisticated modeling of the day-to-day activities of vehicles in an urban agglomeration. We described this model in [5].

With the simulation of ordinary traffic, one can see in figure n°2 an example of the distribution of vehicles in the main roads of the city.

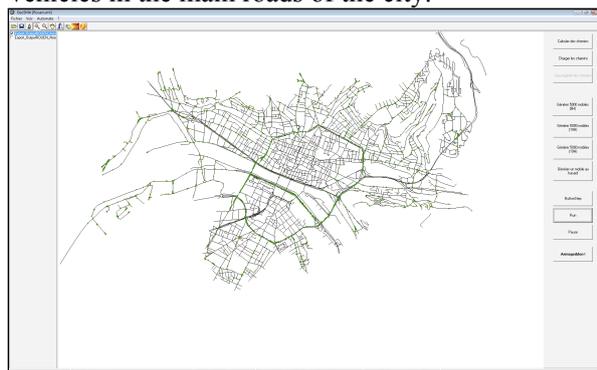


Fig. 2: An example of traffic in a town before an accident occurs.

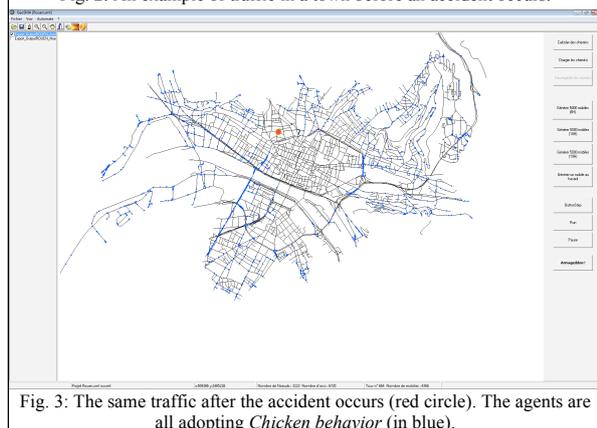


Fig. 3: The same traffic after the accident occurs (red circle). The agents are all adopting *Chicken behavior* (in blue).

Starting at this point, we generate an event in the city that is supposed to represent an accident. This event, for the purpose of the simulation, is perceived by all individuals and is considered as a repulsive event. In figure n°3, this event is a mouse-click event located by the user without any consideration about the reality of the area. As agents perceived the impact zone (in fact XY coordinates), they all change their planned trajectory. Once in a crossroad, all mobiles pick out the *Chicken behaviors* and compute their new XY position using:

$$XY_{(t+1)} = \text{Best value} (\min (VxVy \text{ explosion} - VxVy \text{ edge}))$$

This escaping behavior is for instance not applied

in concurrence with any other mobility strategies or tactical behavior: they have not the possibility to avoid traffic jam or loops. The main effect of the general application of this rule is purely the draining of the transportation network. Of course this "Hollywood" scenario is not relevant in reality but let us test implemented mechanism.

Vulnerability increases when a certain quantity of actors changed their dynamics of mobility, mainly after a shift in their goals. Beliefs, desires and goals are then important to take into account in this kind of model.

V. DISCUSSION

We have defined methods modeling mobility itself, but we now need to model the decision processes for picking or switching between these methods. In an ordinary situation, people follow their own planning and most of the time never deviate of their schedule. But how to justify and explain the fact that in some circumstances, people shift from one behavior to another, from an ordinary behavior to one of the extraordinary described here such as *Sheep* or *Roadrunner* ?

Each agent can be seen as a cognitive agent, where motivation is important in the act of mobility. Motivation is "life dependant", and "contextual dependant": we can say that there is a path dependence of the individual motivations, where the present and future is mainly conditioned by the past; and that sometimes motivations, in a short spatiotemporal range, depart and express something really different. This last conception can be seen as the result of processes such as adaptation, evolution, archaic instinct and so on and so forth. In our debate, this is linked to the fact that people are able to change their plans and that they do not want to keep going to the previously planned destination.

We are not fathoming here the psychological processes that lead from one objective to another, the main point is the result of such behaviors. We have to think of a way to sum-up individual intelligence by simple processes.

Architecture such as Beliefs - Desires - Intentions (BDI) [7] is probably well adapted to this kind of cognitive agent.

- Belief here represents the schedule in normal

situation, information about environment (other mobile agents and road network) and attributes of agents that can describe risk culture, sociability, tolerance level etc. Belief is subject to uncertainty and error. In our model, Belief play a role as representation of industrial accident and its dynamic is important in human behaviors. Both the spatial and temporal distance of the accident can modulate how it is perceived.

- Desires are goals assigned to the agent, they are influenced by beliefs. Desires represent in our model different points to reach in space.

- Intentions represent the priority for an agent to achieve goals: it can represent a sort of utility function where each element is a point in space.

$$f(g)=\{X1Y1; X2Y2;...XiYj\}$$

Intentions are then both a goal and a list of goals.

In a disaster situation, agent receives different kind of information (Beliefs). If following his curiosity or instincts is a predominant goal of the agent (Desires), or fear or cupidity etc. then he will permute elements of his utility function, and even add new elements in order to plan new actions (Intentions).

The main question is then: is it necessary to have a good knowledge of people desires to simulate crowd dynamics and vulnerability of transportation network? In other words, what level of detail is needed in the modeling of individual agent to accurately model the behavior of a crowd of them ?

REFERENCES

- [1] E. Daudé, D. Provitolo, E. Dubos-Paillard, E. Gaillard, E. Eliot, P. Langlois, E. Propeck-Zimmermann, T. Saint-Gérand, "Spatial risks and complex systems: methodological perspectives" in *From System Complexity to Emergent Properties*, Springer series understanding complex systems, in press.
- [2] D. Helbing, I. Farkas, T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, 407, pp. 487-490, Sept. 2000.
- [3] D. Helbing, I. Farkas, P. Molnar, T. Vicsek, "Simulation of pedestrian crowds in normal and evacuation situations," In *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S. D. Sharma, (Eds.). Springer, Berlin, pp. 21-58, 2002.
- [4] J. Epstein, D. Goedecke, F. Yu, R. Morris, D. Wagener, G. Bobashev, "Controlling Pandemic Flu: the value of International Air Travel Restrictions," *PLoS ONE*, 2(5), e401, Mai 2007.
- [5] P. Tranouez, P. Langlois, E. Daudé, "A multiagent urban traffic simulation. Part I: dealing with the ordinary", ICCSA, juillet 2009.
- [6] E. W. Dijkstra: A note on two problems in connexion with graphs. In *Numerische Mathematik*, 1 (1959), S. 269–271.
- [7] M. E. Bratman, *Intentions, Plans, and Practical Reason*, Harvard University Press: Cambridge, MA, 1987.

Use of geosimulations and the complex system theory to better assess flash floods risks in the Paris Basin watersheds (France)

J. Douvinet, *Geophen Laboratory, University of Caen (France)* D. Delahaye, *Geophen Laboratory, University of Caen (France)*, P. Langlois, *MTG Laboratory, University of Rouen (France)*

Abstract—“Flash floods” occurring on the loamy plates of the Paris Basin, in the north of France, are the most dangerous form of floods encountered in this area because of their torrential nature. To better assess occurrence of these events, watersheds have been located thanks to the French CatNat data base. The morphological organization of the watershed between forms, slopes and drainage networks, plays a key role on the “flash floods” dynamics and on the water flow concentration along pathways. Therefore, all the traditional quantitative methods usually describe separately each morphological component and none of them enables to estimate, in a synthetic and dynamic way, impacts of the spatial organization of those three morphological components. Based on the complex systems theory, it is now possible to measure the 3D organization of the catchment area. Methodological implications of this work are very relevant as the tools involved were not currently used by geomorphologists or hydrologists at the beginning of this research. Results obtained and the validity of these spatial analysis methods will have to be discussed at the end of this work, progressively integrating other variables in simulations.

Index Terms— Flash floods, Paris Basin, catchment morphology, Cellular Automata, complex systems.

I. INTRODUCTION

Occurring on the loamy plates of the Paris Basin (France) at the end of spring and during summer, “flash floods” events are the result of violent meteorological events (thunderstorms) concentrated both in time and space. High rainfall intensities (> 50 mm) do not last for more than several hours (< 15 h). These phenomena are the most dangerous and un-known form of floods encountered in the north of France due to their torrential

Manuscript received May, 20, 2009. This work was supported in part by the ACI Complex Systems in SHS, the Geophen Laboratory and the University of Caen – France (sponsor and financial support acknowledgment goes here).

J. Douvinet is ATER in Paris 1 – Panthéon-Sorbonne, in France, 23 rue St Jacques, 75005 Paris, and associated to the Geophen Laboratory, Esplanade de la Paix, F-14035 Caen Cedex (corresponding author to provide phone: 00 (33) 0231566146 ; johnny.douvinet@unicaen.fr).

D. Delahaye is Professor in Geography at the University of Caen (France) and associated to the Geophen Laboratory, Esplanade de la Paix, F-14035 Caen Cedex 2 (daniel.delahaye@unicaen.fr).

P. Langlois is Professor in Geography at the University of Rouen and associated to Graphics Analysis and Modeling (MTG) Laboratory, F-76821 Mont-Saint-Aignan (patrice.langlois@univ-rouen.fr).

behaviour. Human impact on these hazards is indisputable as they usually originate in agricultural ploughed areas. As well as the road network located in the thalwegs of the valleys, concentrating stream flows and aggravating their velocity, the urbanization leads to faster the water flows and increases the vulnerability of societies located in downstream parts (fig. 1).



Fig. 1. Photographs showing the violence and impacts of flash floods occurring in the North of France during summer of at the end of the summer after high rainfall intensities (Douvinet, 2008).

II. METHODS AND DATA

A. A morphological control at regional scale

In a first time, 189 basins subjected to various “flash floods” (1983-2005) have been located thanks to the French disaster “CatNat” data base in order to better assess the spatial and temporal occurrences of the events. Despite rainfall intensities and land use controlling such hazards, the comparative analysis led on all the 189 case studies [1, 2] underlines that the internal morphological organization of a watershed and the relation between the forms, slopes and drainage networks, play a key role on the spatial dynamics of “flash floods” and on the water flow concentration.

At regional scale, two major morphological types have been observed. 94 % of all basins are included in the Type 1: these watersheds are of small size, less than forty square kilometres; they present high-slope incision; as they are situated in the downstream of higher basins, they are characterized by a gap in the Strahler ordination [3]. Other basins representing 6 % form a second group, called the Type 2. The latter present stronger sizes, ranged from 25 to 80 km²; they present weak-slope gradients, and as they are localized in upstream parts of big basins, the Strahler ordination regularly increases.

This result is important as the influence of the topography has always been underestimated in this area characterized by a “no-efficient relief” (fig. 2).

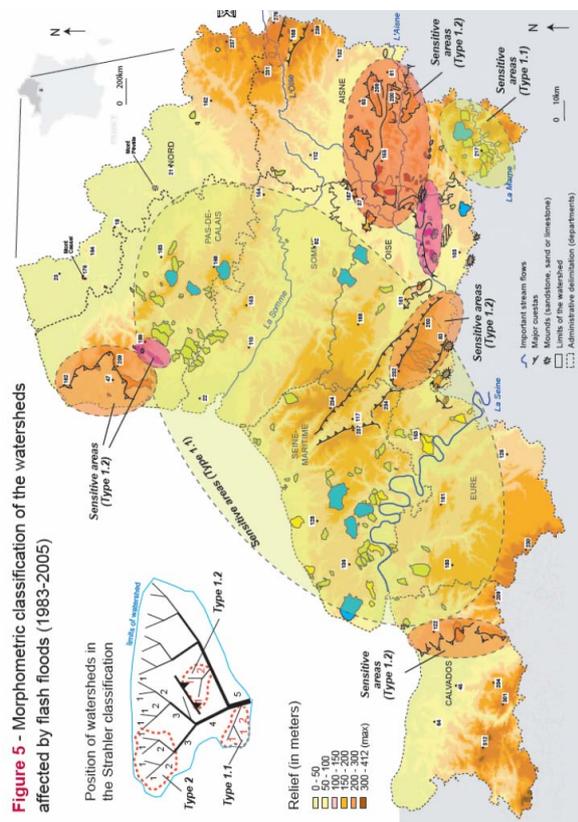


Fig. 1. A morphological control observed at a regional scale; two types have been distinguished on the 189 studied watersheds.

B. Problem statement

In a second time, according to previous results, we decided to better understand the effects of topography on hydrological response on affected basins. Generally, morphometric indexes are used to measure their influence. Therefore, an exhaustive inventory [4, 5] have well shown that this quantitative methods separately describe major morphological component in form indexes, network indexes and slope indexes. Hence, we have index of compacity of Gravelius, Horton ratio ordination,

longitudinal profile section, the hypsometric curve..., but none of these indexes enables to estimate influence of the spatial organization of all morphological components in a synthetic and a dynamic ways [6, 7, 8].

So we develop new methods, especially based on the theory of complex systems and cellular automata, to better analyze the incidence of the network patterns in a certain form and under slopes conditions [9, 10]. Our approach allows us to implement the three morphological components (shape, slope and drainage pattern) in a generalized cellular automaton called RuiCells. It becomes possible to make various simulations according to the implemented parameters. In this study, we focus our attention on results obtained by two simulations: the spatial influence between networks and form in a 2D-dimension; the effects with slopes in a 3D-dimension (fig. 3).

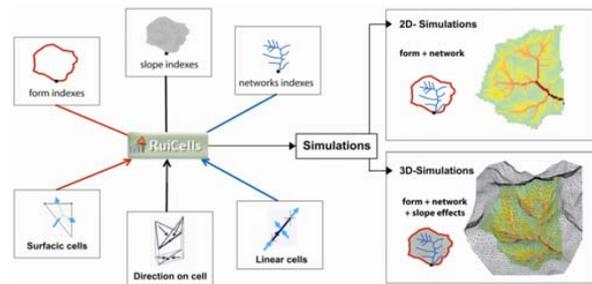


Fig. 3. Cellular Routing Scheme allowing the integration of the morphological component in a dynamic system, and the two simulations types associated.

C. Specific aims of this study

The purpose of this study is to further expand the assessment of catchment morphology on surface hydrological response. The specific aim of our research is threefold: i) to identify the catchment morphology effects on the runoff routing and production; ii) to improve analysis of the relationships between morphological components (catchment shape, slopes, drainage networks) and measure the impacts of the three morphological components on hydrological responses, from local to global watershed scales; iii) to quantify the surface flows variability over space and time on these dry valleys.

III. SIMULATION RESULTS IN 2D-DIMENSION

In 2D-dimensions, the RuiCells Model provides a convenient approach for simulating surface flows on each point of the drainage areas. Studying numerous and complex environmental systems, we notice the importance of the spatial relations between surfaces and networks patterns, but also the minor effects of the catchment shape. Drainage networks effects and slope impacts are increasing as the area also increases. On the other hand, the global basin response can be broke up to identify the evolution of the peak of surface flow on local scales and starting in the downstream. Simulations reveal the “cascading system” between the surfaces.

Use of several points of measurement located on a few sub-catchments allows following the surface flow construction until the overall scale. At this global scale, similar peak flows can be obtained for similar catchment sizes but within various configurations of network patterns. A specific index (fig. 4) identifies the points in upstream of which networks of thalwegs are well organized regarding the basin form and quantifies the structural efficiency in all points of a watershed [1].

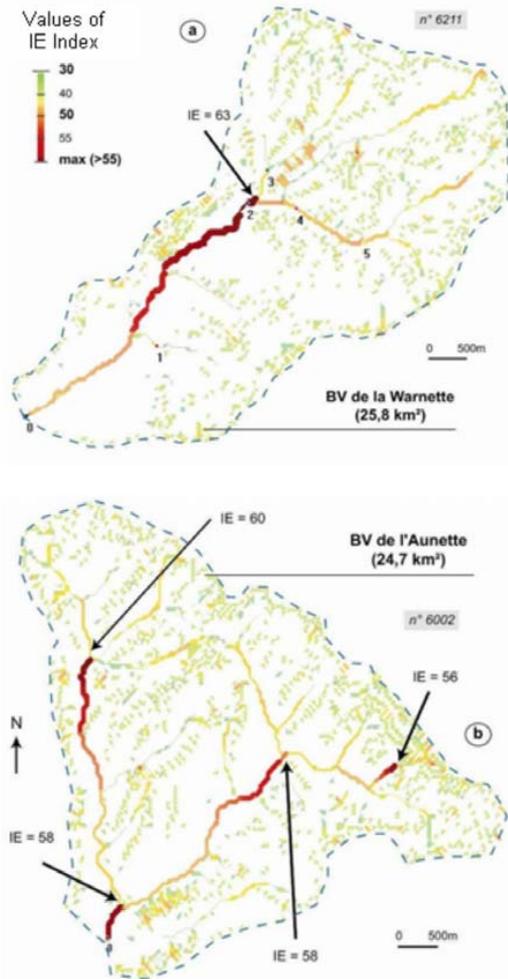


Fig. 4. The IE index underlines the structural efficiency within a watershed and various points can emerge since local to global scales.

So we can identify various configurations explaining such efficiency (fig. 5). This efficiency can be hidden at the final scale whereas it explains the final peak flow and the watershed behavior. Our approach raises the question about the relevance of this global catchment scale: measuring discharges at the downstream of confluences seems to be the “bad solution” if we want to understand the internal behavior and influence of the topography since upstream sources [11, 12].

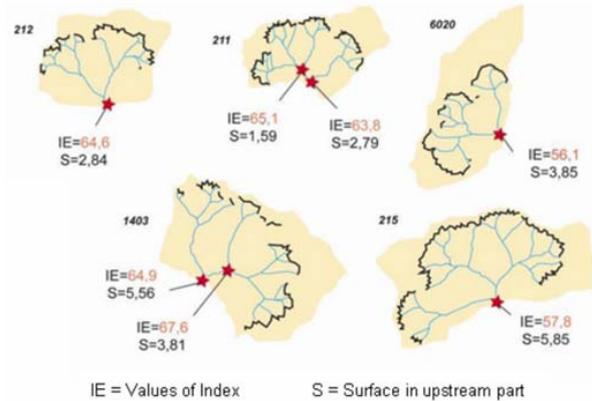


Fig. 5. Various spatial structure and organization between surfaces and network explain high IE Indexes. Various patterns can explain similar efficiency.

This simple and powerful method provides an excellent way to assess the dynamics of “flash floods”, hyperconcentrated flows phenomena more and more occurring on the small dry valleys belong to the Paris Basin, in the North of France.

All the case studies subjected to such flash floods present significant and various morphological signatures that the theory of complex systems and cellular automata make emerging. If this influence was expected, the dynamic and spatial simulations improve their quantification.

IV. SIMULATION RESULTS IN 3D-DIMENSION

Implementing the flow velocity, surface flow graphs have a better hydrological sense as they consider slopes and the water height on each cell. Simulated hydrographs represent now the discharges in $m^3.s^{-1}$ values. As the catchment shape, slopes and the networks are combined, results show a theoretical 3D response of the morphological system. The curves appear really smoothed under the effect of slopes (fig. 6).

Indeed, the larger the basin, the smoother the gauge traces. The simulated curve for the catchment of Aizelles seems to be less disturbed than those for the Les Ouis or the Essômes catchments. The mean average discharge is also the highest for the smallest catchment whereas values decrease on the other areas. Furthermore, although their size is nearly of similar area, surface flow graphs for the Lézarde, Villers and St Martin catchments show major differences.

The peak discharge for the Villers catchments ($10,7 m^3/min$) is greater than those simulated for the St Martin ($9,3 m^3/min$) and for the Lézarde ($8 m^3/min$). A negative relationship, with $r^2 = 0,81$, exists between the time-to-flow-out vs the mean average slopes while the correlation coefficient between peak discharges vs catchment sizes decreases through spatial scales ($r^2 = 0,58$). The time-to-flow out for the Villers, St Martin and Lézarde exceeds those simulated on Essômes while the size of this latter is higher than others.

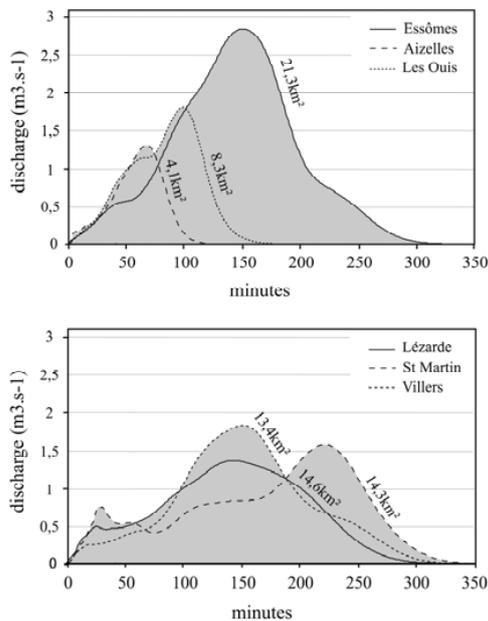


Fig. 6. Simulated graphs obtained with the assumptions of an injection of 1mm in all the cells and with a flow velocity increasing according to the slope angle. Algorithms are more detailed in Langlois and Delahaye (2002).

V. DISCUSSION AND CONCLUSION

Relationships between the catchment morphology and its hydrological response were often difficult to well-establish and this has been one of the main key issues for geomorphology theory since many years [13, 14, 15, 16].

Few studies have shown that the drainage networks structure play a key role on functional entities (Vogt et al., 2003) and so on the theoretical surface flow response. The global response has usually been considered like a linear system and accepts the existence of an instantaneous unit hydrograph (IUH) or of a geomorphological instantaneous unit hydrograph (GIUH) as an global average response [8, 17, 18, 19]. Therefore, in this study, we propose using a simple CA model to integrate the drainage network, but also the surfaces and slopes from the cellular scale to better assess the global response.

Simulation results obtained with RUICELLS bring original replies on hydrological influence of topography. Firstly, they underline the dominant influence of spatial organisations in relation with drainage networks on the surface runoff response, and at the opposite, the minor role of the catchment shape and its size. Second, they indicate a spatial scaling effect on these catchments of 3rd and 4th Strahler-order and highlight similar hydrological behaviours for these catchments independently of their size. Third, they partially explain why these watersheds are prone to flash floods, especially identifying two configurations where the natural hazard risk is dangerous.

REFERENCES

- [1] Douvinet J., Delahaye D., Langlois P., Modélisation de la dynamique potentielle d'un bassin versant et mesure de son efficacité structurelle. *Cybergéo, Revue européenne de géographie* (revue en ligne), 412., 2008 <http://www.cybergeo.eu/index16103.html#tocto1n3>
- [2] Douvinet J., Les bassins versants sensibles aux crues rapides dans le Bassin Parisien – Analyse de la structure et de la dynamique de systèmes spatiaux complexes. *Thèse de Doctorat, Géographie*, Caen, 321 p., 2008
- [3] Strahler A.N., Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, 38, pp. 913-920, 1957.
- [4] Douvinet J., Delahaye D., Langlois P. – Cellular Automata in Physical Geography. *Proc. of the XIVth European Colloquium of Theoretical and Quantitative Geography* (ECTQG), Montreux, 127-132, 2007.
- [5] Douvinet, 2006 (HIC)
- [6] Beven K., Woods E.F., Sivapalan M., 1988. On hydrological heterogeneity – catchment morphology and catchment response. *Journal of hydrology* 100, 353-375.
- [7] Bendjoudi H., Hubert P. (2002) – Le coefficient de compacité de Gravelius : analyse critique d'un indice de forme des bassins versants. *Hydrological Sciences, Journal des sciences hydrologiques*, 47 (6), pp. 921-930.
- [8] Rinaldo, A., Rodriguez-Iturbe, I., Rignon, R., Bras, R.L., Ijjasz- Vasquez, E.J., Mariani, A., Minimum energy and fractal structures of drainage networks. *Water Resources Research* 28, 2183–2195, 1992.
- [9] Delahaye D., Guermont Y., Langlois P., Spatial interaction in the runoff process, *Proc. of the 12th Eur. Coll. on Quantitative and Theoretical Geography*, St Valéry en Caux, France, 2001.
- [10] Langlois P. et Delahaye D., Ruicells, un automate cellulaire pour la simulation du ruissellement de surface. *RIG, Revue Internationale de Géomatique* 12 (4), 461-487, 2002.
- [11] Gupta V.K., Castro S.L., Over T.M., On scaling exponents of spatial peak flows from rainfall and river network geometry. *Journal of Hydrology* 187, 81-104, 1996.
- [12] Vogt J.V., Colombo R., Bertolo F., Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation and environmental characteristics. *Geomorphology* 53, 281-298, 2001.
- [13] Veltri M., Veltri P., Maiolo M., On the fractal dimension of natural channel network. *Journal of Hydrology* 187, 137-144, 1996
- [14] Rodriguez-Iturbe I., Rinaldo A., Fractal river basins, chance and self-organization, Cambridge University Press, Cambridge, 547 p, 1997.
- [15] Schmidt J., Dikau R., Extracting geomorphometric attributes and objects from digital elevation models - Semantics, methods, future needs. In *Dikau, R. & H. Saurer* (eds.): *GIS for Earth Surface Systems - Analysis and Modelling of the Natural Environment*. Schweizbart'sche Verlagsbuchhandlung, pp. 153-173, 1999.
- [16] Mantilla R., Gupta V.K., Mesa O.J. . Role of coupled dynamics and real network structures on Hortonian scaling of peak flows. *Journal of Hydrology* XX, 1-13, 2005.
- [17] Palacio-Vélez O., Gandoy-Bernasconi W., Cuevas-Renaud B., Geometric analysis of surface runoff and the computation order of unit elements in distributed hydrological models. *Journal of hydrology*, 266-274, 1998
- [18] Cudennec C., Calvez R., Pouget J.C., Kingumbi A. Le Goulven P., Constitution et structuration territoriales des ressources, des impacts et des risques hydrologiques au sein du bassin du Merguellil – Perspectives de modélisation hydrologique pour la transposition d'approches de gestion. *Actes du Séminaire PSI, Gestion intégrée de l'eau au sein d'un bassin versant*, Montpellier, France, 2004.
- [19] Nasri S. Cudennec C. Albergel J., Berndtsson R., 2000, Use of a geomorphological transfer function to model design floods in small hillside catchments in semiarid Tunisia. *Journal of hydrology*, 287, pp. 197-213, 2004

Compiled Risks of Spatial Complexity: the Map Algebra Contribution

Emmanuel Bonnet, Thierry Saint-Gérard, Eliane Propeck, David Gaillard Université de Caen, GEOSYSCOM – UMR IDEES 6266 CNRS

ABSTRACT— *This paper proposes a spatial risk analysis method founded on the concept of risk situations. the method is illustrated via the example of the seine estuary and its major industrial risks.*

I. INTRODUCTION

In the scientific community today, it is well understood that the Geographical information System (GIS) makes up one of the most pertinent approaches for discussing the complexity of the real world and the multitude of phenomena which develop, are found in and interact with each other in the real world.

Most of these phenomena can be observed in simple, combined and organised spatial forms which make up what we could call « spatial space ». (1) At any time, this spatial property which they share gives them the capacity to be described and analyzed via simple fundamental “WH” questions such as Where? What? When? and How? These questions, sorted out according to their thematic, temporal and geographic scales, aid in understanding how the spatial phenomena organize and maintain their particular spatial relationships.

This being the case, it then becomes possible to identify the spatial processes which lead to interactive situations. One manner in which to specifically explain and understand space is to create spatial models and theoretical explanations. These models and theories can then be employed to reveal the processes which in turn lead to spatial phenomena. Quantitative geography is full of models (Von Thunen, Modèles gravitaires, Weber, Christaller, Lösch, Lowry, Zipf...) generally defined by theoreticians in the framework of a

particular spatial thematic study. However, they are rarely available as such in the panoply of operators offered by the GIS programs. The stakes are very high when mobilizing these models by activating them using GIS via the spatial analysis functions which the latter regroup.

The objective of this contribution is to demonstrate how the GIS can aid in understanding the complexity of the real world by deconstructing, questioning and combining the geographic information which informs it.

Beyond this first objective, the intent is to show how the usage of map algebra permits us to illustrate spatial phenomena. At the same time, we can bring information to the dynamic systems capable of responding to the question: “What if?” - which is the missing link in the GIS tools.

The final goal is to illustrate the contribution of the GIS, not only in risk analysis situations, but also in the elaboration of new modeling and simulation resources.

The thematic example upon which this work rests is borrowed from the industrial risk thematic; hence, from a sector which is constantly seeking conceptual and methodological advances with which to improve crisis management capacity in the case of uncertainties.

First, we will present the concepts which will be illustrated based on industrial risk research carried out in the Seine estuary.

II. SPATIAL ANALYSIS AND MAP ALGEBRA

Our talk will essentially be constructed around the algebraic mapping methods based on a georeferential network called "GRID". The construction is done via an intermediary of geotreatments which can be called upon, most notably in the form of ArcInfo from the ESRI company.

Map algebra is a method used to achieve spatial modeling with the aid of high level programming language. This language also provides the possibility to perform simple calculations such as mathematical operations, arithmetic, trigonometry, Boolean equations, simple statistics, etc. There is also the possibility of performing complex calculations such as multidimensional statistics, 3D spatial models and hydraulic spatial models on a local as well as a global level.

The usage of map algebra necessitates a specific informational structure from the very beginning. The geographic and thematic information must form a georeferential cell which has specific topological characteristics associated with a dual graph, the summit of which corresponds to the center of each link.

Upon exiting, the new information retrieved from the algebraic mapping operations will present itself in this same cell GRID format. This information will correspond to the results derived from one or several GRID, cell by cell, at local, focal, zonal and global levels.

These local, zonal and global functions were conceptualized by Dana Tomlin in 1990 during his work which took into account spatial variations during the treatment of phenomena. (2)

The local functions take into consideration the phenomena only at the level of a cell.

The focal functions take into account only the phenomena of the neighboring cells. In this type of treatment, the neighboring cells are defined by a

mobile window in which the treatment is realized. The result is allocated to the cell which is in the center of the window.

Zonal Method: The value of the cell in question is attributed in relation to the function of the calculations which are, in turn, based on the values of the cells in a neighboring zone which has been predefined by the user.

There is a fourth method which can be referred to as "global", which takes into account the combined value of the cells in GRID. The geotreatment is applied to the totality of the resulting links.

III. MAP ALGEBRA FOR ANALYSING INDUSTRIAL RISK. EXAMPLE OF THE SEINE ESTUARY

The Seine estuary constitutes one of the largest French concentrations in terms of establishments classed SEVESO 2. About twenty industries are dispersed in the oval portion of the estuary between two principle industrial areas : Le Havre and Port Jérôme (3).

Research financed by the French Minister of Ecology and Sustainable Development, via the Risk, Decision and Territory program, has explored the exposition modalities, the risks and the vulnerabilities to which the population has been exposed.

This analysis, based on a formalization of a previous concept of risk situations, permits us to better understand how spatial phenomena, once combined, can lead to a (or several) risk(s) situation(s) with varying degrees of seriousness.

This concept considers a risk situation to be like a combination and a variation, in a given amount of space. This takes into account various hazards, exposure potential, risks, vulnerabilities, anticipation and recovery.

Initially collected through vector analysis (in vector form), the information was first subjected to

GRID modeling. This operation structured the information on the same spatial cut out, at the same level, in order to proceed to the available informational combinations in a coherent fashion.

Map algebra was especially useful during the creation of computer generated maps with the building of combination analysis on several levels of geographic information.

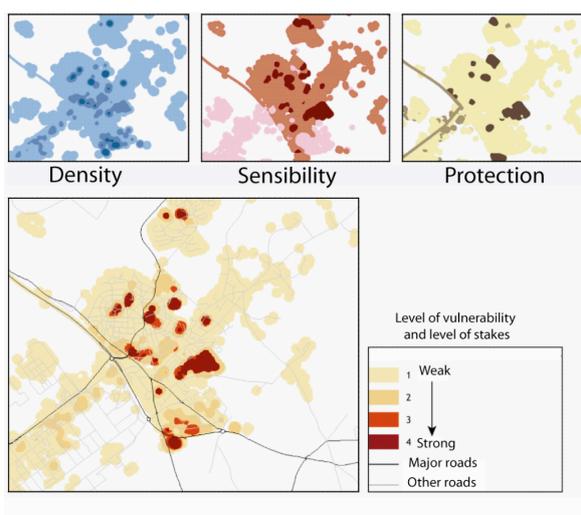


Fig. 1 : Level of vulnerability and level of stakes

The example in figure 1 shows the combination working on several levels of geographical information. The operations created with map algebra permit us to lay out the computer generated map showing the levels of what is at risk as well as the vulnerabilities.

The usage of these techniques permits us to layout a computer generated spatial synthesis which is useful to decision makers and managers as the most exposed areas are vividly pinpointed. These locations simultaneously take into account three criteria.

We can clearly see the manifold possibilities of map algebra.

The possibilities correspond to distinct and pertinent logic used to reconstitute and explain their spatial solidarity. Each one in its own way, the

elements enter into the complexity of risk situations such as exist on the real terrain.

The first and most classic possibility consists in applying operative calculations to the heart of a more or less elaborate equation formula to the thematic values attached to the GRID cells.

In this context, this procedure allows us to calculate (for each cell resulting from the GRID) the sum of the severity of the risk as relates to the corresponding cell in each GRID relating a specific type of risk.

Therefore, we can obtain the global synthetic values of which cartography indicates the sectors of concentration and dispersion inclusive of all risks.

The second possibility consists of creating qualitative information (coding) which express the combination of different types of risk concerning the cell in question: the expression of the risk level yields to the expression of the diversity of the combined risks in one location: this information is qualitative and focuses on a combination.

Finally, another usage consists of using spatial generalization techniques of the cell value in function with the distance and the neighboring criteria specified by the user.

These methods, for example the algorithm of the dominant neighbor, allow us to generate cells which contain the information of the neighboring cells. This information is particularly useful in reconstituting areas. The decision makers need this specific information in risk prevention scenarios.

IV. CONCLUSION

In conclusion, this example applied to the domain of industrial risk shows the GIS offers multiple spatial analysis resources to treat the complexity of spatial phenomena both in terms of their comprehension and management.

However, the map algebra technique is under employed in France.

The potential of map algebra represents a level of interest well above that of the themes touched upon as an example in this extract.

The potential indicated in this extract merits validation by any sector requiring centralized information that can be employed in territorial decision-making.

V. REFERENCES

- [1] F. Pirot ; T. Saint-Gérand and al., «L'analyse spatiale versus algèbre de cartes sous ArcGIS. Des exemples en sciences de l'homme et de la société » in Conférences ESRI, 2005 15 p
- [2] Tomlin C. Dana, 1990
- [3] E. Propeck and al., « Probabilités, risques et gestion territoriale : champs d'action des PPRT », in Géocarrefour, vol. 82/1-2 – Risque : de la recherche à la gestion territorialisée, 2007, 25p

Modeling and Simulation of Pedestrian Behaviors in Transport Areas: The Specific Case of Platform/Train Exchanges

Jérémy Fiegel, Arnaud Banos, Cyrille Bertelle

Abstract—The RATP (Paris Public Transport System) needs a software to optimize the passengers' exchanges between the trains and the platforms in order to improve the train frequency and to reduce risks within passengers congestion. The SimTRAP prototype (Simulation of exchanges between TRains And Platforms) is an agent based model, using a microscopic approach, which is being built for this task.

Index Terms—Agent based models, microscopic approach, self-organization, pedestrian behaviors.

I. INTRODUCTION

WITH more than 1.8 billion passengers a year on its train network (in 2005), the RATP (Paris Public Transport System) is continuously confronted to problems of management of crowds. Concerned by the quality of service, the RATP manages this phenomenon by optimizing the various steps met in a trip. From this perspective, it started a plan aiming at modeling and simulating passengers' exchanges between a train

Manuscript received May 18, 2009.

J. Fiegel is with the RATP (Régie Autonome des Transports Parisiens), Paris, France, and the LITIS (Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes), University of Le Havre, Le Havre, France (phone: 01-58-76-45-41; e-mail: jeremy.fiegel@ratp.fr).

A. Banos is with the LIV (Laboratoire Image et Ville), University of Strasbourg, Strasbourg, France (e-mail: arnaud.banos@lorraine.u-strasbg.fr).

C. Bertelle is with the LITIS (Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes), University of Le Havre, Le Havre, France (e-mail: cyrille.bertelle@univ-lehavre.fr).

and a platform. The resulting tool could help the company to optimize the arrangements on these spaces, to lower the risks of incidents like congestions or falls on tracks and therefore to higher trains frequency and general impression of people towards RATP. Improvements in the evacuation of trains and platforms could also be found.

II. THE EXCHANGES BETWEEN TRAINS AND PLATFORMS

Most of the difficulties which may alter the functioning of trains services occur during the train / platform exchanges: congestions in front of doors, passengers who obstruct the closing of doors... These situations, magnified by the increasing density of passengers, increase the stopping time of the trains and therefore, cause delays.



Fig. 1. Congestion in front of doors during an exchange between a train and a platform

RATP specialists have been using models of simulation of exchanges for a long time. However, these models mostly belong to a macroscopic approach of the phenomenon: they are based on the management of crowds rather than the management of individuals (resulting in losses of precision, notably).

Our objective is to propose a microscopic approach, therefore focusing on individual behaviors, which would provide better estimates, both at micro and macro levels. More generally, we assume that pedestrian behaviors can be seen as a self-organized process, merely based on microscopic interactions in a constrained environment.

III. MODELING AND SIMULATING PEDESTRIAN BEHAVIORS AT MICROSCOPIC LEVEL

Following [1], we identify five key issues to be addressed:

1. Defining a detailed environment with an adapted scale;
2. Reaching adapted spatial and temporal precision;
3. Managing a realistic number of simulated pedestrians;
4. Introducing physiological and behavioral heterogeneity;
5. Combining both local and global interactions.

N. Pelechano et al., in [2], distinguish three main types of modeling approaches: physical models, cellular automata models and rule based models.

1. Physical models, like the famous “Social Force Model” [3] and its recent extensions [2] are able to reproduce some of the self-organizing components of crowds behaviors, but require a large computation effort even for simple environments;
2. Cellular automata models [4]-[5] focus on local interactions between neighboring spatial entities, in which are included desired individual behaviors. They are easier to develop and run faster than the physical models. However, the homogeneous behavior of the individuals

within spatial entities and the limitation of their interactions in relations of spatial nearness can not reflect the real pedestrian behaviors, as concludes [6];

3. Rule based models, like agent based models, are able to deal with more complex environments and behaviors [1]-[7]-[8].

Our SimTRAP prototype directly belongs to that last family.

IV. THE SIMTRAP PROTOTYPE (SIMULATION OF EXCHANGES BETWEEN TRAINS AND PLATFORMS)

The two versions of SimTRAP we created deal with detailed environments (platforms and trains), composed of both static and dynamic objects (trains, doors and folding seats) and are built on the NetLogo system¹.

A. First approach

In this first version, all objects are defined by a point (their center) and a rectangular shape (their bounding box). Some of them have other attributes and can use procedures which allow them to move or which allow the passengers to interact with them:

- a door can open or close with a given speed and until a maximum size;
- a folding seat can open or close when a passenger is sitting on or leaving it;
- all objects of the train can move together with a given speed and a given direction;
- all seats can be free or not.

Passengers are represented by circular agents having, as a first approach, the same internal structure and behaviors. They are defined by their destination, their direction, their speed, their position, their field of vision.

Their destination is determined by their goal which is one of the following:

- Find a good location on the platform (standing or sitting) (1);
- Find a way to enter the train (2);

¹ NetLogo: Web site: ccl.northwestern.edu/netlogo
© 1999-2008 Uri Wilensky. All rights reserved.

- Find a good location in the train (standing or sitting) (3);
- Find a way to get of the train (4);
- Find a way to an exit of the platform (5).

The (1) to (3) goals are only for the “entering” passengers (starting with (1)) and the last ones for the “leaving” passengers (starting with (4)). (2) begins when the train opened its doors, (3) when a passenger entered in the train and (5) when a passenger got of the train. If a passenger has not any goal, he is waiting and does nothing.

The current destination, direction and speed of a passenger can be modified in some cases by the local density (according to the number of other passengers in his field of vision) and by obstacles (an object or a passenger). When an obstacle has been detected, a new direction is computed with a shortest path algorithm.

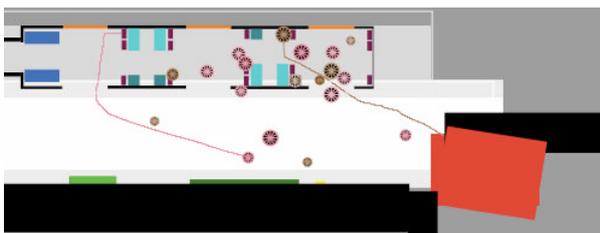


Fig. 2. Screenshot of the first version of SimTRAP showing passengers leaving and entering the train

B. Second approach

A second version of SimTRAP was created to improve the spatial precision. The rectangular and circular shapes used in the first approach were too approximative and not effective.

This time, all objects from trains and platforms which can be generated by the prototype, have exactly the same forms than in reality. Even the trains with two levels can be used in this model.

For this, we use the GIS extension of NetLogo which can load vertexes from a shape file (in

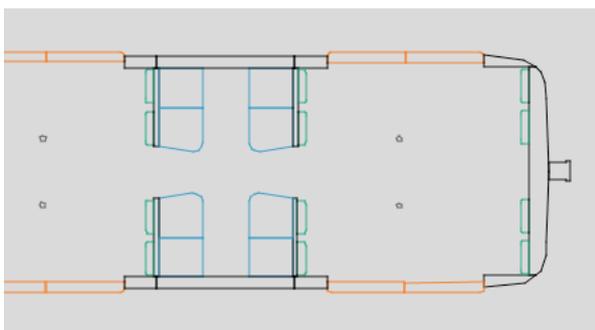


Fig. 3. Details of a train in the second version of SimTRAP

which a fully detailed representation of a train or a platform is stored). Furthermore, the objects are here represented by a point (its center) and a list of its vertexes. Each vertex is linked with two other vertexes of the list in a specific order and knows its owner.

Passengers are represented here by an oval-shaped body and two shoulders (linked to the body). They are twice larger than thick, according to the canons of drawing, and are defined by their gender, which determines which one of the two given sizes they have to use (several men are larger than women).



Fig. 4. Representation of a passenger in the second version of SimTRAP

For example, with these representations, tests of collision are easier (we can use some built-in procedures of NetLogo with its GIS extension).

V. FIRST RESULTS

The first version of SimTRAP allows testing scenarios, for a given set of parameters. For example, figure 5 shows the number of exchanges (passengers entering into plus passengers getting of the trains) in 5 seconds real time, when the number of passengers on the platform and in the train varies.

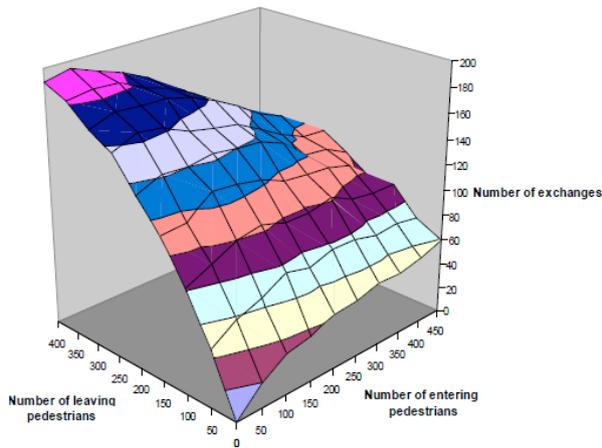


Fig. 5. Simulated number of exchanges in 5 seconds real time according to entering and leaving passengers (MP89 train).

Video analysis where also conducted, in order to calibrate some key parameters. Figure 6 shows the distribution of passengers waiting for the train, according to their distance to the entrance of the platform.

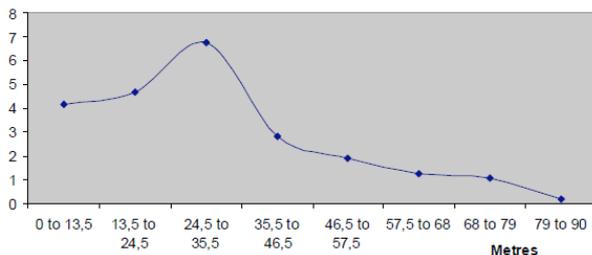


Fig. 6. Video analysis of the distribution of passengers on a given platform

The second version allows testing the capacity of a train. For example, table 1 shows a comparison between theoretical results and SimTRAP results of a test of capacity of a MF67 train (5 wagons). The passengers' characteristics used in these SimTRAP tests are:

- 47% men, 53% women;
- 62 cm for a man's broad;
- 55 cm for a woman's broad².

TABLE I
TEST OF CAPACITY OF A MF67 TRAIN (5 WAGONS)

	Theoretical results	SimTRAP results	Difference
Surface (standing zone)	113.8 m ²	115.2 m ²	1.2%
4 passengers/m² (standing + sitting)	575	582	1.2%
6 passengers/m² (standing + sitting)	806	813	1.1%
Max. passengers/m ²	Never tested	1022 (7.8 p/m ²)	

These tests prove that our environments are realistic. The results depend on the quality of the original maps, but even in the worst cases, the differences between theoretical and simulated capacities are not over 5%.

VI. ECO-PROBLEM-SOLVING

Today, we focus on the passengers' behaviors. We actually try to adapt the eco-problem-solving system, introduced by J. Ferber in 1989 [10].

In this system, the agents are reactive ones and have only two behaviors:

- they search for a state of "satisfaction" ;
- they flee from states in which they can not be satisfied.

In the case of the platform/train exchanges, the agents can be satisfied by modifying their standing or sitting location, but also by staying where they are and trying to pressure the close and disturbing passengers.

So, each agent computes two costs each time it is not satisfied anymore:

- a cost of displacement, which represents the "willpower" required to move to another location ;
- a cost of position, which represents the "willpower" required to stay at the same location.

The first cost is function of the attractiveness of the targeted location, the length of the planned path and the number of agents who are on it (and who must be disturbed).

The second cost is function of the attractiveness of the current location and the power of the disturbance generated by the neighboring passengers.

Like humans, the agents do not have an infinite “willpower”. We put a limit for each cost : when an agent's cost of displacement or position is beyond its limit, this person do not want at all to go to the targeted location or can not stay longer at the same place. If a passenger really has to move or to stay, because the other choice is worst, he will pressure much more his neighbors.

Some useful tips can be used with these rules. For example, if we want to set passengers' goal to “enter the train which just opened its doors”, we just have to put their cost of position beyond the limit while they are still staying on the platform (to do so, we could change the attractiveness of every places in the train by a much higher value and reduce the attractiveness of the platform).

Furthermore, the limits can be personalized for each agent. So, we can simulate more complex behaviors with an easier way. For example, selfish persons can be represented by agents whose limit of the cost of position is higher than the others' limit (the pressure does not disturb them much: they just want to stay at the same location).

This system has not been tested yet and still needs some improvements, but we can see that it offers some great possibilities.

- [1] A. Banos, A. Charpentier, “Simulating pedestrian behavior in subway stations with agents”, *Proceedings of the 4th European Social Simulation Association*, Toulouse, France, September 10-14, 2007, pp. 611-621.
- [2] N. Pelechano, J. M. Allbeck, N. I. Badler, “Controlling individual agents in high-density crowd simulation”, *Eurographics 2007, ACM SIGGRAPH Symposium on Computer Animation*, 2007, 10 p.
- [3] D. Helbing, P. Molnar, I. Farkas, K. Bolay, “Self-organizing pedestrian movement”, *Environment and Planning B*, n°28, 2001, pp. 361-383.
- [4] V. Blue, J. Adler, “Emergent fundamental pedestrian flows from cellular automata microsimulation”, *Transportation Research Record*, vol. 1644, 1998, pp. 29-36.
- [5] M. Muramatsu, T. Irie, T. Nagatani, “Jamming Transition in Pedestrian Counter Flows”, *Physica A: Statistical Mechanics and its Applications*, vol. 267, 1999.
- [6] K. Teknomo, *Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model*, Ph. D. Dissertation, Tohoku University, Japan, 2002, 141 p.
- [7] M. Batty, *Cities and complexity*, MIT Press, Cambridge, 2005, p. 565.
- [8] S. Paris, J. S. Pettré, S. Donikian, “Pedestrian reactive navigation for crowd simulation: a predictive approach”, *Eurographics 2007*, vol. 26, n°3, 2007, pp. 665-674.
- [9] ObEpi – Roche survey, “4e enquête épidémiologique nationale sur l'obésité et le surpoids en France”, 2006.
- [10] J. Ferber, “Objets et agents : une étude des structures de représentation et de communications en Intelligence Artificielle”, Ph. D., University of Paris VI, 1989.

REFERENCES

Agent-Oriented Approach for Detecting and Managing Risks in Emergency Situations

Fahem Kebair and Frédéric Serin

Abstract—This paper presents an agent-oriented approach to build a decision support system aimed at helping emergency managers to detect and to manage risks. We stress the flexibility and the adaptivity characteristics that are crucial to build a robust and efficient system, able to resolve complex problems. The system should be independent as much as possible from the subject of study. Thereby, an original approach based on a mechanism of perception, representation, characterisation and assessment is proposed. The work described here is applied on the RoboCupRescue application. Experimentations and results are provided.

Index Terms—Assessment agents, clusters, decision support system, factual agents.

I. INTRODUCTION

The use of Decision Support Systems (DSSs) has considerably increased, during the last decade, due to the complexity of the problems faced by the decision makers. Indeed, the need for decision support tools should be, if anything, increasing [10]. In some domains or circumstances, making a decision is an arduous task that requires some abilities exceeding the human capacities. We can think decision-making in Simon's decision making model, which consists in intelligence, design and choice [11]. Based on this model, the complexity of decision making lies in the difficulty to get a clear insight into the problem to resolve, to process the vast amount of collected information, to make the right choice in time and to harmonise finally the set of decisions made by the decision makers or the organisations. Therefore, computer-based systems may be very helpful to support decision making, especially when the environment problem is complex, dynamic and partially known. Processing and managing information issued from such an environment represents a challenge to the DSS developers. However, DSS are well known to be customized for a specific purpose and can rarely be reused. Moreover, DSSs only support circumstances which lie in the known and knowable spaces and do not support complex situations sufficiently [4]. This led us to think DSSs must be flexible and adaptive to be effective in solving

complex problems as the risk and crisis management. Flexibility allows the use of the system in different subject of studies with minor changes. In other words, the system operates in a generic manner and relies on specific knowledge that are defined by experts of the domain. Adaptivity is an essential characteristic to build intelligent information systems which draws increasingly the attention of the scientists in computer science and in artificial intelligence. Thanks to the adaptivity, the system may adapt its behaviour autonomously by altering its internal structure and changing its behaviour to better respond to the change of its environment. The multiagent systems technology is an appropriate solution to achieve these two objectives. Intelligent agents [13] are able to self-perform actions and to interact with other agents and their environment in order to carry out some objectives and to react to changes they perceive by adapting their behaviours.

In this paper we propose an agent-oriented approach aimed at building a DSS that has as role to help emergency managers to detect and to manage risks in emergency situations. The system perceives facts occurred in the environment, represents them and analyses them to assess the current situation. To evaluate the situation, the system uses an analogical reasoning based on the following postulate: if a given situation A seems like a situation B, then it is likely that the consequences of the situation A will be similar to those of B. Consequently, the risk appeared in B become a potential risk of A. An internal multi-level kernel is used to insure the whole decision-support process. We utilise an earthquake scenario using the RoboCupRescue Simulation System (RCRSS) [7][9] in order to illustrate our approach. Experimentations and results are provided and discussed. spacing.

II. DECISION SUPPORT SYSTEM FOR RISK DETECTION AND MANAGEMENT

A. Definitions and Approaches

The Risk is a concept that denotes a potential negative impact to an asset or some characteristic of value that may arise from some present process or future event. There are many more and less precise definitions of risk. They do depend on specific applications and situational contexts. It can be assessed qualitatively or quantitatively. In our context, we are interested in natural and technological risks. The management of these risks often represented a large-scale challenge for the individuals and the organisations, since they are hard to predict and their occurrences are much sudden. The

F. Kebair is PhD student in computer science with LITIS--Laboratoire d'Informatique de Traitement de l'Information et des Systèmes, University of Le Havre, 25 rue Philippe Lebon, 76058, Le Havre, Cedex, France, e-mail: fahem.kebair@univ-lehavre.fr

F. Serin is professor assistant in computer science with LITIS, e-mail: frederic.serin@univ-lehavre.fr

risk management may be defined as the systematic application of management policies, procedures and practices to the tasks of establishing the context, identifying, analysing, evaluating, treating, monitoring and communicating risk [1]. This process is complex and exceeds widely the human abilities. The use of the DSS in this case is indispensable. Indeed, DSSs are interactive, computer-based systems that aid users in judgment and choice activities. They provide data storage and retrieval but enhance the traditional information access and retrieval functions with support for model building and model-based reasoning. They support framing, modeling, and problem solving [2]. In the context of the risks and crisis management, the DSS must insure the following functionalities:

- Evaluation of the current situation, the system must detect/recognize an abnormal event;
- Evaluation/Prediction of the consequences, the system must assess the event by identifying the possible consequences;
- Intervention planning, the system must help the emergency responders in planning their interventions thanks to an actions plan (or procedures) that must be the most appropriate to the situation.

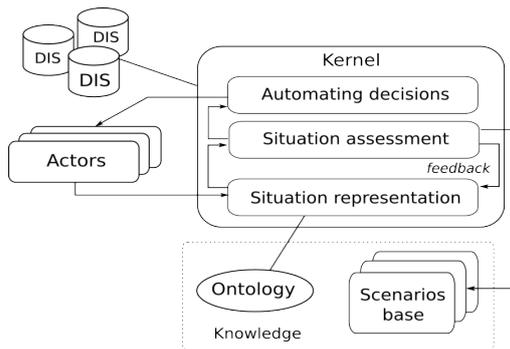


Fig. 1. Whole DSS architecture

The kernel is the main part of the DSS and has as role to manage all the decision-support process. The environment includes essentially the actors and Distributed Information Systems (DIS) and feeds permanently the system with information describing the state of the current situation. In order to apprehend and to deal with these information, specific knowledge related to the domain as ontologies and proximity measures are required. The final goal of the DSS is to provide an evaluation of the situation by comparing it with past experimented situations stored as scenarios in a Scenario Base (SB).

The kernel is a MAS operating on three levels. It intends to detect significant organisations that give a meaning to data in order to support finally the decision making. We aim, from such a structure, to equip the system with an adaptable and a partially generic architecture that may be easily adjusted to new cases of studies. Moreover, its suppleness makes the system able to operate autonomously and to change its behaviour according to the evolution of the problem

environment. As follows a description of each level:

- *Situation representation*: One fundamental step of the system is to represent the current situation and its evolution over time. Indeed, the system perceives the facts that occur in the environment and creates its own representation of the situation thanks to a factual agents organisation. This approach has as purpose to let emerge subsets of agents.
- *Situation assessment*: A set of assessment agents are related to scenarios stored in a SB. These agents scrutinise permanently the factual agents organisation to find agents clusters enough close to their scenarios. This mechanism is studied "manually" by an expert of the domain and is similar to a Case-Based Reasoning (CBR) [8], except it is dynamic and incremental. According to the application, one or more most pertinent scenarios are selected to inform decision-makers about the state of the current situation and its probable evolution, or even to generate a warning in case of detecting a risk of crisis. The evaluation of the situation will be then reinjected in the perception level in order to confirm the position of the system about the current situation. This characteristic is inspired from the feedbacks of the natural systems. In that manner, the system learns from its successes or from its failures.
- *Automating decisions*: Outcomes generated by the assessment agents are captured by a set of performative agents and are transformed in decisions that may be used directly by the final users.

B. RoboCupRescue Case Study

The RCRSS is an agent-based simulator which intends to reenact the rescue mission problem in real world. An earthquake scenario is reproduced including various kinds of incidents as the traffic after earthquake, buried civilians, road blockage, fire accidents, etc. A set of heterogeneous agents (RCR agents) coexist in the disaster space: rescue agents that are fire brigades, ambulance teams and police forces, and civilians agents. We focus, in this application, on the development of the rescue agents behaviours. Our final goal is to use the DSS in order to improve their decision-making ability and to support them during their rescue operations.

A model of the RoboCupRescue disaster space and the properties of its components, and the RCR agents are detailed in [12]. We use this model in order to extract knowledge and to formalise information.

III. DYNAMIC REPRESENTATION OF THE SITUATION: FACTUAL AGENTS

The system perceives and represents the facts occurred in the situation in an original manner using factual agents. Factual agents are reactive and proactive agents according to the agents definition given in [13]. Each agent carries an elementary datum that represents an observed fact and that aims to manage it over time. This information is presented in the shape of a Factual Semantic Feature (FSF), more details

about this structure and how it is formalised and managed by a factual agent is provided in [6].

The objective by using factual agents in the representation situation level is to reflect the dynamic change of the situation and to let emerge, from this view, agents subsets. These subsets may be representative of some situations that are close to some others encountered in the past. The analysis of these agents groups is based on geometric criteria, insuring thus the independence of the treatment from the subject of study. Each factual agent exposes behavioural activities that are characterised thanks to numerical indicators. The latter form a behavioural vector that draws, by its variations, the dynamics of the agent during its live. This gives a meaning to the state of the agent inside its organisation and consequently to the prominence of the semantic character that it carries.

The goal of our approach is to characterise the factual agents organisation by forming dynamically agents clusters and comparing them with stored scenarios. The clustering algorithms seem appropriate to this objective, since they are able to create objects groups in an unsupervised way. However, these methods present some deficiencies in our case. The main ones are the need to specify some parameters as the minimal distance between two objects, required by density-based algorithms [3]; or the minimal length of a cluster, required by Kmeans algorithms [5]. Moreover, the experimentations we led using these methods showed us that we are unable to analyse instantaneously the obtained clusters neither to reproduce them. We changed therefore our way for proceeding by confiding this task to the assessment agents. These agents will search through the factual agents in order to form clusters, that should be the closest to the scenarios to which they are linked. We think this approach is more suitable for our problem, since it does not require specific knowledge and we are certain that the obtained clusters have probably a meaning and may be easily interpreted. In addition we may exploit the assets of the agents, especially their adaptivity and their communication abilities.

IV. SITUATION ASSESSMENT

A. Assessment Agents

Each assessment agent is linked to a scenario stored in the SB (see Fig. 2). Each scenario is composed of one or more factual agents clusters, this depends on the treated application. A cluster is made up of a set of elements, each one includes an FSF, the indicators values of the factual agent associated to this FSF and the size of its Acquaintances Network (AN). Thus, a cluster element has the following structure: $FSF:V_{I1}...V_{In}:S_{AN}$, with V_I a value of indicator I , and an example of an FSF is (fire, intensity, strong, location, 2nd street, time, 10:00 pm).

The role of the assessment agents is to scrutinise permanently the organisation of the factual agents in order to extract agents clusters that should be similar as much as possible to their scenarios. A relevance, which is the sum average of all the similarities values of a created cluster elements, is attributed to each cluster to indicate its proximity

to a stored scenario. This value is included in a range of [0,1]. The more the relevance is near to 1, the more the cluster is close to its scenario maker and vice versa. The clusters, and consequently the assessment agents, are sorted according to their relevances and the selected agents depend on their rank and the size of their clusters i.e. the first agents covering the bulk of the situation are selected.

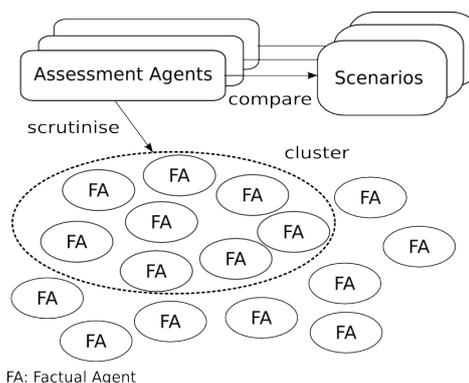


Fig. 2. Role of the assessment agents in the DSS

To find close elements in the factual agents organisation, the assessment agents look only at the numeric properties of the agents and disregard the semantic characters that they carry. This insures the genericity of the mechanism. The assessment agents compare the elements of their scenarios with those carried by the factual agents by computing distances between them. The compared data are vectors defined by the n indicators of the factual agent and its AN size. The cosine similarity measure is used in order to compute the similarity between these vectors. The similarity value is included in a range of [0,1]. A value of 1 means the perfect equality between the two vectors, whereas 0 means their total divergence.

$$CS_{V_1, V_2} = \frac{x_1 x_2 - y_1 y_2 - z_1 z_2}{\sqrt{x_1^2 - y_1^2 - z_1^2} \sqrt{x_2^2 - y_2^2 - z_2^2}}$$

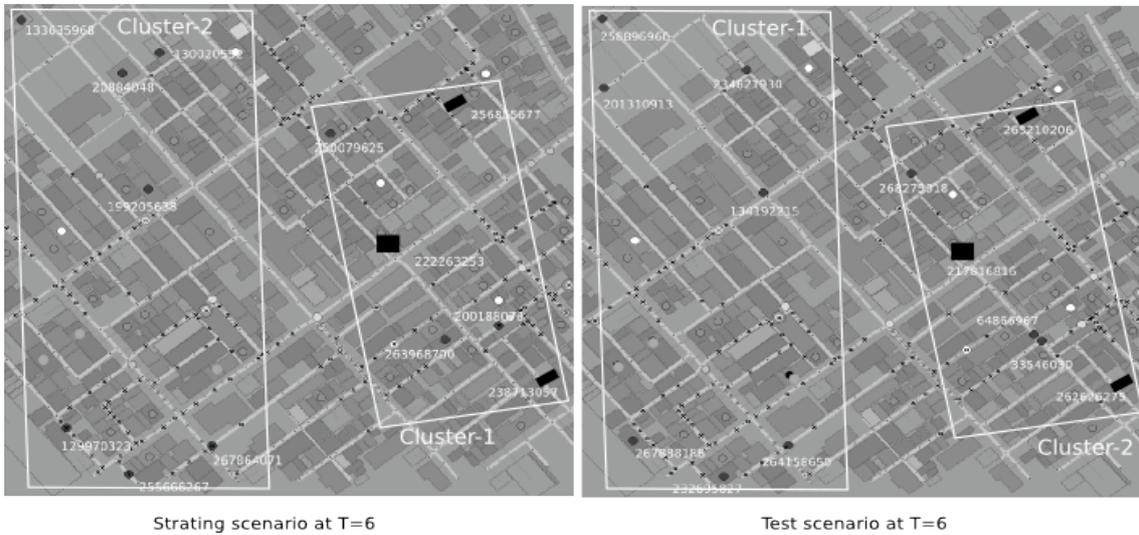


Fig. 3. First test example at the beginning of the RCR simulation

B. Experimentations

We have made experimentations on the RCR application dealing with fires situations. We have developed a prototype allowing the representation and the assessment of risks. The perceived facts in the disaster space are related to the fires propagation and to the fire brigades activities that try to extinguish these fires. The system includes a factual agents organisation for the perception and the representation of the situation and a set of assessment agents to deal with the facts evolution. At this progression stage of our work, the assessment situation is limited to the recognition of factual agents clusters according to past ones defined and experimented beforehand. We have defined therefore, from a starting scenario, a clusters set that we intend to regain in other similar scenarios by forming similar clusters. To modify an RCR scenario, we change the strategy applied by the fire brigades. This allows to have a different perception of the environment and different behaviours of the agents.

Fig. 3. shows two views of the disaster space state at the beginning of the simulation--at the 6th second. The left view belongs to the starting scenario, the right one belongs to a scenario test. What interests us in these views are the fire brigades agents represented by black ellipses and the fires represented by black rectangles. Both objects have white identifiers (IDs), we note that the RCRSS gives randomly new IDs for all the RCR objects in each new simulation. These two elements are represented in the system by two different kinds of factual agents. We have identified two factual agents clusters at this step. Cluster-1 includes starting fires and the first fire brigades having perceived these fires and which are the most able to put out them. Cluster-2 contains however the rest of the fire brigades that are in a passive state.

TABLE I

CREATED CLUSTERS AT THE 6 TH SECOND OF THE RCR SIMULATION		
Stored clusters	Assessment Agents	Similar clusters
Cluster-2: fireBrigade#267864071 fireBrigade#130020552 fireBrigade#129970323 fireBrigade#255666267 fireBrigade#199205638 fireBrigade#20884048 fireBrigade#133635968	Agent-2	Cluster-1, $r=0.99$ fireBrigade#267888188 fireBrigade#264158650 fireBrigade#201310913 fireBrigade#134192215 fireBrigade#234821930 fireBrigade#232695827 fireBrigade#258896960
Cluster-1: fireBrigade#200188078 fireBrigade#250079625 fireBrigade#263968700 fire#238713057 fire#222263253 fire#256855677	Agent-1	Cluster-2, $r=0.89$ fireBrigade#64866967 fireBrigade#268275018 fireBrigade#33546030 fire#265210206 fire#262626275 fire#217816816
Cluster-4	Agent-4	Cluster-3, $r=0.80$
Cluster-3	Agent-3	Cluster-4, $r=0.67$

Table I presents a test example. For this example we have four assessment agents, each one is associated to one cluster in the base. The table shows both the stored clusters elements and those created by the assessment agents. As we see, the two first agents (Agent-2 and Agent-1) regained two analogous clusters with relatively high relevances (r) in the test scenario and cover all the perceived facts of the situation. These two agents are therefore selected as the best candidates to provide the final decisions.

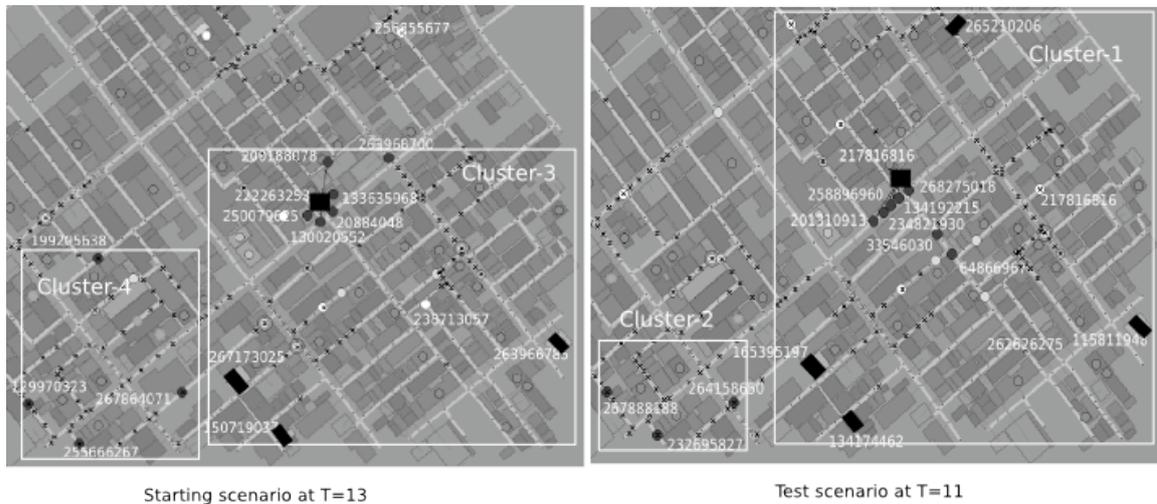


Fig. 4. Second test example in the middle of the RCR simulation}

TABLE II
CREATED CLUSTERS AT THE 11TH SECOND OF THE RCR SIMULATION

Stored clusters	Assessment Agents	Similar clusters
Cluster-3: fireBrigade#200188078 fireBrigade#263968700 fireBrigade#133635968 fireBrigade#20884048 fireBrigade#130020552 fireBrigade#250079625 fire#222263253 fire#263966785 fire#267173025 fire#150719037	Agent-3	Cluster-1, $r=0.83$ fireBrigade#201310913 fireBrigade#134192215 fireBrigade#234821930 fireBrigade#268275018 fireBrigade#64866967 fireBrigade#258896960 fire#265210206 fire#217816816 fire#134174462 fire#165395197 fire#115811948
Cluster-4: fireBrigade#199205638 fireBrigade#267864071 fireBrigade#25566267 fireBrigade#129970323	Agent-4	Cluster-2, $r=0.80$ fireBrigade#264158650 fireBrigade#267888188 fireBrigade#232695827
Cluster-1	Agent-1	Cluster-3, $r=0.78$
Cluster-2	Agent-2	Cluster-4, $r=0.44$

The second example (see Fig. 4) concerns another scenario in an advanced stage of the RCR simulation--at the 13th second of the simulation--in which fires are more important and the fire brigades are more active. At this step, two starting clusters have been identified and stored. Cluster-3 includes fire brigades in full fight with fires and other important starting fires. Cluster-4 presents some isolated fire brigades blocked by debris and that are unable to move. A similar situation is perceived at the 11th second of the test scenario. The most relevant assessment agents are Agent-3 and Agent-4 that succeed in creating two similar clusters, whereas Agent-1 and Agent-2 have retrogressed in the relevances rank.

V. CONCLUSION

We have described in this paper an agent-based approach that aims to build a DSS. The system intends to help emergency planners to detect risks and to manage crisis situations by perceiving, representing and assessing a current situation. We think this approach may be adjusted easily to different problems types and enables the system to have an adaptive behaviour thanks to a multiagent multilevel kernel. We are working currently on the assessment level of the system mechanism. We have presented here first results applied on the RoboCupRescue. We intend to apply this approach on different subjects of studies in order to better improve its generic aspect. We aim also to generalise this approach by setting up a generic modelling of factual agents clusters that will enhance their formalisation and their management.

REFERENCES

- [1] Australian Standard, AS/NZS 4360:2004: *Risk management*, 2004.
- [2] M. J. Druzdzel and R. R. Flynn, *Decision Support Systems*. In *Encyclopedia of Library and Information Science*, vol. 67, pp. 120–133, 2000.
- [3] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), Varna, Bulgaria, pp. 226–231, 1996.
- [4] S. French and C. Niculae, *Believe in the Model: Mishandle the Emergency*. Journal of Homeland Security and Emergency Management, Springer-Verlag New York, Inc., Secaucus, NJ, USA, vol. 2, pp. 1–18, 2005.
- [5] J. A. Hartigan and M. A. Wong, *A k-means clustering algorithm*. Applied Statistics, vol. 28, pp. 100–108, 1979.
- [6] F. Kebair and F. Serin, *Information Modeling for a Dynamic Representation of an Emergency Situation*. Proceedings of the 4th IEEE International Conference on Intelligent Systems IS'08, Varna, Bulgaria, vol. 1, pp. 2–7, 2008.
- [7] H. Kitano, S. Tadokoro, H. Noda, I. Matsubara, T. Takahashi, A. Shinjou, and S. Shimada, *RoboCup Rescue: search and rescue in large-scale disasters as a domain for autonomous agents research*. Proceedings of the IEEE Conference on Systems, Man, and Cybernetics (SMC-99), vol. 6, pp. 739–743, 1999.
- [8] K. Kolodner, *Case-based reasoning*. Morgan Kaufmann, Boston, 1993.
- [9] RoboCupRescue Official Web Site. <http://www.robocuprescue.org/>.
- [10] M. J. Shawn, D. M. Gardner, and H. Thomas, *Research Opportunities in*

- Electronic Commerce. Decis. Support Syst*, 1997, vol. 21, pp. 149–156.
- [11] H. A. Simon, *The New Science of Management Decision*. Prentice Hall PTR, 1977.
- [12] T. Takahashi, *RoboCupRescue Simulation Manual*. Available: <http://sakura.meijo-u.ac.jp/takaHP/kiyosu/robocup/Rescue/manual-English-v0r4/index.html>.
- [13] M. Wooldridge, *An Introduction to MultiAgent Systems*. John Wiley &

Validation of an ontology of risk and disaster through a case study of the 1923 Great Kanto Earthquake

D. Provitolo, J.P. Müller, E. Dubos-Paillard

Abstract— This paper seeks to validate a factual ontology derived from an ontology of the domain of risk and catastrophe (Provitolo, Müller, Dubos-Paillard, 2009). The factual ontology is that part describing the structure and dynamics of the system, that is, the representation of an event and of on what it has a bearing. The objective is to show that the ontology is able to account for and allow comparison between complex stories (because of the diversity of event types and of their multi-scale description). The event used as the basis for validation of the ontology is the Great Kanto Earthquake of 1923. The account is that of P. Hadfield (1991) that provides a detailed description of the earthquake. Judgements by the actors involved of the system elements and events are excluded from this exercise, which is a first stage in validating the ontology.

Index Terms— instantiation, Kanto earthquake, modelling, ontology of disaster, validation

I. INTRODUCTION

IN the field of risk and catastrophe, a great deal of research has been conducted into the concepts [1]-[2]-[3]-[4]-[5]-[6]-[7]-[8]-[9]-[10]-[11] and the analysis of accidents or catastrophes. On the basis of that work, we have proposed a formalized ontology of risk and catastrophe [12].

The most common definition of an ontology is that of [13] who defines it as a specification of the conceptualization of a domain. An ontology is therefore a structure for describing knowledge in a given field. A distinction is generally drawn between the conceptual ontology defining the terminology employed and the concrete ontology, which uses that terminology to describe an actual situation. The formalized ontology that we proposed [12 op. cit.] is essentially a conceptual ontology in which we distinguish the terminology used to depict what happened (e.g. 30 died on some bridge) from the characterization of what happened (there has been an accident, a serious accident or catastrophe) from the standpoint of the various actors of the system.

This work is supported by ANR COSMAGEMS.

E. Dubos-Paillard is Assistant professor with UMR ThéMA, Franche-Comté University (email: edwige.dubos-paillard@univ-fcomte.fr)

J.P. Müller is Senior scientist with CIRAD and LIRMM, Montpellier (email: jean-pierre.muller@cirad.fr; jean-pierre.muller@lirmm.fr)

D. Provitolo is Researcher CNRS with UMR Geoazur, Nice Sophia-Antipolis University (e-mail: Damienne.provitolo@geoazur.unice.fr)

Here we begin with this formalized ontology and look more especially at the part describing the system's structure and dynamics. Henceforth we call this the factual ontology as it allows us to describe what happened without taking up any particular standpoint.

The overall aim of this factual ontology is to be able to handle stories made complex by the diversity of types of events and their multi-scalar character. A conceptual framework needs to be provided within which to analyse the various types of events, whether localized or dispersed, natural, industrial/technological or social. This factual ontology purports also to be suitable for studying events on different scales (micro, meso, macro). This should make it easier both to put into perspective different events that *a priori* share few common features, and to come up with a method for comparing events.

The paper aims first to validate such an ontology by instantiating it based on the factual description of an event: the Great Earthquake of Kanto of 1923 as related by P. Hadfield [14]. This concrete case was chosen because the account mostly describes facts and not standpoints or judgements of those facts by those involved. We are looking here, then, at the representation of the story. This will allow us both to test out our factual ontology and to question the validation process of this sort of ontology.

II. A FACTUAL ONTOLOGY OF RISK AND CATASTROPHE

Factual ontology is that part that allows us to describe the structure and dynamics of a system. As stated in the introduction, the factual ontology is a part of the conceptual ontology of risk and catastrophe presented in [12 op. cit.]. The conceptual ontology enabled us:

- to return to the essential concepts allowing us to characterize risk, accident, catastrophe and the associated notions and to organize these concepts in terms of the relations among them;

- to identify four subsystems making up the conceptual model: *Structure*, *Dynamics*, *Actor* and *Characterization* (Fig. 1). The *Characterization* subsystem enables us to specify how different actors characterize the system's structure and dynamics.

In this paper we propose an initial validation of the structure and dynamics part of this ontology based on the account by P.

Hadfield [14 op. cit.] of ‘The Great Kanto Earthquake’ by reproducing the account using the concepts of this ontology. Actor judgements of the system’s elements and events are excluded from this first validation exercise.

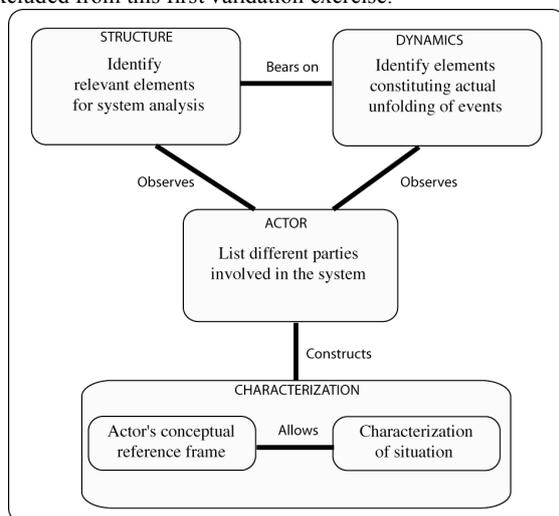


Fig. 1. Systemic division of the meta-model into four subsystems

We therefore present only the Structure and Dynamics subsystems of the conceptual model of risk and catastrophe. This model has been represented by UML (Unified Modelling Language) class diagrams [15]. Ontologies do not have standard graphical means of representation although a graphic depiction is often a very powerful means of communication. UML provides such representations and especially class diagrams representing concepts and their attributes by rectangular boxes and relations between concepts by arrows. Two types of relation are distinguished: the generalization/specialization relation for expressing that one concept is more general/more specific than another (arrow with white triangle pointing to the more general concept) and the semantic relations bearing the name of the relation and the associated cardinalities (how many objects may be related to how many others).

In what follows, we begin with the ontology (and therefore the class diagram) proposed in [12 op. cit.] that we specialize to take account of the specific account. For example, the general notion of event will be specialized into collapse, fire, etc. as particular cases of events.

Thereafter we shall instantiate these notions to reproduce the account proper. Thus, several cases of collapse are mentioned in the account and shall therefore be so many instances of the general notion of collapse. Another UML diagram -the object diagram- shall be used to depict these instances and their structural and temporal linkages. An object diagram represents each instance by a box mentioning the instantiated notion possibly with a name by which to identify it and each connection by an arrow. These connections are

themselves instances of semantic relations between notions. For example, it can be said in conceptual terms that one event may cause another (that there is a semantic relation of cause and effect between events) and the connections will be able to express which events actually did cause other events according to the account.

A. The structure of the system

The *Structure* identifies the relevant elements for analysing a system open to potentially catastrophic events. The elements are the parts forming the system’s structure (Fig. 2). The system is open to its environment (in the systemic meaning of the term). It is therefore also composed of exogenous elements that are by definition outside of the field of study. The ‘Element’ class generalizes the ‘Living Element’, ‘Physical Element’, ‘Organization’ and ‘Infrastructure’ classes that appeared to us to be the relevant categories to be distinguished in the case of risk and catastrophe:

- living element includes all human beings and natural populations such as plants and animals;
- physical element corresponds to the description of the earth’s surface (oceanography, hydrography, pedology, relief, etc.) and does not directly pertain to human activities;
- organization is a structure for responding to needs and achieving set objectives. Organization integrates systems for preventing and managing events.
- miscellaneous infrastructures encompass built areas, facilities, networks, etc.

Instantiating the conceptual model has required the inclusion of new relations:

- the first to indicate that a system may be a particular case of an element. Adding this relation makes it possible to represent the interleaving of spatial levels (scale) within a system (systems are thus composed of systems, which are themselves made up of systems, etc.);
- the second to indicate the existence of neighbourhood relations between elements of the system. The concept of neighbourhood refers to a topological space. It provides more possibilities than the simple use of distance between spatial entities (metric space). It allows us to form spatial subsets by neighbourhood (first-, second-order continuity, etc.). We shall see that this neighbourhood relation allows us to make a territorial analysis of the event, or more accurately the series of events, that occurred in the city of Tokyo.

This first instantiation therefore enhanced the structure of the system.

The system structure is related to the dynamic subsystem since both events and damage bear on the ‘element’ class.

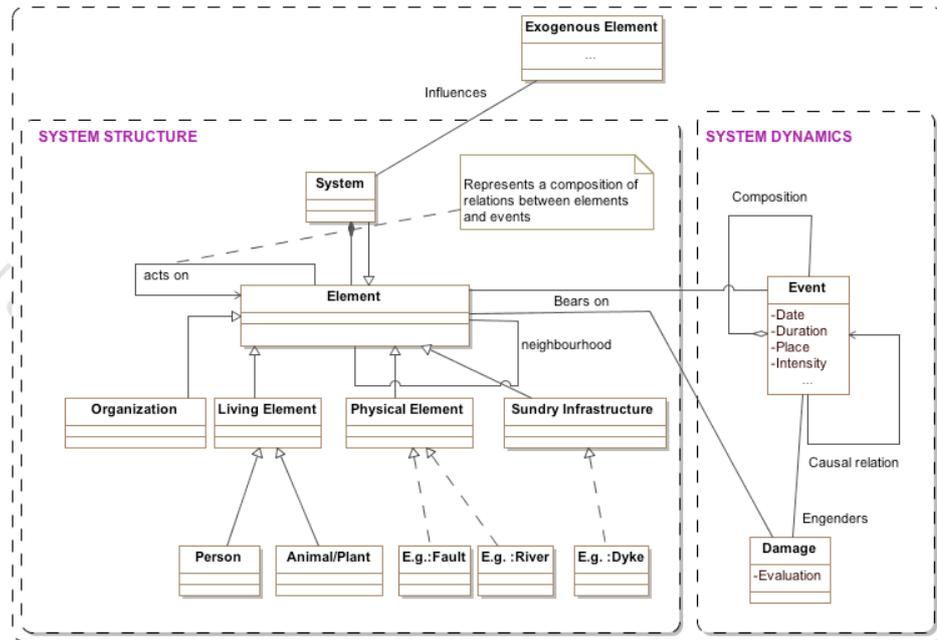


Fig. 2 The system structure and its relations with the dynamic subsystem

B. System dynamics

Just as the system structure is composed of elements, the system dynamics (Fig. 3) is composed of elementary structures in the form of events. Each event may be ascribed a date (at a given level of granularity) and a duration. Each

event may be made up of events, which provides an understanding of the interleaving of temporal levels in the event structure in parallel with the interleaving of spatial levels.

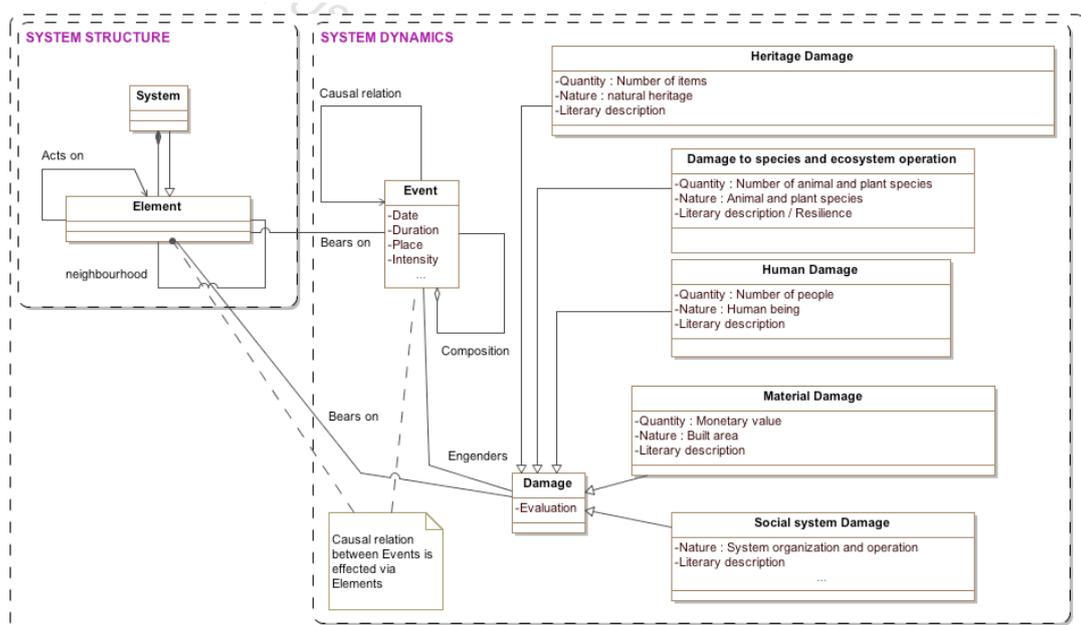


Fig. 3 System dynamics and its relations with the Structure subsystem

An event has a bearing on one or more elements of the system. The 'bears on' relation allows us to make a link, the cogwheel between and event and an element. Elements are in fact nothing other than the 'matter' of which events are part.

An event may also cause another event. Events are then tied to each other by causal relations that are achieved via elements: one speaks of causal chains between events. This causal chain corresponds to domino effects that are often cited in the literature on risk and catastrophe, especially by [16]-[17]-[18]-[19]-[20]-[21]. It shall be seen later in this paper that it is a set of causal chains that allows the Kanto earthquake to be characterized as 'The Great Kanto Earthquake'.

Lastly, the event can engender damage of different kinds and in variable amounts. Damage bears on elements, which explains why the classification of damage is based on that of elements. The classes titled 'human damage, 'damage to species and ecosystem operation', 'material damage', 'damage to social system' and 'damage to heritage' specialize the 'Damage' concept. Such damage may be the subject of quantifications or of literary descriptions. It is generally defined in human or material terms. But it may also bear on economic and financial systems and on heritage whether natural or cultural.

III. INITIAL VALIDATION OF THE FACTUAL ONTOLOGY BASED ON THE ACCOUNT OF THE GREAT KANTO EARTHQUAKE (1923)

The aim is to determine to what extent the ontology developed can account for complete and complex stories (structural complexity related to the variety of type of event, elements, complexity of spatial scales and levels of organization).

In this paper, validation is effected by instantiation on a wealth of situations within the chosen account so as to check whether the model constructed applies to different types of concrete situation. Thus, to test whether the factual ontology is robust, it was instantiated (using *Magic Draw*) from P. Hadfield's account of the Great Kanto Earthquake of 1923. That account is rich enough for the factual ontology to be tested. It allows it to be tested on a large number of system elements (living elements, physical elements, infrastructure), on different types of event, whether localized (earthquake/natural origin) or widespread (fire/technological origin) and at different spatial scales (cities, districts) and at different levels of organization (population, individual). By contrast, this account makes no mention of the characterization of these events by their various actors.

A. Presentation of the case study: the Great Kanto Earthquake, 1923

The Great Kanto Earthquake (1923) is a well documented event, especially in P. Hadfield's 'Sixty Seconds That Will Change the World: The Coming Tokyo Earthquake' [14 op. cit]. P. Hadfield draws on Japanese records to provide a detailed literary description of the event, of the domino effects

and the damage. The description highlights the complexity of a catastrophe of natural origin in an urban environment.

The major earthquake that struck the Kanto region on 1 September 1923 shortly before noon killed thousands and caused serious damage in the cities of Yokohama and Tokyo. The shock waves that lasted less than one minute destroyed two-thirds of Tokyo and four-fifths of Yokohama. Numerous fires broke out in both cities, because the event occurred when the inhabitants were beginning to warm their braziers and light their cookers to prepare meals. At the time, Tokyo, the capital of a little developed country, counted 2.5 million inhabitants. In Tokyo the braziers set light to the wooden houses, gas and hydrocarbon depots and tanks exploded, gas mains broke and the broken water mains made fire-fighting impossible. The mostly agricultural Japanese economy was badly hit. It was estimated that 9000 factories were destroyed by fire. More than 120 000 were killed either by buildings collapsing or by fires, or by crowd panic. As the fire in Tokyo could not be brought under control, to escape the advancing flames, many victims tried to cross the River Sumida that skirts central Tokyo. But when the bridge between the two banks broke, hundreds became panic-stricken and toppled into the water, where they met their deaths.

This account emphasizes both the speed of the event and the tragic consequences of the fires that broke out in Tokyo and Yokohama immediately after the earthquake. The traditional wooden buildings facilitated the outbreak of many fires [22].

B. Instantiation of the ontology

The ontology is instantiated in two stages by our method:

- specialization of generic concepts (element, event, damage) by specifying what types of elements, events and damage are spoken of in this story;
- development of an object diagram representing the account.

1) Specialization of the conceptual model

The first stage of the instantiation consisted in analysing the account of the 'Great Kanto Earthquake' to extract the terms that specialize the 'Element', 'Event' and 'Damage' classes (Figs 4 and 5). This analysis uses the generalization/specialization relation of concepts. For example, geological faulting is one sort of event; a bridge is one sort of infrastructure.

For the needs of the description, the concepts of 'City' and of 'District' will specialize the concept of systems (we consider them, then, as particular systems in that they are themselves made up of elements). This structuring/representation of information will enable us to understand the interleaving of the spatial levels and so the complexity of the spatial scale of a phenomenon: description at the scale of a city, a district or a set of districts making up a city (Fig. 4).

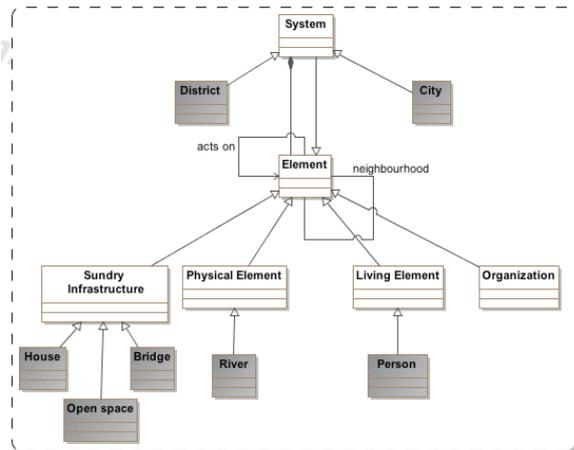


Fig. 4. Specialization of the system structure

The elements on which events bear are the 'House', 'Open Space' and 'Bridge' elements for the 'Infrastructure' class; the 'River' class for the class named 'Physical Element'; the 'Person' class for 'Living Element' (Fig. 4). The 'Faulting',

'Earthquake', 'Lighting Brazier', 'Fire', 'Collapse House', 'Assembly', 'Flight' and 'Crushing' classes specialize the 'Event' (Fig. 5).

The 'Human Damage' class is specialized into the classes 'Injury', 'Death', 'Injured' and 'Dead', while the 'Material Damage' class generalizes the 'Destruction House', 'Destruction Houses' and 'Destruction Bridge' classes (Fig. 5).

At this level of analysis, we can already see the complexity of the system explicitly appearing because of the variety and the number of components, the presence of individual and collective structures: the generic concept 'Human Damage' may concern one person (death, injury) or a population (dead, injured); the 'Material Damage' class may generalize destruction of a specific house or bridge or of several houses on the scale of a district or a city. Implicitly, complexity is also engendered by the interleaving of various levels of organization [Pavé, 1994]. The 'Assembly' class, for example, is the result of individuals clustering. We shall see later that such interleaving of levels is reflected by the emergence of new properties.

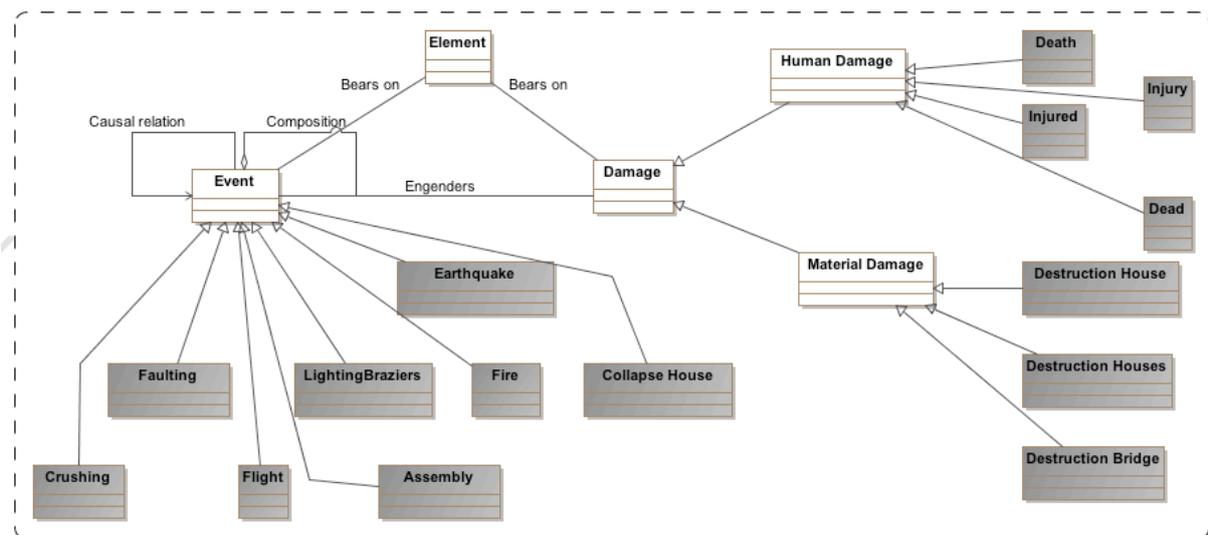


Fig. 5. Specialization of the system dynamics

It will be noticed that the complexity arising from the interleaving of spatial levels appears clearly in the system structure but is less directly apparent for events. This observation can be explained by:

- the actual composition of the ontology which means that events relate to structure;
- deliberately ignoring interactions governing each type of event. Such interactions usually give rise to new properties, especially in the field of catastrophes [20 op. cit]. Each type of event identified is therefore the outcome of a process of interaction. To take an example, friction between tectonic plates engenders stress that builds slowly and eventually

causes a sudden release which, by domino effect and force transfer, may create another until a chain reaction is produced that is the origin of an earthquake [16 op. cit.].

2) The construction of the concrete model

To speak of the particular event of the 'Great Kanto Earthquake' we instantiate the abstract concepts identified generally in the conceptual model and more precisely in the specification of the model.

a) Representation of the spatial and temporal context

From P. Hadfield's account we extract information about the identification and situation of elements of the system in

space and their spatial arrangement. This information allows us to identify the general spatial context on which the events will bear. The instantiation of the ontology on the basis of the textual data allows us to consider a particular system composed of two cities, Yokohama and Tokyo. A neighbourhood relation (from the 'Element' class relation) is established between these two cities because Yokohama is located 'a few kilometres south-west of Tokyo' (Fig. 6).

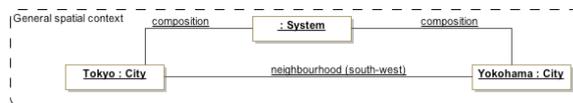


Fig. 6. The spatial context

These two cities experience a same event named the Great Kanto Earthquake. This event may, as need be, be considered as a point in time (1 September 1923) or as an episode, that is, in terms of its unfolding, its dynamics over time: from 11.58 am until sunset (Fig. 7).

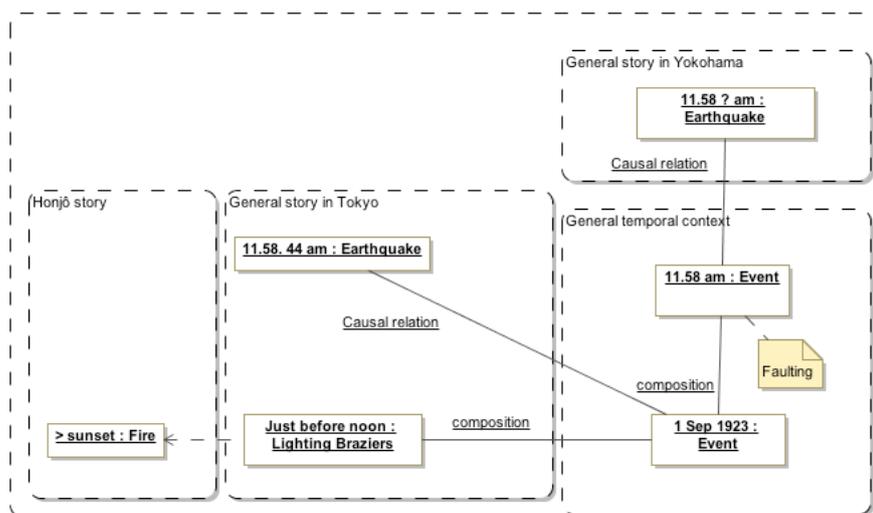


Fig. 7. The temporal context

b) Representation of the event at different scalar levels

Instantiation of the ontology allows us to account for the scales of analysis of the event and the variability in the fineness of the level of detail of the event depending on the selected scale level. Two examples are covered to further validate the ontology: the transition from the macro scale to the micro scale (analysis of the event on the scale of the city of Yokohama, then on the scale of one person), and the transition from the macro scale to the meso scale (analysis of the event on the scale of the city of Tokyo and then of its districts). The multiple scales of these two examples clearly show that the level of detail of the unfolding of the event varies with the spatial scale level selected for recounting the event. Instantiation of the account thus shows several changes of scale all providing clarifications about the unfolding of the catastrophe.

The visual representation (of the graphic modelling type) of this instantiation also helps to identify the multiple scales.

- (1) From the macro scale (city) to the micro scale (person): the transition from an overview to a close-up view of the event

Instantiation of the account allows us to distinguish the

general history of the earthquake that unfolds on the scale of the city of Yokohama (Fig. 8) and a particular story of one building and one person (Fig. 9). We thus have an overview and a close-up view of the unfolding of the event, enabling us to test one of the objectives of the ontology, the multi-scale analysis.

The event of 11.58 am, that we characterized as faulting, causes an earthquake at 11.58 am and that bears on the city of Yokohama. On the scale of the city of Yokohama, we learn that the earthquake is the cause of a domino effect, a fire. These two events engender material damage such as the destruction of housing, that has a bearing on the city (Fig. 8).

But in his account, Hadfield also describes the story of the earthquake on an individual scale. That is a particular story in Yokohama that retranscribes the event for a building (a restaurant) and a clearly identified person ('a pretty waitress') (Fig. 9).

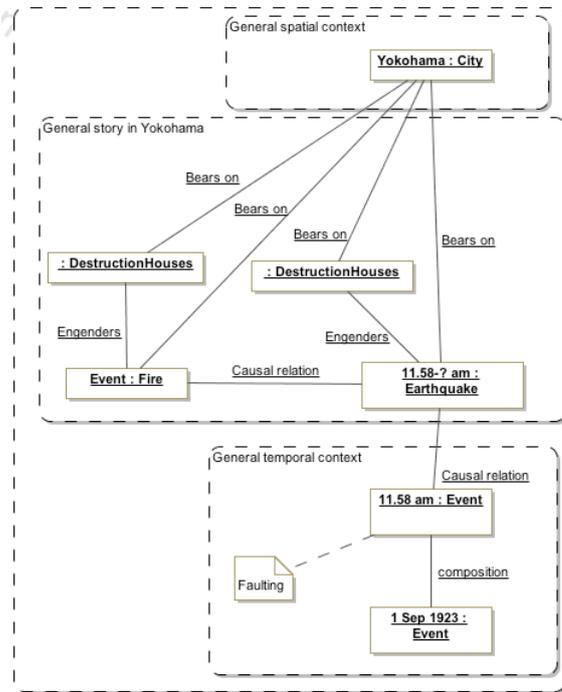


Fig. 8. Instance of the factual ontology: a general story in Yokohama

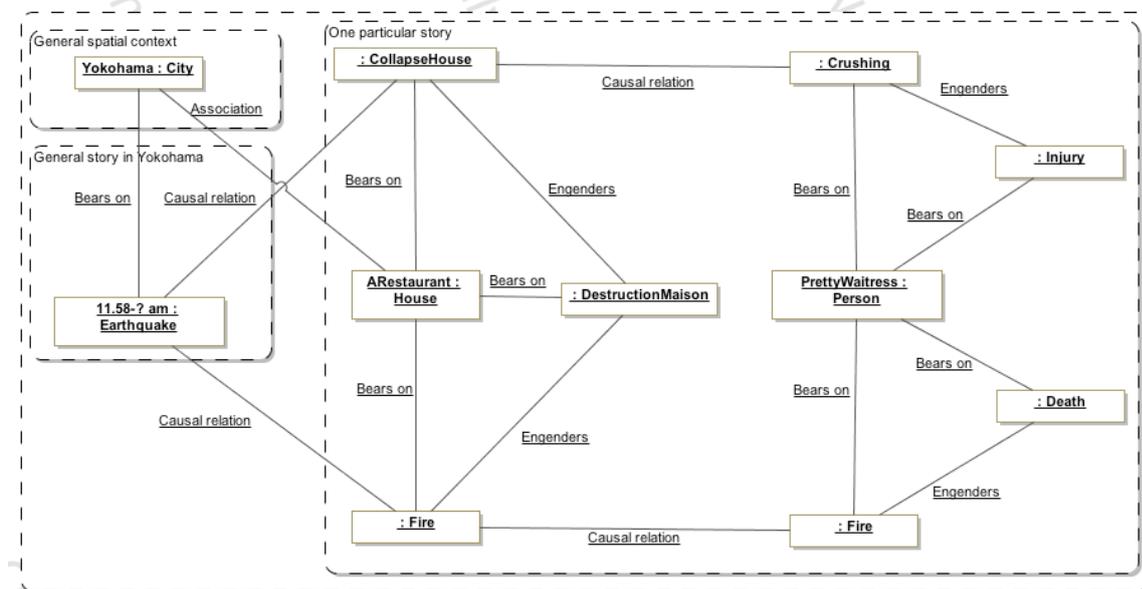


Fig. 9. Instance of the factual ontology: an individual story in Yokohama

The earthquake triggers two causal chains of events on the micro scale, a fire and the collapse of a house that bear on a restaurant and that engender damage of the ‘House Destruction’ class. That damage bears on a well identified establishment: a restaurant. The instantiation presented show therefore that we have a set of events that bears on individual

entities (a restaurant, a young girl) that engender damage that becomes more serious with time. The young girl is first injured by being crushed in the collapse of a restaurant, and then killed by a fire that breaks out inside the restaurant. This instance shows that the fatal injury results from the combination of several events.

This first instantiation allows us to validate the ontology on a multi-scale analysis, in this case the analysis at the macro scale and the micro scale. The example covered below also confirms the validation of the ontology for the transition from the macro to the meso scale and reveals the complexity of the system due to the interleaving of spatial levels.

- (2) The interleaving of spatial levels: analysis of the unfolding of the event on the scale of the city of Tokyo and its districts

Here the event is analysed at two levels: the city of Tokyo

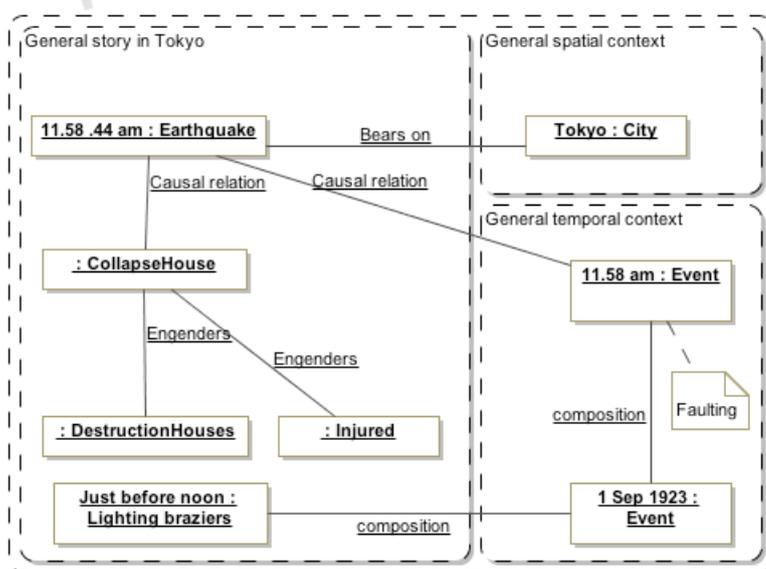


Fig. 10. The earthquake of 11.58 am and 44 seconds on the scale of the city of Tokyo

The story of the centre (Fig. 11) requires different types of event to be differentiated: the fire, the flight and the assembly. One bears on the open space of the Imperial Palace, the other on the open space of the River Sumida. The gathering of a crowd outside the Imperial Palace did not cause any human damage. However, the assembly that bears on the River Sumida was the scene of many human losses connected with a series of relations of contextual causality and panic behaviour. Those behaviours were part of a chain of causal relations between events (fire, flight, assembly) and a chain of perceptions (visual, auditory, etc.) of the situation. Those behaviours could not be transcribed in the factual ontology because the ontology in its current state does not provide the conceptual tools for dealing with perception.

(Fig. 10) and its districts (Figs 11 and 12).

The event (now characterized as the Great Tokyo Earthquake of 1 September 1923) is composed of two events that took place in one case at 11.58 am (the faulting) and just before noon (the lighting of the braziers). On the scale of the city of Tokyo (Fig. 10), two events are instantiated: the earthquake at 11.58 and 44 seconds and the collapse of houses. These events are part of a chain of events arising from the general temporal context. The 'Collapse House' event engenders two types of damage: material damage (destruction of houses) and human damage (the injured).

Instantiation of the story of Honjō (Fig. 12) highlights a contextual system comprising three districts—Nihonbashi, Asakusa, Honjō—and of the River Sumida which, because of their neighbourhood relation, triggered an event chain of fires in space and time (at 2.00 pm, 4.00 pm) entailing behaviours of flight and of assembly in an open space: a former army depot that had become wasteland. This space was not spared by the advancing fires. After sunset, a new fire broke out, bearing on the open space and engendering many human losses because of a factual causal relation between two events: the assembly brought about by the earlier fires and this new fire. There was therefore a series of events which, because of their diffusion in space and time, became multipolar.

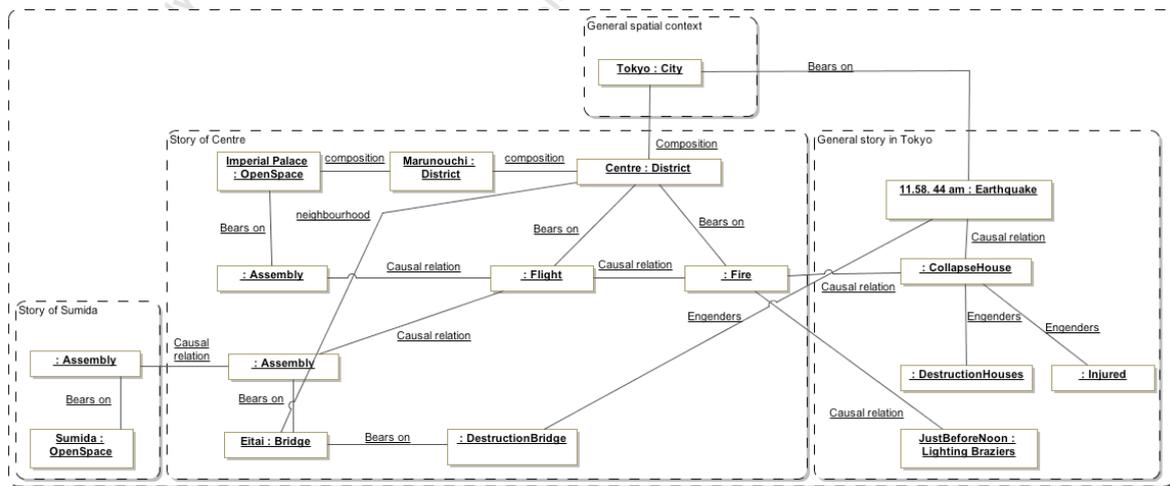


Fig. 11. Instantiation of the story of the Centre in Tokyo

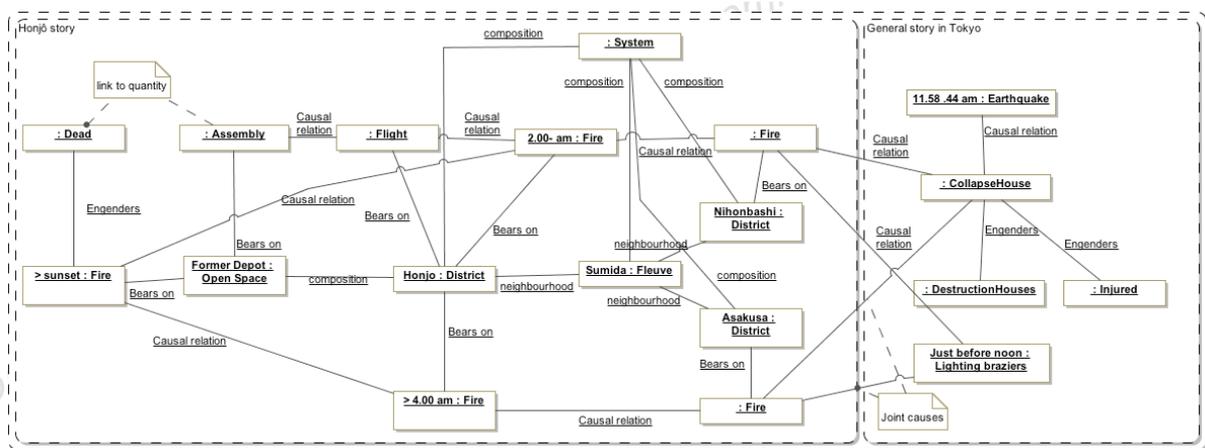


Fig. 12. Instantiation of the story of Honjō in Tokyo

Instantiation of this story shows that it is relatively easy to give an account of a catastrophe type of event using the factual ontology ‘filter’ proposed by [12 op. cit.]. Instantiation of the Great Kanto Earthquake has also showed the need for a multi-scale approach to reconstruct the account as well as possible because each scale provides specific information about the event. This vision of an event on different scales has many advantages in terms of comparison: comparison of accounts pertaining to different catastrophes, but also comparison of events of the same scale localized in different places and part of the same catastrophe.

IV. DISCUSSION

A first important element is that UML diagrams make it possible to visualize not just the articulation between concepts but also the structure of events and of concrete systems. For example, in the previous figure, the cascade of events, here the propagation of the fire, by making it clear which districts it

bears on, brings out a topology within the city of Tokyo that is never described as such in the account. The graphical representation brings it out quickly. In the same way, concurrent event chains appear that jointly contribute to a new event (e.g. the fire in the former depot) and the structural and event hierarchy that reciprocally make visible the implicit structures.

Whether the example of the Honjō district or that of the waitress in Yokohama (on very different scales), by following the ties between instances, the representation of the account brings out structurally, the domino effects that are often mentioned in the literature. More than domino effects, it is multiple effects by separate causal chains that allow us to account for events (and therefore for vulnerability) that would not otherwise be identifiable. The catastrophe therefore became multipolar and evolved in space and time.

The case of Honjō is especially interesting, for all of the districts involved, on either side of the river, only formed a system on the occasion of the events that occurred there. Three

points can be learned from this:

- the existence of particularly substantial vulnerability of a district that does not seem to have been identified as such;
- the relative vulnerability of the centre and of Honjō tied to the dynamics of the process that makes what are *a priori* similar elements -the imperial palace and wasteland on the site of a former depot- play very different roles since the population can take refuge there in one case and dies there in the other;
- the fact that a system does not necessarily exist *a priori* but is identified depending on the events that occur there. This highlights the interest of making simulations under various scenarios so as to identify the systems that emerge and their associated vulnerabilities. A system and its vulnerability do not exist of themselves but only as a function of the events that may occur there.

As concerns the validation approach itself, there are two points of interest. First, it has enabled us to complete the conceptual ontology, not just to add more specific concepts such as the idea of city or district, but also linkages we had not thought of *a priori* such as the neighbourhood relation and the hierarchical breakdown of systems. Conceptual ontology virtually acquires the status of theory, that can be revised after the experimentation constituted by the instantiation on a concrete case. It has also enabled us to show that the whole of a complex account through the diversity of the events described and its multi-scalar aspect could be represented (not everything has been presented here for reasons of space).

Admittedly, that does not prove that any and all accounts can be modelled with the concepts proposed, nor does it prove the relevance of these concepts to the other part of the conceptual ontology that is specific to the characterization of facts such as catastrophes by various criteria. There too, a mixture of tests of internal coherence and instantiation of accounts and multiple testimonies is required.

V. CONCLUSION AND PERSPECTIVES

Although it must still be subjected to other instantiations, whether in terms of structure and of system dynamics or in terms of the representation borne by actors on the event, the example proposed has allowed us to emphasize that factual ontology may provide a framework for formalizing knowledge and so facilitating comparisons. It has also shown the essential character of a multi-scale approach when accounting for an event. Lastly, it seems to us that instantiation provides insight into the timing of events on a given scale (it is not always specified in the account) and their spatiality.

At this stage, we cannot yet validate the ontology of risk and catastrophe proposed by [12 op. cit.] but this first instantiation has shed some interesting light on things. Other examples should be covered to confirm or possibly fine-tune the proposed ontology.

Nonetheless, although the instantiation teaches us a great deal, we are aware of the limits of the ontology in its current shape. Thus, the panic phenomenon referred to in the text titled 'The Great Kanto Earthquake', could not be treated in

the context of this ontology. Panic is part of a chain of causal relations between events (fire, flight, assembly) and a chain of perception (visual, auditory, etc.) of the situation. Panic is therefore dependent on a series of events but also on the perception people have of the event. The model proposed does not as it stands provide the conceptual tools for dealing with perception.

We need, therefore, to go beyond this limit because panic behaviour is not an epi-phenomenon. Panic is unusual in more ways than one: it is neither localized, nor confined to a particular environment, nor the result of any one specific event.

REFERENCES

- [1] A. Dauphiné, *Risques et catastrophes*. Paris: Armand Colin, 2003.
- [2] R. D'Ercole et al., "Les vulnérabilités des sociétés et des espaces urbanisés: concepts, typologie, modes d'analyse," *Revue de géographie alpine*, tome LXXXII, n°4, pp. 87-96, 1994.
- [3] C. Gilbert, "La vulnérabilité, une notion à explorer", *Pour la Science*, n° 51, pp. 116-120, 2007.
- [4] G.-Y. Kervern. *Éléments fondamentaux des Cindyniques*. Paris: Economica, 1995.
- [5] F. Leone, *Caractérisation des vulnérabilités aux catastrophes « naturelles » : contribution à une évaluation géographique multirisque*. HDR, Université Paul Valéry – Montpellier III, 2008.
- [6] V. November, "Le risque comme objet géographique", *Cahiers de géographie du Québec*, 50(141) pp. 289-296, 2006.
- [7] P. Pigeon, *Géographie critique des risques*. Paris: Economica, 2005.
- [8] D. Provitolo, "La vulnérabilité aux inondations méditerranéennes: une nouvelle démarche géographique", *Annales de Géographie*. pp. 23-40, 2007
- [9] J.I. Uitto, "The geography of disaster vulnerability in megacities: a theoretical framework," *Applied Geography*, Tome 18, n°1, pp. 7-16. 1998.
- [10] Y. Veyret, *Les risques*, Paris: Bréal, 2004.
- [11] B. Wisner, "There are worse things than earthquakes: hazard, vulnerability and mitigation capacity in greater Los Angeles," in Mitchell J.-K., *Crucibles of Hazard: Disasters and Megacities in Transition*, Tokyo, New York, Paris: United Nations University Press, 1999.
- [12] D. Provitolo, J.P. Müller and E. Dubos-Paillard, "Vers une ontologie des risques et des catastrophes: le modèle conceptuel," in Actes du colloque, *Ontologie et dynamiques des systèmes complexes*, XVI èmes rencontres de Rochebrune, janv 2009, Available <http://gemas.msh-paris.fr/dphan/rochebrune09/papiers/ProvitoloDamien ne.pdf>.
- [13] T.R. Grüber, "Toward principles for the design of ontologies used for knowledge sharing", in N. Guarino and R. Poli (Eds.), *International Workshop on Formal Ontology*, Padova, Italy, (1993). In *International Journal of Human-Computer Studies*, Volume 43(5-6), pp 907-928, 2005.
- [14] P. Hadfield, *Sixty Seconds That Will Change the World: the Coming Tokyo Earthquake*. London: Sidgwick & Jackson, 1991.
- [15] S. Cranefield and M. Purvis, "UML as an ontology modelling language," in Proc. of the Workshop on Intelligent Information Integration, 16th International Joint Conference on Artificial Intelligence (IJCAI-99), 1999, Available: <http://www.aifb.uni-karlsruhe.de/WBS/dfe/iii99/crainfield-ijcai99-iii.pdf>
- [16] P. Bak, *How Nature Works*. New York: Springer-Verlag, 1996.
- [17] P. Blaikie, T. Cannon, I. Davis and B. Wisner, *At risk: natural hazards, people's vulnerability, and disasters*. London: Routledge, 1994.
- [18] C. Chaline and J. Dubois Maury, *Les risques urbains*. Paris: Armand Colin, 2004.
- [19] D. Provitolo, "Un exemple d'effets de dominos : la panique dans les catastrophes urbaines", *Cybergéo*, n° 328. 2005. Available: <http://www.cybergeu.eu/index2998.html>
- [20] D. Provitolo "A proposal for a classification of the catastrophe systems based on complexity criteria," in *From System Complexity to Emergent Properties*, Springer, Series understanding complex systems, in press.

- [21] E. Daudé, D. Provitolo, E. Dubos-Paillard, D. Gaillard, E. Eliot, P. Langlois, E. Propeck-Zimmermann and Th. Saint-Gerand, “Spatial risks and complex systems: methodological perspectives” in *From System Complexity to Emergent Properties*, Springer, Series understanding complex systems, in press.
- [22] B. De Vanssay, “Du séisme de Kanto au séisme de Kobé. Utilité et limites de la prévision des catastrophes”, *Futuribles analyse et prospective*, pp. 25-43, 1997.
- [23] A. Pavé A., *Modélisation en biologie et en écologie*. Lyon: Aléas, 1994.

Exploitation of a displacement survey to detect road network use vulnerability

Michel Nabaa, Cyrille Bertelle, Antoine Dutot, Damien Olivier

Université du Havre

LITIS

25, rue Philippe Lebon

F-76085 Le Havre

{Michel.Nabaa; Cyrille.Bertelle; Antoine.Dutot; Damien.Olivier;
}@univ-lehavre.fr

Patrick Lions

CODAH

19, Rue Georges Braque

F-76085 Le Havre

Patrick.Lions@agglo-havraise.fr

Abstract—Many agglomeration are faced to multiple technological and natural hazards. The use of the road network by vehicles may cause big problems especially under evacuation situations. An accident on a main road may endanger many lives. A household survey was conducted in CODAH¹ about people displacement in the seine estuary. In this paper, we extracted and treated the survey result. We also compared results with a detection of organizations in large graphs algorithm.

Index Terms—vulnerability, risk, self organization, detection of organizations, GIS, household displacement survey.

I. INTRODUCTION

AS many agglomerations, CODAH is faced to natural and technological hazards. 33 establishments are classified SEVESO² with high threshold. To minimize risk effects, we must localize the population at any time. In this context, the major risk management team (DIRM) of CODAH has developed a model to estimate the diurnal / nocturnal population distribution at buildings scale (PRET RESSE) [1]. This model does not take into account the population displacement on the road network. In previous works, we proposed an approach to dynamically assess the vulnerability related to the road network use by vehicles [2], [3]. To complete this work, we have exploited a household survey about the displacement of the population of CODAH in the Seine estuary. In this paper, we exploit the survey and we compute vehicles path declared displacements to dispatch the population on the network. We present also a detection of organizations on large graphs algorithm. The aim of this algorithm is to detect zones

¹Communauté De l'Agglomération Havraise: it groups 17 communes including Le Havre. The goal is to develop many common projects in territory management, health and hygiene, public transport, risk management...

²Directive SEVESO is an European directive, it lays down to the states to identify potential dangerous site. It intends to prevent major accidents involving dangerous substances and limit their consequences for man and the environment, with a view to ensuring high levels of protection throughout the Community.

susceptible to be congested, this by taking into account the topological aspect of the graph. Finally, the results of the survey will be compared to those obtained from the algorithm.

II. RISK MANAGEMENT

Real exercises in an evacuation planning are so expensive (people, resources, logistics...) and sometimes not realistic : they can not take into account each individual behavior as the panic effect and initial response to an evacuation in a dangerous area. Panic generally results from the lack of coordination and dialog between individuals. Hence the use of the simulation and in particular Multi-Agent Systems (MAS) is necessary to model each agent behavior and more understand the evolution of a critical situation due to interactions between evacuees and the propagation of a danger : an accident on the road network may cause dangerous traffic jams especially in a case of a danger that spreads quickly. Once we understand the evolution of a critical situation, we try to fight against organizations of vehicles and try to allocate vehicles on different roads to avoid bottlenecks. Recently, many researchers tried to couple a MAS with a Geographical Information System (GIS), especially in risk management and vulnerability assessment. For computer modelers, this integration provides the ability to have agents linked to real geographical locations. For GIS users, it provides the ability to model the emergence of phenomena by various interactions of agents in time and space by using a GIS [4]. Thanks to a GIS, one can combine different vulnerability layers (a layer per danger : geophysical one, social one, access to the network...) to construct a global layer which leads to a better assessment of vulnerability; we think particularly about the work of [5]. In geography, the representation of a MAS coexists n levels of organizations and use several classes of agents (e.g. Level 1: individuals or companies, Level 2 and three: economic, urban communities). There will be rules at

every level and the approach is not necessarily a bottom up one as in the models of self-organization [6]. In [7], the author modeled the spread of the panic from a group of individuals in danger situation to non panicked ones. She used a dynamic system (differential equations) to simulate the behavior of individuals. Many execution scenarios showed that the emergence of panic has not occurred in all scenarios when changing some parameters. The emergence depends on the rate of transmission from a population susceptible to panic to a panicked population, the time taken to return to a normal behavior (the population is more panicked after the disaster) and the number of initially panicked people. The principle of resilience was also discussed; it represents the time that the system makes to return to its initial state after a period of instability due to a disaster.

At neighborhood scale, many researches were developed models based on microsimulation. In their paper [8], the authors have presented a model to estimate the time required to evacuate a neighborhood according to the population, the number of vehicles and roads network capacity. The model is based on optimization in order to find the dangerous area around a critical point. This model was coupled with a GIS (ArcInfo) to visualize the results (identifying evacuation plans) and to establish an evacuation map for the town (Santa Barbara). The same authors [9] opened the way to the study based on geographic information systems to evacuate people. Their study identified communities that may face transportation difficulties during an evacuation. They modeled the city by a graph to estimate lanes (Lane) occupation by vehicles during an evacuation. A graph partitioning model was adopted in order to detect vulnerable neighborhoods around each node of the graph and to build a vulnerability map around the nodes of the graph. A constructive heuristic was used to calculate the best cluster around each node. The result was displayed on a map with ARCINFO GIS. An area is increasingly vulnerable according to the number of evacuees per lane (number of people to evacuate in a neighborhood / number of exit roads). The author believes that the combining of an evacuation based vulnerability layer and a hazard one allows to build a general risk map, so we can explore the various issues related to risk. However, in this approach, we predefine the maximum number of nodes in a neighborhood, which may not always be realistic and does not take into account the traffic evolution during evacuation. Secondly, the city of Santa Barbara is not a typical example of urban metropolitan areas in the United States with an important population facing great difficulties during an evacuation.

III. DISPLACEMENT SURVEY

Lifestyles and travel practices are changing. Traveled distances have increased. New infrastructures have been established and new services are proposed: these developments must meet the expectations and new patterns of migration of our territory inhabitants. That is why a household displacement survey was conducted, with a standard model developed by the Center for Studies on networks, transport, urban planning and public constructions (the CERTU). The survey will pro-

vide shared data on current displacement practices and their evolution.

A. Survey utility

The last displacement survey occurred in 1992 at Le Havre agglomeration. Over time, taken paths become more complex and new infrastructures exist; hence the need to a displacement survey to respond to the population expectations, this by :

- considering the travel conditions between the city and its catchment area of life: people come from further far away to work, study or consume.
- Interview people about their expectations for transport.
- defining the transport policy of the CODAH for the coming years: how to increase the use of public transport, where and how to adapt the supply, how to relate the supply of transport lanes with the rest of the territory, what are the new flows, new trends, future developments areas of housing and economy ...
- promote environmental issues, defined in the law of Solidarity and Urban Renewal. Communities undertake increasingly the development of a sustainable mobility.

In this work, the main utility of the survey is to localize the population on the network at any time of the day. This will help us to detect the congested areas and to estimate the vulnerability related to the road network use by vehicles.

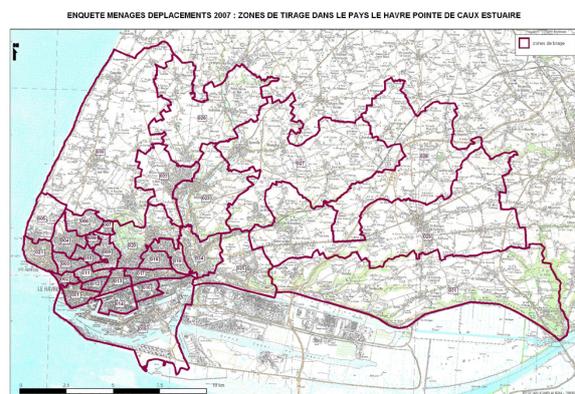


Fig. 1: zoning the Seine estuary communes

The territory of the survey covers the communities of the Seine estuary (CODAH, Saint Romain de Colbosc community) and Tancarville community or a territory of 34 communes. This territory was divided into 102 zones to meet the survey needs. Each surveyed person moves from a zone to the same or another zone. These zones are illustrated in figure 1.

B. Survey details

5194 persons representing 2224 households were surveyed from different zones of CODAH. After a generalization of the survey, we have around 320000 motorized displacements per day. The adopted model is illustrated in the following UML schema (figure 2).

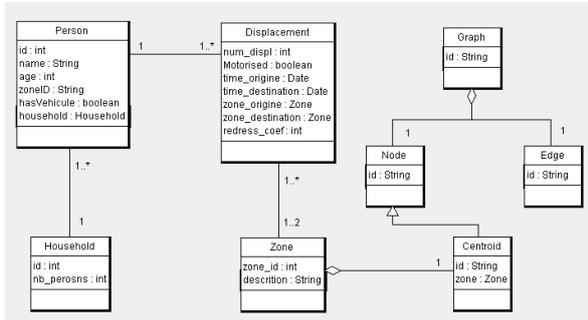


Fig. 2: UML schema

Each zone has at least a centroid which is a node strongly connected to other nodes. So each person declaring a displacement from a zone $Z1$ to a zone $Z2$ must move from a centroid of $Z1$ to a centroid of $Z2$. The *redress_coef* attribute in the *Household* class represents the redress coefficient of the household. So, if a person p belonging to a household h having the redress coefficient c moves from a zone a at time $t1$ to a zone b at time $t2$, we must generalize this by moving c persons from a to b at the same time. Most of people round up their time of departure and arrival. One declares leaving home at 08h00 while he did it at 07h56. This has caused a problem of routes saturation because of moving many vehicles from the same origin to the same destination, at the same time. To face this problem, we used a normal distribution around departure time while respecting the duration of a displacement declared by each person.

C. Environment modeling

The road network is integrated as a layer in the Geographic Information System (GIS). From this layer, we extract the data by using the open source java GIS toolkit Geotools. This toolkit provides several methods to manipulate Geo spatial data and implements Open Geospatial Consortium (OGC) specifications, so we can read and write to ESRI shapefile format. Once data road network are extracted, we use the GraphStream tool [10] developed within LITIS laboratory of Le Havre to construct a graph corresponding to the GIS network layer. This tool is designed for modeling; processing and visualizing graphs.

The data extracted from network layer contains the roads circulation direction, roads id, roads type, their lengths and geometry.

The extracted multigraph $G = (V, E)$ represents the road network where V is the set of nodes and E the set of arcs. We deal with a multigraph because we have sometimes more than one oriented arc in the same direction between two adjacent nodes due to multiple routes between two points in the Seine Estuary road network. GraphStream facilitates this task because it is adapted to model and visualize multigraphs. In the constructed multigraph:

- The nodes represent roads intersections,
- The arcs represent the roads taken by vehicles,

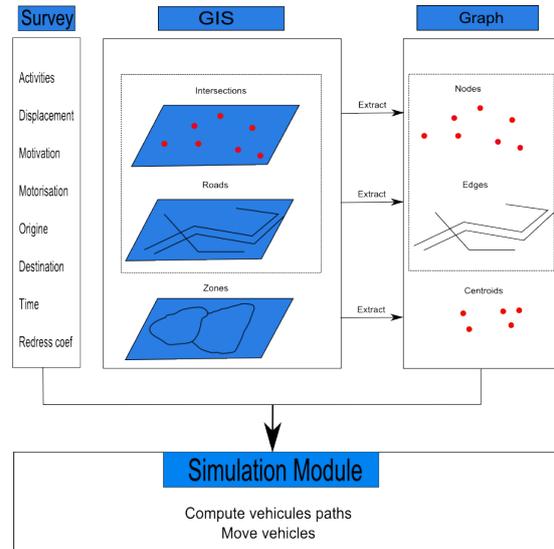


Fig. 3: Extract the graph and compute vehicles path

- The weight on each arc represents the needed time to cross this arc. Each arc has a maximum capacity of vehicles ($Arc\ length / Vehicle\ length$)

D. Compute vehicles paths

In our work, we must repatriate people having their personal vehicle on the road network. The main goal is to detect routes with high vehicles density. It helps us in the detection of organizations and estimation of the vulnerability related to the road network using. Generally, people tend to choose the shortest path in time when using their own vehicles. This assumption, which may seem audacious and even unrealistic, however, is found in almost all mobility models implemented in most countries since 1960 [11]. For each person displacement from a centroid $c1$ to a centroid $c2$, we compute the shortest path between the two nodes by using A-Star algorithm. It has the advantage to be fastest than Dijkstra algorithm when using an efficient heuristic. The last is an exact algorithm which compute the shortest path between a source node to all other nodes in a graph. In future work, we will use the dynamic version of A-Star (D-star) to avoid congested routes when computing paths. Another problem occurred when defining just one centroid per zone; obviously the must connected one. In figure 4, taken routes by vehicles are colored in red. At the left of the figure, we adopt one centroid per zone : we have 4772 taken routes and those routes are quickly congested. At the right, we used 15 centroids per zone to overcome this limitation : 17684 routes was taken. It is the maximum centroids number per zone if we use a machine having 2GO of ram. We can conclude that the more routes are taken and the less maximum roads capacity is violated.

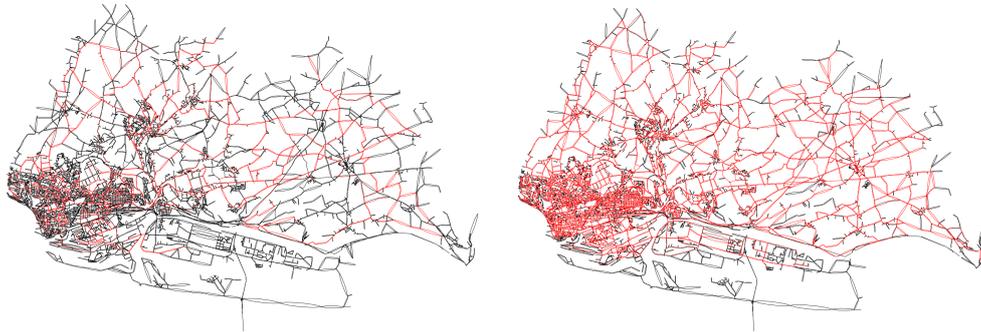


Fig. 4: Taken routes with one centroid and 15 centroids per zone

IV. LOOKING AT OFTEN USED PATHS IN THE CITY ROAD NETWORK

We conducted a simple experiment using random walks inside the city road network, seen as a graph. In this experiment we take a given number of entities placed randomly on nodes of the city graph. Then we run an iterative loop where, at each step, these entities either traverse a randomly chosen edge or wait.

Each time an entity crosses an edge $e \in E$, a special variable e_{np} (number of passes) bound to the edge is incremented. At the end of such a simulation, the number of passes on each edge by entities will therefore be stored in these variables.

The entities may either cross an edge or wait because the displacements must consider road traversing times. The time an entity waits is proportional to the average time used to travel on the corresponding road: for two edges of the same length, an entity will wait longer if the road is a city road than if it is an highway.

After an entity crossed an edge, it chooses the next edge to cross randomly, but do not consider any more the edge it came from. The edges the entity used are stored in a memory, and the entity tries to avoid reusing them. This memory is a FIFO stack and ultimately, roads in the memory are forgotten, so that the entity can reuse them later.

As the road graph we use is bounded, some roads on the border are cut. If an entity reaches such a dead end (or a real bag end), it "jumps" in another position in the graph, chosen randomly.

In the figure 5 we show one run of this experiment. We in fact did several runs with a varying random seed, and always obtained very similar results. Edges are colored using a color graduation going from blue (almost no pass) to red (a lot of passes) passing by green, yellow and orange. The scale is geometric. To pass from blue to green you have to double the number of passes, and to pass from green to yellow you anew have to double this number of passes.

There are as many entities as nodes in the network. After some setup, we set the entity memory to 40 nodes. We stop the algorithm when a given maximum number of passes is reached on one of the edges, here 4000. We chosen such a high number to ensure the exploration of entities is significant enough.



Fig. 5: Random walks in Le Havre, blueish to greenish roads are less used and yellowish to reddish roads are the more used.

For a better understanding, the same results are shown on figure 6 with labels that help to locate roads. The road that are highly used are labelled and highlighted. They correspond to town centers and highways. The A29 and A13 are the two main highways passing by and going to Le Havre. The road labelled "main city entry" is one of the most used road when coming to Le Havre (due to the fact the city is a harbour and have its south and west sides bordered by water, and at the north a plateau ("ville haute") with smaller accesses. The "Breque" label indicates a very large interchange that is, for the same reasons, one of the most used section of road when coming to and leaving Le Havre.

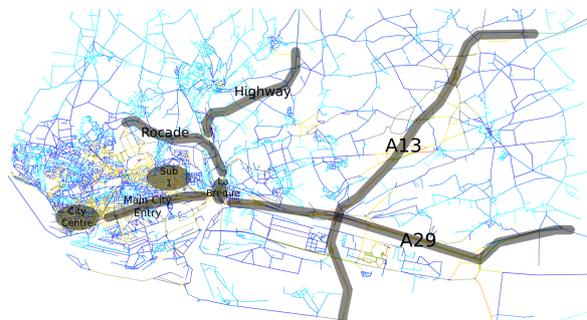


Fig. 6: Figure 5 with labels on important areas.

The number of passes found with this random walk is

an indication of the roads in the network where vehicles are inclined to pass, due to the topology of the road graph. This can be an indicator on areas where vehicle will most probably pass often.

We also run this simulation anew, but with some roads removed. Namely, we removed the A29 and A13 highways, as well as the "main city entry". The figure 7 shows the result. These road "absorbed" a large part of the traffic, and therefore all the entities use a lot more other roads. The "Breque" interchange is completely saturated. This place is well known by Le Havre inhabitants for the traffic jams occurring in the morning and the evening when people go to or leave their work. "La Breque" is one of the mandatory road to use in order to leave or enter the city.

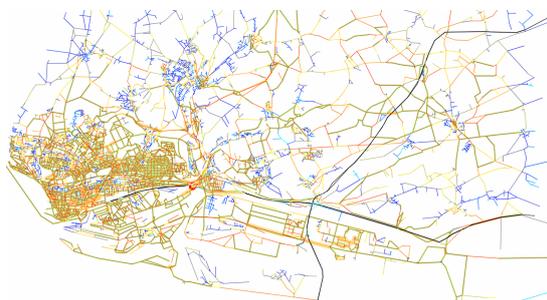


Fig. 7: Random walks in Le Havre, with the A29 and A13 highways removed as well as the main city entry.

If we run the same experiment on figure 8, but only disabling the "main city entry", the results are more similar to the one without any road removed. This may mean that this road could be removed and people redirected to adjacent roads without impacting largely the traffic.



Fig. 8: Random walks in Le Havre, with only the main city entry removed.

V. CONCLUSION

We have extracted a household survey on the motorized population displacement in the Seine Estuary. For each displacement from a zone to another one, we computed a shortest path from a centroid of the first zone to a centroid to the other zone. We used an A-Star algorithm to do that. Results showed that the more we have centroids per zone, the more we explore more routes and the less we violate routes capacity. We also

applied on the Seine estuary graph, a detection of organizations algorithm based on random walks. The last showed that we can predict the possibility to have traffic jams in some areas on the basis of the network structure, this without having any data on the traffic. In the near future, we will adopt a microsimulation of the traffic with the use of D-Star algorithm to dynamically compute paths and to respect the road capacities at any time.

REFERENCES

- [1] J. Bourcier and P. Mallet, "Allocation spatio-temporelle de la population exposée aux risques majeurs. contribution à l'expologie sur le bassin de risques majeurs de l'estuaire de la seine: modèle pret-resse," *Revue internationale de Géomatique*, vol. 16(10), pp. 457–478, 2006.
- [2] C. Bertelle, A. Dutot, M. Nabaa, D. Olivier, and P. Mallet, "Modeling of the vulnerability related to the dynamic road traffic," in *ESM*, Le Havre, 2008.
- [3] C. Bertelle, A. Dutot, M. Nabaa, and D. Olivier, "Detection of organizations in large graphs : Application on risk management," in *Agent Based Spatial Simulation*, 2008.
- [4] R. Najlis and M. J. North, "Repast for gis." University of Chicago and Argonne National Laboratory, IL, USA: In Proceedings of Agent 2004 : Social Dynamics: Interaction, Reflexivity and Emergence, 2004.
- [5] S. Cutter, J. Mitchell, and M. Scott, "Revealing the vulnerability of people and places: a case study of georgetown county, south carolina," *Annals of the Association of American Geographers*, vol. 90(4), pp. 713–737, 2000.
- [6] E. Daudé, *Systèmes multi-agents pour la simulation en géographie : vers une Géographie Artificielle*. in Y. Guermont (dir.), *Modélisation en Géographie : déterminismes et complexités*, 2005, ch. 13, pp. 355–382.
- [7] D. Provitolo, "A proposition for a classification of the catastrophe systems based on complexity criteria," in *European Conference Complex Systems-EPNACS'07, Emergent Properties in Natural and Artificial Complex Systems, Dresden*, 2007, pp. 93–106.
- [8] R. Church and T. Cova, "Mapping evacuation risk on transposition networks using a spatial optimization model," *Transportation Research Part C: Emerging Technologies*, vol. 8(1-6), pp. 321–336, 2000.
- [9] T. Cova and R. Church, "Modelling community evacuation vulnerability using gis," *International Journal of Geographical Information Science*, vol. 11(8), pp. 763–784, 1997.
- [10] A. Dutot, F. Guinand, D. Olivier, and Y. Pigné, "Graphstream: A tool for bridging the gap between complex systems and dynamic graphs," in *EPNACS: Emergent Properties in Natural and Artificial Complex Systems*, 2007.
- [11] A. Banos, "À la conquête de la fourmière urbaine : quand le géographe se fait explorateur de mondes artificiels," *Tech. Rep.*, 2005. [Online]. Available: http://fig-st-die.education.fr/actes/actes_2005/index.htm

Evolution Problems and Complexity

Organizing Committee

- Rabah LABBAS
University of Le Havre, France
Rabah.Labbas@univ-lehavre.fr
- Salim MESSAOUDI
King Fahd University, Saudi Arabia
messaoud@kfupm.edu.sa

Description

In the complex systems, there are many concrete situations described by different evolution problems. Our session is intended to give some tools and results in some problems governed by partial differential equations. Includes

- Semi group theory
- Partial differential equations
- Asymptotic behaviour of the solutions
- Inverse problems

Contents

Oscillatory behavior of solutions for a class of second order nonlinear differential equation with perturbation Quanxin ZHANG , Lei WANG	159
On quasistatic models of contact phenomena Anna OCHAL , Stanislaw MIGORSKI	164
On an isotropic differential inclusion Ana Cristina BARROSO , Gisella CROCE , Ana Margarida RIBEIRO	169
On some transmission problems in Hölder spaces Ahmed MEDEGHRI	174
Coefficients de singularités géométriques pour des problèmes d'évolution Mohand MOUSSAOUI	174
Sturm-Liouville problems for a complete abstract second order differential equation of elliptic type in UMD spaces Stéphane MAINGOT	174
A general decay result in a viscoelastic Timoshenko system Salim MESSAOUDI	174
On some transmission problems with boundary Dirichlet conditions Fatimetou MINT AGHRABATT	175
Complete abstract differential equations of elliptic type on the half-line : application of Dore-Venni and Da Prato-Grisvard sum theory in L_p-spaces Amine ELTAIEF	175
On some abstract fourth differential equation with transmission and boundary conditions Hassan DIARAMOUNA SIDIBE	175

Oscillatory Behavior of Solutions for a Class of Second Order Nonlinear Differential Equation With Perturbation

Quanxin Zhang
Binzhou University
Department of Mathematics
Binzhou, Shandong, China
3314744@163.com

Lei Wang
Shandong University
School of Mathematics
Binzhou University
Department of Mathematics
Binzhou, Shandong, China
keli21cn@163.com

Abstract

In this paper, we discuss the oscillatory behavior of solutions for a class of second order nonlinear differential equation with perturbation and establish two theorems which develop and generalize some known results.

1 Introduction

In the past few years, the oscillation problem for the following second order nonlinear differential equation with damping

$$(a(t)\psi(x(t))x'(t))' + p(t)x'(t) + q(t)f(x(t)) = 0, \quad ' = \frac{d}{dt} \quad (E_1)$$

has been studied [1,2], and the oscillation of the following second order nonlinear differential equations

$$(a(t)\psi(x(t))x'(t))' + q(t)f(x(g(t))) = 0 \quad (E_2)$$

and

$$(a(t)\psi(x(t))x'(t))' + q(t)f(x(t)) = 0 \quad (E_3)$$

have been investigated in [3,4]. And Jurang Yan[5] has given the oscillation theorems for a second order linear differential equations with damping

$$(r(t)x'(t))' + p(t)x'(t) + q(t)x(t) = 0. \quad (E_4)$$

In this paper, we discuss the oscillatory behavior of the solutions of the second order nonlinear differential equation with perturbation of the form

$$(a(t)\psi(x(t))x'(t))' + Q(t, x(t)) = P(t, x(t), x'(t)), \quad ' = \frac{d}{dt}. \quad (1)$$

where we let

(A₁) $a : [t_0, +\infty) \rightarrow R$ ($R = (-\infty, +\infty)$) is positive continuously differentiable;

(A₂) $\psi : R \rightarrow R$ is continuously differentiable and $\psi(u) > 0$ for $u \neq 0$;

(A₃) $Q : [t_0, +\infty) \times R \rightarrow R$ is continuous, and there exists a continuous function $q(t)$ and continuously differentiable function $f(x)$ such that $\frac{Q(t, x)}{f(x)} \geq q(t)$ for $x \neq 0$,

where $q : [t_0, +\infty) \rightarrow R$, $f : R \rightarrow R$, $uf(u) > 0$ and $f'(u) > 0$ for $u \neq 0$;

(A₄) $P \in ([t_0, +\infty) \times R^2 \rightarrow R)$, and there exists $p(t) \in ([t_0, +\infty) \rightarrow R)$ such that $\frac{P(t, x(t), x'(t))}{f(x)} \leq p(t)$ for $x \neq 0$.

In the condition that

$$\lim_{t \rightarrow +\infty} \int_{t_0}^t [q(s) - p(s)] ds = +\infty \quad (2)$$

is not establishment, we discuss the oscillatory behavior of Eq.(1) and establish two theorems which develop and generalize some known results.

Throughout by a *solution* of Eq.(1) we shall mean a function which exists on $[t_0, +\infty)$ satisfies Eq.(1) and $x(t) \neq 0, t \in [T, +\infty)$. As usual, a solution of Eq.(1) is said to be oscillatory if it has arbitrarily large zeros; otherwise it is said to be *nonoscillatory*. A nonoscillatory solution of Eq.(1) is said to be *weakly oscillatory* if $x'(t)$ changes sign for arbitrarily large values of t (see [3, 6]). Eq.(1) is called *oscillatory* if all its solutions are oscillatory.

With respect to their asymptotic behavior, all the solutions of Eq.(1) can be divided into the following four types:

$S^+ = \{x = x(t)$ solution of Eq.(1): there exists $t_x \geq t_0$ such that $x(t)x'(t) \geq 0$ for $t \geq t_x\}$;

$S^- = \{x = x(t)$ solution of Eq.(1): there exists $t_x \geq t_0$ such that $x(t)x'(t) < 0$ for $t \geq t_x\}$;

$S^O = \{x = x(t)$ solution of Eq.(1): there exists $t_n, t_n \rightarrow +\infty$, such that $x(t_n) = 0\}$;

$S^{WO} = \{x = x(t)$ solution of Eq.(1): $x(t) \neq 0$ for t sufficiently large and for all $t_\alpha > t_0$ there exist $t_{\alpha_1} > t_\alpha$, $t_{\alpha_2} > t_\alpha$ such that $x'(t_{\alpha_1})x'(t_{\alpha_2}) < 0\}$.

With very simple argument we can prove that S^+, S^-, S^O, S^{WO} are mutually disjoint. By the above definitions, it turns out that solutions in the class S^+ are eventually either positive nondecreasing or negative nonincreasing, solutions in the class S^- are eventually either positive nonincreasing or negative nondecreasing, solutions in the class S^O are oscillatory, and finally, solutions in the class S^{WO} are weakly oscillatory.

2 Main Results

In this section, we establish two oscillatory theorems of Eq.(1).

Lemma 1 *If for sufficiently large T such that*

$$\liminf_{t \rightarrow +\infty} \int_T^t [q(s) - p(s)] ds \geq 0, \quad (3)$$

then $S^{WO} = \emptyset$ for Eq.(1).

Proof Suppose that Eq.(1) has a solution $x(t) \in S^{wo}$. There is no loss of generality in assuming that there exists $t_1 \geq t_0$ such that $x(t) > 0$ for all $t \geq t_1$. (For $x(t) < 0$, the proof is similar.) Thus for all $t_\alpha > t_1$ there exist $t_{\alpha_1}, t_{\alpha_2} > t_\alpha$, such that $x'(t_{\alpha_1})x'(t_{\alpha_2}) < 0$. Therefore there exists the sequence $\{C_n\} \rightarrow +\infty$ such that $x'(C_n) < 0$. Choosing sufficiently large N such that C_N satisfies the condition (3). i.e.,

$$\liminf_{t \rightarrow +\infty} \int_{C_N}^t [q(s) - p(s)] ds \geq 0.$$

Consider the function

$$W(t) = \frac{a(t)\psi(x(t))x'(t)}{f(x(t))}, \quad t \geq t_1.$$

Then it follows from Eq.(1) when $t \geq t_1$

$$\begin{aligned} W'(t) &= -\frac{Q(t, x(t))}{f(x(t))} + \frac{P(t, x(t), x'(t))}{f(x(t))} \\ &\quad - a(t)\psi(x(t))f'(x(t))\frac{x'^2(t)}{f^2(x(t))} \\ &\leq -q(t) + p(t) \\ &\quad - a(t)\psi(x(t))f'(x(t))\frac{x'^2(t)}{f^2(x(t))} \\ &\leq -q(t) + p(t), \end{aligned}$$

For all $b \geq t_1$, integrating the above inequality from b to t_1 , we have

$$\frac{a(t)\psi(x(t))x'(t)}{f(x(t))} \leq \frac{a(b)\psi(x(b))x'(b)}{f(x(b))} - \int_b^t [q(s) - p(s)] ds. \quad (4)$$

Then for the above C_N when $t \geq C_N$ we have

$$\frac{a(t)\psi(x(t))x'(t)}{f(x(t))} \leq \frac{a(C_N)\psi(x(C_N))x'(C_N)}{f(x(C_N))} - \int_{C_N}^t [q(s) - p(s)] ds.$$

Therefore

$$\begin{aligned} \limsup_{t \rightarrow +\infty} \frac{a(t)\psi(x(t))x'(t)}{f(x(t))} &\leq \frac{a(C_N)\psi(x(C_N))x'(C_N)}{f(x(C_N))} \\ &\quad + \limsup_{t \rightarrow +\infty} \left\{ - \int_{C_N}^t [q(s) - p(s)] ds \right\} < 0. \end{aligned}$$

Then for all $t \geq C_N$ we obtain $x'(t) < 0$, which gives a contradiction since $x'(t_{\alpha_1})x'(t_{\alpha_2}) < 0$. The proof is complete.

Lemma 2 *If*

$$\int_{t_0}^{+\infty} [q(s) - p(s)] ds < +\infty \quad (5)$$

and

$$\lim_{t \rightarrow +\infty} \int_{t_0}^t \frac{1}{a(s)} \int_s^{+\infty} [q(\tau) - p(\tau)] d\tau ds = +\infty, \quad (6)$$

furthermore $\frac{f(u)}{\psi(u)}$ is strongly sublinear, that is,

$$\int_\varepsilon^{+\infty} \frac{\psi(u)}{f(u)} du < +\infty, \quad \int_{-\infty}^{-\varepsilon} \frac{\psi(u)}{f(u)} du > -\infty, \quad (7)$$

for all $\varepsilon > 0$, then for Eq.(1), we have $S^+ = \emptyset$.

Proof Suppose that Eq.(1) has a solution $x(t) \in S^-$. There is no loss of generality in assuming that there exists $t_1 \geq t_0$ such that $x(t) > 0, x'(t) \geq 0$ for all $t \geq t_1$. (For $x(t) < 0, x'(t) \leq 0$, the proof is similar.) As in the proof of lemma 1 we can acquire (4). From (5) we obtain ($t \geq b, x'(t) \geq 0$),

$$0 \leq \frac{a(b)\psi(x(b))x'(b)}{f(x(b))} - \int_b^{+\infty} [q(s) - p(s)] ds.$$

Thus for all $t \geq b$ we have

$$\int_t^{+\infty} [q(s) - p(s)] ds \leq \frac{a(t)\psi(x(t))x'(t)}{f(x(t))}.$$

So we can obtain

$$\int_b^t \frac{1}{a(s)} \int_s^{+\infty} [q(\tau) - p(\tau)] d\tau ds \leq \int_b^t \frac{\psi(x(s))x'(s)}{f(x(s))} ds,$$

let $t \rightarrow +\infty$, which contradicts condition (6) and (7). The proof is complete.

Theorem 1 *If the condition (3),(5) and (7) hold and assume that*

$$\lim_{t \rightarrow +\infty} \int_{t_0}^t \frac{1}{a(s)} ds = +\infty \quad (8)$$

is satisfied. Then Eq.(1) is oscillatory.

Proof It follows from Lemma 1 and Lemma 2, $S^+ = S^{WO} = \emptyset$ for Eq.(1). Therefore, to prove Theorem 1, it suffices to show that $S^- = \emptyset$ for Eq.(1). Let $x(t)$ be a solution of class S^- of Eq.(1). There is no loss of generality in assuming that there exists $t_1 \geq t_0$ such that $x(t) > 0, x'(t) \leq 0$ for all $t \geq t_1$. (For $x(t) < 0, x'(t) \geq 0$, the proof is similar.) It follows from (3) there exists $t_2 \geq t_1$ such that

$$\int_{t_2}^t [q(s) - p(s)] ds \geq 0$$

for $t \geq t_2$. From Eq.(1), $x'(t) \neq 0$ for $t \geq t_2$. In fact, If $t \geq t_2$ then $x'(t) \equiv 0$, because of (A_3) and (A_4) we have $q(t) \leq p(t)$, which contradicts condition (3). So $x'(t) \neq 0$ for $t \geq t_2$. There exists $t_3 \geq t_2$ such that $x'(t_3) < 0$. Integrating Eq.(1) from t_3 to t , we have

$$\begin{aligned} a(t)\psi(x(t))x'(t) &= a(t_3)\psi(x(t_3))x'(t_3) \\ &+ \int_{t_3}^t P(s, x(s), x'(s)) ds - \int_{t_3}^t Q(s, x(s)) ds \\ &\leq a(t_3)\psi(x(t_3))x'(t_3) + \int_{t_3}^t p(s)f(x(s)) ds \\ &\quad - \int_{t_3}^t q(s)f(x(s)) ds. \\ &= a(t_3)\psi(x(t_3))x'(t_3) \\ &\quad - \int_{t_3}^t [q(s) - p(s)]f(x(s)) ds \\ &= a(t_3)\psi(x(t_3))x'(t_3) \\ &\quad - f(x(t)) \int_{t_3}^t [q(s) - p(s)] ds \\ &\quad + \int_{t_3}^t f'(x(s))x'(s) \int_{t_3}^s [q(\tau) - p(\tau)] d\tau ds \\ &\leq a(t_3)\psi(x(t_3))x'(t_3) = k \quad (k < 0). \end{aligned}$$

Consequently, for all $t \geq t_3$, we have

$$\int_{x(t_3)}^{x(t)} \psi(u) du \leq k \int_{t_3}^t \frac{1}{a(s)} ds,$$

Noting the condition (8) and the fact $0 < x(t) \leq x(t_3)$, implies the left of this inequality, that is $\int_{x(t_3)}^{x(t)} \psi(u) du$ has lower bounded, but the right of it tend towards minus infinity. So a contradiction exists. The proof is complete.

Lemma 3 *If the condition (7) hold and*

$$\lim_{t \rightarrow +\infty} \int_T^t [q(s) - p(s)] \int_T^s \frac{1}{a(\tau)} d\tau ds = +\infty \quad (9)$$

is satisfied, then $S^+ = \emptyset$ and $S^{WO} = \emptyset$ for Eq.(1).

Proof (I) Suppose that Eq.(1) has a solution $x(t) \in S^+$. There is no loss of generality in assuming that there exists $t_1 \geq t_0$ such that $x(t) > 0, x'(t) \geq 0$ for all $t \geq t_1$. (For $x(t) < 0, x'(t) \leq 0$, the proof is similar.) Consider the function

$$W(t) = \frac{-a(t)\psi(x(t))x'(t)}{f(x(t))} \int_{t_1}^t \frac{1}{a(s)} ds, \quad t \geq t_1.$$

Then it follows from Eq.(1) that

$$\begin{aligned} &W'(t) \\ &= \frac{-(a(t)\psi(x(t))x'(t))'}{f(x(t))} \int_{t_1}^t \frac{1}{a(s)} ds \\ &\quad + \frac{a(t)\psi(x(t))x'^2(t)f'(x(t))}{f^2(x(t))} \int_{t_1}^t \frac{1}{a(s)} ds \\ &\quad - \frac{\psi(x(t))x'(t)}{f(x(t))} \\ &= \left[\frac{Q(t, x(t))}{f(x(t))} - \frac{P(t, x(t), x'(t))}{f(x(t))} \right. \\ &\quad \left. + \frac{a(t)\psi(x(t))x'^2(t)f'(x(t))}{f^2(x(t))} \right] \int_{t_1}^t \frac{1}{a(s)} ds \\ &\quad - \frac{\psi(x(t))x'(t)}{f(x(t))} \\ &\geq [q(t) - p(t) + a(t)\psi(x(t))f'(x(t)) \frac{x'^2(t)}{f^2(x(t))}] \\ &\quad \cdot \int_{t_1}^t \frac{1}{a(s)} ds - \frac{\psi(x(t))x'(t)}{f(x(t))} \\ &\geq [q(t) - p(t)] \int_{t_1}^t \frac{1}{a(s)} ds - \frac{\psi(x(t))x'(t)}{f(x(t))}. \end{aligned}$$

So

$$W(t) \geq \int_{t_1}^t [q(s) - p(s)] \int_{t_1}^s \frac{1}{a(\tau)} d\tau ds - \int_{t_1}^t \frac{\psi(x(s))x'(s)}{f(x(s))} ds. \quad (10)$$

Noting the condition (7) and (9) we obtain

$$\lim_{t \rightarrow +\infty} W(t) = +\infty,$$

which contradicts the assumption $W(t) < 0$.

(II) Suppose that Eq.(1) has a solution $x(t) \in S^{WO}$. There is no loss of generality in assuming that there exists $t_1 \geq t_0$ such that $x(t) > 0$ for all $t \geq t_1$ since the proof is similar if $x(t) < 0$ for all large t . Since for all $t_\alpha > t_1$ there exist $t_{\alpha_1}, t_{\alpha_2} > t_\alpha$, such that $x'(t_{\alpha_1})x'(t_{\alpha_2}) < 0$. Proceeding as in the proof of the above (I), we obtain (10), thus,

$$\liminf_{t \rightarrow +\infty} W(t) \geq \liminf_{t \rightarrow +\infty} \int_{t_1}^t [q(s) - p(s)] \cdot \int_{t_1}^s \frac{1}{a(\tau)} d\tau ds + \liminf_{t \rightarrow +\infty} \left\{ - \int_{t_1}^t \frac{\psi(x(s))x'(s)}{f(x(s))} ds \right\}.$$

Noting the condition (7),

$$\limsup_{t \rightarrow +\infty} \int_{t_1}^t \frac{\psi(x(s))x'(s)}{f(x(s))} ds$$

has upper bound. In fact, from the condition (7) we know it has upper bound for $x'(s) > 0$. And for $x'(s) < 0$ we know 0 is upper bound. Then

$$\liminf_{t \rightarrow +\infty} \left\{ - \int_{t_1}^t \frac{\psi(x(s))x'(s)}{f(x(s))} ds \right\}$$

has lower bound. Noting the condition (9) we have $x'(t) < 0$ for all large t , which gives a contradiction since $x'(t_{\alpha_1})x'(t_{\alpha_2}) < 0$. The proof is complete.

Theorem 2 *If the assumptions (7),(8) and (9) are satisfied, then Eq.(1) is oscillatory.*

Proof It follows from Lemma 3 that $S^+ = S^{WO} = \emptyset$. Therefore, to prove Theorem 2, it suffices to show that $S^- = \emptyset$ for Eq.(1). Let $x(t)$ be a solution of type S^- of Eq.(1). Without loss of generality, we may assume that there exists $t_1 \geq t_0$ such that $x(t) > 0, x'(t) \leq 0$ for all $t \geq t_1$. (For $x(t) < 0, x'(t) \geq 0$, the proof is similar.) Consider the function

$$W(t) = \frac{-a(t)\psi(x(t))x'(t)}{f(x(t))} \int_{t_1}^t \frac{1}{a(s)} ds, \quad t \geq t_1.$$

As in the proof of Lemma 3 we obtain (10), i.e.,

$$W(t) \geq \int_{t_1}^t [q(s) - p(s)] \int_{t_1}^s \frac{1}{a(\tau)} d\tau ds - \int_{t_1}^t \frac{\psi(x(s))x'(s)}{f(x(s))} ds.$$

In view of condition (9), $W(t) \rightarrow +\infty$ for $t \rightarrow +\infty$. Then there exists $t_2 \geq t_1$ such that $W(t) \geq 1$ for $t \geq t_2$. Therefore

$$\frac{\psi(x(t))x'(t)}{f(x(t))} \leq -\frac{1}{a(t) \int_{t_1}^t \frac{1}{a(s)} ds}, \quad t \geq t_2.$$

Let $A(t, t_1) = \int_{t_1}^t \frac{1}{a(s)} ds$, from the above we can acquire $x'(t) < 0$ for $t \geq t_2$ and

$$\int_{x(t_2)}^{x(t)} \frac{\psi(u)}{f(u)} du \leq - \int_{t_2}^t \frac{1}{a(s)A(s, t_1)} ds = - \ln \frac{A(t, t_1)}{A(t_2, t_1)} \rightarrow -\infty \quad (t \rightarrow +\infty).$$

Then $x(t) \rightarrow 0$ ($t \rightarrow +\infty$). It also follows from (9) that there exists $t_3 \geq t_2$ such that

$$\int_{t_3}^t [q(s) - p(s)] \int_{t_3}^s \frac{1}{a(\tau)} d\tau ds \geq 0, \quad t \geq t_3.$$

Integrating Eq.(1) we have

$$\begin{aligned} & \int_{t_3}^t (a(s)\psi(x(s))x'(s))' \int_{t_3}^s \frac{1}{a(\tau)} d\tau ds \\ &= \int_{t_3}^t P(s, x(s), x'(s)) \int_{t_3}^s \frac{1}{a(\tau)} d\tau ds \\ &- \int_{t_3}^t Q(s, x(s)) \int_{t_3}^s \frac{1}{a(\tau)} d\tau ds. \end{aligned}$$

Then integrating the left and in the same time

$$\begin{aligned} & a(t)\psi(x(t))x'(t) \int_{t_3}^t \frac{1}{a(\tau)} d\tau \\ &\leq \int_{x(t_3)}^{x(t)} \psi(u) du \\ &- \int_{t_3}^t [q(s) - p(s)] f(x(s)) \int_{t_3}^s \frac{1}{a(\tau)} d\tau ds \\ &= \int_{x(t_3)}^{x(t)} \psi(u) du \\ &- f(x(t)) \int_{t_3}^t [q(s) - p(s)] \int_{t_3}^s \frac{1}{a(\tau)} d\tau ds \\ &+ \int_{t_3}^t f'(x(s))x'(s) \cdot \int_{t_3}^s [q(u) - p(u)] \int_{t_3}^u \frac{1}{a(\tau)} d\tau du ds \\ &\leq \int_{x(t_3)}^{x(t)} \psi(u) du. \end{aligned} \tag{11}$$

Because $x(t) \rightarrow 0$ for $t \rightarrow +\infty$, there exists $t_4 \geq t_3$ such that $x(t) < \frac{x(t_3)}{2}$ for all $t \geq t_4$. Thus, there exists constant $L > 0$, making that

$$\int_{x(t_3)}^{x(t)} \psi(u) du < -L.$$

Therefore, when $t \geq t_4$, from the above formula (11), we can obtain

$$\psi(x(t))x'(t) \leq -L \frac{1}{a(t) \int_{t_3}^t \frac{1}{a(\tau)} d\tau}.$$

Consequently,

$$\int_{x(t_4)}^{x(t)} \psi(u) du \leq -L \ln \frac{A(t, t_3)}{A(t_4, t_3)} \rightarrow -\infty \quad (t \rightarrow +\infty).$$

But the left of this inequality, that is $\int_{x(t_4)}^{x(t)} \psi(u) du$ has lower bounded. So a contradiction exists. The proof is complete.

3 Examples

In this section, we give two illustrative examples.

Example 1 Consider the equation

$$\begin{aligned} & \left(\frac{1}{2t^2} x^{-2}(t) x'(t) \right)' + \frac{1}{t^2} x(t) \\ & + \frac{1}{4(1+t)^2} x(t) [x'(t)]^2 = 0, \quad (t > 0). \end{aligned} \quad (12)$$

where $a(t) = \frac{1}{2}t^{-2}$, $\psi(u) = u^{-2}$, let $q(t) = t^{-2}$, $p(t) = 0$, $f(u) = u$. It is easy to verify that Eq.(12) satisfies the conditions of Theorem 1. Therefore, Eq.(12) is oscillatory. However, using any known results, we can not obtain the conclusion.

Example 2 Consider the equation

$$\begin{aligned} & \left(\frac{1}{2t^{\frac{3}{2}}} x^{-\frac{4}{3}}(t) x'(t) \right)' + \frac{1}{t^{\frac{3}{2}}} x^{\frac{1}{3}}(t) \\ & + \frac{1}{4(1+t)^{\frac{3}{2}}} x^{\frac{1}{3}}(t) [x'(t)]^2 = 0, \quad (t > 0). \end{aligned} \quad (13)$$

where $a(t) = \frac{1}{2}t^{-\frac{3}{2}}$, $\psi(u) = u^{-\frac{4}{3}}$, let $q(t) = t^{-\frac{3}{2}}$, $p(t) = 0$, $f(u) = u^{\frac{1}{3}}$. It is easy to check that Eq.(13) satisfies all the conditions of Theorem 2. Therefore, Eq.(13) is oscillatory. However, using any known results, we can not obtain the conclusion.

References

- [1] Q.X.Zhang and J.R.Yan, Oscillatory behavior of a second order nonlinear differential equation with damping, *Journal of Systems Science and Mathematical Science*, 2004,24(3): 296-302.
- [2] J.R.Yan and Q.X.Zhang, Oscillatory theorems for second order nonlinear differential equations with damping, *Journal of Systems Science and Mathematical Science*, 1993,13(3): 276-278.
- [3] M. Cecchi and M. Marini, Oscillatory and nonoscillatory behavior of a second order functional differential equation, *Rocky Mount. J. Math.*, 22(1992), 1259-1276.
- [4] Yu.V. Rogovchenko, On oscillation of a second order nonlinear delay differential equation, *Funkcial. Ekvac.*, 43(2000), 1-29.
- [5] J.R.Yan, Oscillation theorems for second order linear differential equations with damping, *Proc. Amer. Math. Soc.* 98(1986), 276-282.
- [6] G. S. Ladde, V. Lakshmikantham and B. G. Zhang, *Oscillation Theory of Differential Equations with Deviating Arguments*, Marcel Dekker, New York, 1987.

On quasistatic models of contact phenomena

Anna Ochal and Stanislaw Migorski

Abstract—We deal with a quasistatic model for contact problem which involves general nonmonotone and multivalued laws and subdifferential boundary conditions. We study the asymptotic behavior of a weak solution to an abstract second order nonlinear evolution hemivariational inequality modeling contact problems, as the power of the acceleration forces vanishes.

First we prove the existence and regularity result for the evolution hemivariational inequality. Then the crucial step is to derive a suitable uniform estimate on the sequence of solutions and the existence of a limit for this sequence which is a solution of a parabolic model for a quasistatic hemivariational inequality. Finally, we present an application of the abstract theorem to a quasistatic viscoelastic contact problem.

Index Terms—Contact problem, asymptotic behavior, hemivariational inequality, solid mechanics.

I. INTRODUCTION

THE aim of this paper is to study the asymptotic behavior of a weak solution to an abstract second order nonlinear evolution hemivariational inequality modeling contact problems when the power of the acceleration forces vanishes. We consider the following evolution inclusion being an equivalent form of a hemivariational inequality

$$\varepsilon u''(t) + A(t)u'(t) + Bu(t) + \gamma^* \partial J(t, \gamma u(t)) \ni f(t) \quad (1)$$

where $\varepsilon > 0$ is proportional to the mass density, $A(t)$ and B are linear continuous operators from a reflexive Banach space V into its dual V^* , γ is a linear continuous operator, γ^* is its adjoint, ∂J denotes the Clarke subdifferential of a locally Lipschitz time-dependent function $J(t, \cdot)$ and $f: (0, T) \rightarrow V^*$ is given. Moreover, we associate with (1) appropriate initial conditions.

The study of this type of inclusions is motivated by many contact problems of solid mechanics. It is well known that the dynamic equation of motion, representing momentum conservation, that govern the evolution of the state of the body, is of the form $u'' - \text{Div } \sigma = f$, where u is a displacement, σ is the stress tensor, Div is the divergence operator and f is the density (per unit volume) of applied forces, such as gravity. This equation is valid for all systems and materials, since they are derived from the fundamental principle of momentum conservation.

We are interested in situations in which the system configuration and the external forces and tractions evolve slowly

Manuscript received April 6, 2009. This work was done within the project Polonium "Nonsmooth Analysis with Applications to Contact Mechanics" under contract no. 7817/R09/R10 between the Jagiellonian University, Kraków, Poland and Université de Perpignan, France. The first two authors were supported in part by the Ministry of Science and Higher Education of Poland under grant no. N201 027 32/1449.

A. Ochal and S. Migorski are with Jagiellonian University, Faculty of Mathematics and Computer Science, Krakow, Poland, (e-mail: {ochal,migorski}@softlab.ii.uj.edu.pl).

in time in such a way that the accelerations in the system are rather small and negligible, so that the inertial terms (the second order time derivative) can be neglected. In this case, we obtain the quasistatic approximation (equilibrium equation) for the equation of motion, $-\text{Div } \sigma = f$. Rigorous mathematical treatment of quasistatic problems is recent, e.g. Han and Sofonea [5], Shillor, Sofonea and Telega [13], Sofonea, Han, Shillor [14] and the references therein. When we assume that the process is slow and the accelerations are negligible, mathematically it means that the system changes character, from being of a hyperbolic type to an elliptic or a parabolic type.

Our main interest lies in general nonmonotone and multivalued laws and subdifferential boundary conditions. Such relations were considered for the first time in the early eighties by P.D. Panagiotopoulos [11] who introduced the notion of hemivariational inequality as a generalization of variational inequality. For examples and applications of hemivariational inequalities, we refer to Panagiotopoulos [11], [12] and Naniewicz and Panagiotopoulos [9].

Our motivation to study the model (1) comes from a quasistatic viscoelastic contact problem. We deal with a deformable viscoelastic body which occupies an open bounded subset Ω of \mathbb{R}^d ($d = 2, 3$ in applications). The boundary Γ of Ω is supposed to be Lipschitz continuous and Γ consists of three mutually disjoint measurable parts Γ_D , Γ_N and Γ_C such that $m(\Gamma_D) > 0$. Setting $Q = \Omega \times (0, T)$, we are looking for the displacement field $u: Q \rightarrow \mathbb{R}^d$ which solves the following mechanical problem

$$-\text{Div } \sigma(t) = f_0(t) \quad \text{in } Q \quad (2)$$

$$\sigma(t) = \mathcal{C}(t)e(u'(t)) + \mathcal{G}e(u(t)) \quad \text{in } Q \quad (3)$$

$$u(t) = 0 \quad \text{on } \Gamma_D \times (0, T) \quad (4)$$

$$\sigma(t)n = f_1(t) \quad \text{on } \Gamma_N \times (0, T) \quad (5)$$

$$-\sigma_\nu(t) \in \partial j_\nu(t, u_\nu) \quad \text{on } \Gamma_C \times (0, T) \quad (6)$$

$$-\sigma_\tau(t) \in \partial j_\tau(t, u_\tau) \quad \text{on } \Gamma_C \times (0, T) \quad (7)$$

$$u(0) = u_0 \quad \text{in } \Omega. \quad (8)$$

Here σ and e denote the stress and strain tensors, \mathcal{C} and \mathcal{G} are viscosity and elasticity operators, ∂j_ν and ∂j_τ are the Clarke subdifferentials of prescribed locally Lipschitz functions j_ν and j_τ . The detailed description of the contact problem (2)–(8) is given in Section V.

In the paper we start with an existence and regularity result for the evolution hemivariational inequality. Such a result holds for every fixed $\varepsilon > 0$ and it is obtained (cf. Theorem 4) by using the theory of hemivariational inequalities developed recently in [6], [7], [8]. Then, the crucial step is to derive a suitable uniform estimate on the sequence of solutions to (1). The main result concerns the existence of a limit as $\varepsilon \rightarrow 0$, for

this sequence. Our asymptotic analysis shows that a parabolic model for a quasistatic hemivariational inequality of the form

$$A(t)u'(t) + Bu(t) + \gamma^* \partial J(t, \gamma u(t)) \ni f(t) \quad \text{a.e. } t$$

is solvable as a limit of the dynamical one, as the mass coefficient in the inertial term tends to zero. We also point out that the results of this paper can be applicable to several quasistatic models of contact phenomena. Finally, we remark that the question of uniqueness of solutions to both dynamical and quasistatic hemivariational inequality is left open.

II. NOTATION

Let (V, Z, H, Z^*, V^*) be an evolution fivefold, i.e. V and Z are reflexive, separable Banach spaces, H is a Hilbert space and $V \subset Z \subset H \approx H^* \subset Z^* \subset V^*$ with all embeddings dense and continuous. Suppose that V is compactly embedded in Z . We denote by $\langle \cdot, \cdot \rangle$ the duality of V and V^* and the pairing between Z and Z^* as well, by $\|\cdot\|_E$ the norm in a space E , being, respectively V , Z , Z^* and V^* and by $|\cdot|_H$ the norm in H . Moreover, the notation $\mathcal{L}(E, F)$ stands for the space of linear bounded operators from a Banach space E to a Banach space F and w - E denotes the space E endowed with the weak topology. For $U \subset E$, we also write $\|U\|_E = \sup\{\|u\|_E : u \in U\}$.

Given a fixed number $0 < T < \infty$ and $2 \leq p < \infty$, we introduce the following function spaces $\mathcal{V} = L^p(0, T; V)$, $\mathcal{Z} = L^p(0, T; Z)$, $\mathcal{H} = L^2(0, T; H)$, $\mathcal{Z}^* = L^q(0, T; Z^*)$, $\mathcal{V}^* = L^q(0, T; V^*)$ with $1/p + 1/q = 1$ and $\mathcal{W} = \{w \in \mathcal{V} : w' \in \mathcal{V}^*\}$, where the time derivative is understood in the sense of vector-valued distributions. The latter is a separable, reflexive Banach space with the norm $\|w\|_{\mathcal{W}} = \|w\|_{\mathcal{V}} + \|w'\|_{\mathcal{V}^*}$. We have

$$\mathcal{W} \subset \mathcal{V} \subset \mathcal{Z} \subset \mathcal{H} \subset \mathcal{Z}^* \subset \mathcal{V}^*$$

with continuous embeddings, cf. [3], [4]. The pairing of \mathcal{V} and \mathcal{V}^* and also the duality between \mathcal{Z} and \mathcal{Z}^* are denoted by $\langle \langle \cdot, \cdot \rangle \rangle$.

Next, we recall the definitions of the generalized directional derivative and the generalized gradient of Clarke for a locally Lipschitz function $\theta: X \rightarrow \mathbb{R}$, where X is a Banach space (see [2], [3]). The generalized directional derivative of θ at $x \in X$ in the direction $v \in X$, denoted by $\theta^0(x; v)$, is defined by

$$\theta^0(x; v) = \limsup_{y \rightarrow x, \lambda \downarrow 0} \frac{\theta(y + \lambda v) - \theta(y)}{\lambda}.$$

The generalized gradient of θ at x , denoted by $\partial\theta(x)$, is a subset of a dual space X^* given by

$$\partial\theta(x) = \{\zeta \in X^* : \theta^0(x; v) \geq \langle \zeta, v \rangle \text{ for all } v \in X\}.$$

A locally Lipschitz function θ is called regular (in the sense of Clarke, cf. [2]) at $x \in X$ if for all $v \in X$ the one-sided directional derivative $\theta'(x; v)$ exists and satisfies $\theta^0(x; v) = \theta'(x; v)$ for all $v \in X$.

III. HEMIVARIATIONAL INEQUALITY

We provide a result on the existence of solutions to an evolution hemivariational inequality of the form

$$\begin{cases} \varepsilon u''(t) + A(t)u'(t) + Bu(t) + \gamma^* \partial J(t, \gamma u(t)) \ni f(t) \\ \text{a.e. } t \in (0, T) \\ u(0) = u_0, \quad \sqrt{\varepsilon} u'(0) = u_1. \end{cases} \quad (9)$$

In this section we suppose that $\varepsilon > 0$ is fixed. We remark that, by the definition of the Clarke subdifferential, problem (9) is equivalent to the following inequality

$$\begin{cases} \langle \varepsilon u''(t) + A(t)u'(t) + Bu(t) - f(t), v \rangle + \\ + J^0(t, \gamma u(t); \gamma v) \geq 0 \text{ for all } v \in V, \text{ a.e. } t \\ u(0) = u_0, \quad \sqrt{\varepsilon} u'(0) = u_1, \end{cases}$$

where J^0 stands for the generalized directional derivative of $J(t, \cdot)$. For this reason problem (9) is called a hemivariational inequality.

DEFINITION 1: A function $u \in \mathcal{V}$ is called a solution to (9) if and only if $u' \in \mathcal{W}$ and there exists $\eta \in \mathcal{Z}^*$ such that

$$\begin{cases} \varepsilon u''(t) + A(t)u'(t) + Bu(t) + \eta(t) = f(t) \text{ a.e. } t \\ \eta(t) \in \gamma^* \partial J(t, \gamma u(t)) \text{ a.e. } t \in (0, T) \\ u(0) = u_0, \quad \sqrt{\varepsilon} u'(0) = u_1. \end{cases}$$

REMARK 2: If u is a solution to (9), then $u \in W^{1,p}(0, T; V)$. Since the embeddings $W^{1,p}(0, T; V) \subset C(0, T; V)$ and $\mathcal{W} \subset C(0, T; H)$ are continuous (cf. Proposition 8.4.14 in [4]), the initial conditions $u(0)$ and $u'(0)$ have a meaning in V and H , respectively.

Since ε is fixed, we write for simplicity u for the solution u_ε of (9). We assume that X is a Banach space and the following hypotheses hold.

$H(A)$: $A \in L^\infty(0, T; \mathcal{L}(V, V^*))$ is an operator such that $A(t)$ is coercive, i.e. there is a constant $\alpha > 0$ such that for a.e. $t \in (0, T)$, $\langle A(t)v, v \rangle \geq \alpha \|v\|_V^p$ for all $v \in V$.

$H(B)$: $B \in \mathcal{L}(V, V^*)$ is nonnegative and symmetric.

$H(J)$: $J: (0, T) \times X \rightarrow \mathbb{R}$ is a function such that

- (i) $J(\cdot, x)$ is measurable on $(0, T)$ for all $x \in X$;
- (ii) $J(t, \cdot)$ is locally Lipschitz on X for a.e. $t \in (0, T)$;
- (iii) $\|\partial J(t, x)\|_{X^*} \leq \bar{c}(1 + \|x\|_X^{2/q})$ for all $x \in X$, a.e. $t \in (0, T)$ with $\bar{c} > 0$.

$H(\gamma)$: $\gamma \in \mathcal{L}(Z, X)$.

(H_0) : $f \in \mathcal{V}^*$, $u_0 \in V$, $u_1 \in H$.

(H_1) : if $p = 2$, then $\alpha > 2\bar{c}T\beta^2 \|\gamma\| \max\{1, \|\gamma\|\}$, where $\beta > 0$ is an embedding constant of V into Z and $\|\gamma\| = \|\gamma\|_{\mathcal{L}(Z, X)}$.

The crucial step in the proof of the existence result for (9) is to derive the following uniform estimate.

LEMMA 3: Let $\varepsilon > 0$ be fixed. Assume hypotheses $H(A)$, $H(B)$, $H(J)$, $H(\gamma)$ and (H_0) , and let u be a solution to (9). If $p > 2$, then there exists a constant $C > 0$ independent of ε such that

$$\begin{aligned} \|u\|_{C(0, T; V)} + \|u'\|_{\mathcal{V}} + \sqrt{\varepsilon} \|u'\|_{L^\infty(0, T; H)} + \varepsilon \|u''\|_{\mathcal{V}^*} &\leq \\ &\leq C \left(1 + \|u_0\|_V^{2/q} + |u_1|_H^{2/q} + \|f\|_{\mathcal{V}^*}^{2/q} \right). \end{aligned}$$

Moreover, this estimate still holds for $p = 2$ provided (H_1) is satisfied.

The following is the existence result for a hemivariational inequality (9).

THEOREM 4: If hypotheses $H(A)$, $H(B)$, $H(J)$, $H(\gamma)$, (H_0) and (H_1) hold, then for every fixed $\varepsilon > 0$ the problem (9) admits at least one solution.

Proof. The idea of the proof follows Theorem 4 of [10]. Using an integral operator K we reduce the problem (9) to an evolution inclusion of the first order

$$\begin{cases} f(t) \in \varepsilon z'(t) + A(t)z(t) + BKz(t) + \gamma^* \partial J(t, \gamma Kz(t)) \\ z(0) = \tilde{u}_1, \end{cases}$$

where $\tilde{u}_1 = u_1/\sqrt{\varepsilon}$. We proceed with two cases. First consider the case $u_1 \in V$. By employing the surjectivity result for the sum of a linear, densely defined and maximal monotone operator and bounded, coercive and L -pseudomonotone one, we obtain the existence of solutions of the first order problem. The crucial point here is to establish the L -generalized pseudomonotonicity of corresponding evolution operator. Finally, by a density argument we remove the restriction $u_1 \in V$ and prove the result in the general case $u_1 \in H$. \square

IV. ASYMPTOTIC ANALYSIS

We consider the asymptotic behavior of a sequence of solutions to (9) when a small parameter in the inertial term tends to zero.

Consider the following evolutionary hemivariational inequality of the form:

$$\begin{cases} A(t)u'(t) + Bu(t) + \gamma^* \partial J(t, \gamma u(t)) \ni f(t) \text{ a.e. } t \\ u(0) = u_0. \end{cases} \quad (10)$$

DEFINITION 5: A function $u \in L^\infty(0, T; V)$ is called a solution to (10) if and only if $u' \in \mathcal{V}$ and there exists $\eta \in \mathcal{Z}^*$ such that

$$\begin{cases} A(t)u'(t) + Bu(t) + \eta(t) = f(t) \text{ a.e. } t \in (0, T) \\ \eta(t) \in \gamma^* \partial J(t, \gamma u(t)) \text{ a.e. } t \in (0, T) \\ u(0) = u_0. \end{cases}$$

If u is a solution to (10), then $u \in L^\infty(0, T; V) \subset \mathcal{V}$, i.e. $u \in W^{1,p}(0, T; V)$. Since $W^{1,p}(0, T; V) \subset C(0, T; V)$ continuously, the initial condition $u(0)$ has a meaning in V .

THEOREM 6: Let the hypotheses $H(A)$, $H(B)$, $H(J)$, $H(\gamma)$, (H_0) and (H_1) hold. For every ε fixed, let u_ε be a solution to (9) given by Theorem 4. Then there exists $u \in L^\infty(0, T; V)$ such that $u' \in \mathcal{W}$ and

$$\begin{aligned} u_\varepsilon &\rightarrow u \text{ weakly* in } L^\infty(0, T; V), \\ u'_\varepsilon &\rightarrow u' \text{ weakly in } \mathcal{V}, \\ \sqrt{\varepsilon}u'_\varepsilon &\rightarrow 0 \text{ weakly* in } L^\infty(0, T; H), \\ \varepsilon u''_\varepsilon &\rightarrow 0 \text{ weakly in } \mathcal{V}^*, \end{aligned}$$

as $\varepsilon \rightarrow 0$. Moreover, the limit function u is a solution to (10).

Proof. Let $\{u_\varepsilon\}$ be a sequence of solutions to the evolution problem (9). From Lemma 3, we deduce that there is a constant $c > 0$ independent of ε such that

$$\|u_\varepsilon\|_{C(0, T; V)} + \|u'_\varepsilon\|_{\mathcal{V}} + \sqrt{\varepsilon}\|u'_\varepsilon\|_{L^\infty(0, T; H)} + \varepsilon\|u''_\varepsilon\|_{\mathcal{V}^*} \leq c.$$

Hence, taking a subsequence of $\varepsilon \rightarrow 0$ if necessary, we have

$$u_\varepsilon \rightarrow u \text{ weakly* in } L^\infty(0, T; V), \quad (11)$$

$$u'_\varepsilon \rightarrow u' \text{ weakly in } \mathcal{V}, \quad (12)$$

$$\sqrt{\varepsilon}u'_\varepsilon \rightarrow \zeta \text{ weakly* in } L^\infty(0, T; H), \quad (13)$$

$$\varepsilon u''_\varepsilon \rightarrow \rho \text{ weakly in } \mathcal{V}^* \quad (14)$$

with $u \in L^\infty(0, T; V)$, $u' \in \mathcal{V}$, $\zeta \in L^\infty(0, T; H)$ and $\rho \in \mathcal{V}^*$. It is clear that (11) implies $\sqrt{\varepsilon}u_\varepsilon \rightarrow 0$ weakly* in $L^\infty(0, T; V)$ which together with (13) gives $\zeta = 0$. Consequently from (13) and (14) we have $\rho = 0$.

The convergences (11) and (12) entail $u_\varepsilon \rightarrow u$ weakly in $W^{1,p}(0, T; V)$, so we may show that

$$u_\varepsilon(t) \rightarrow u(t) \text{ weakly in } V \text{ for all } t \in [0, T] \quad (15)$$

and because $V \subset Z$ compactly, we have

$$u_\varepsilon(t) \rightarrow u(t) \text{ in } Z \text{ for all } t \in [0, T]. \quad (16)$$

Since u_ε is a solution to (9), we have

$$\begin{cases} \varepsilon u''_\varepsilon(t) + A(t)u'_\varepsilon(t) + Bu_\varepsilon(t) + \eta_\varepsilon(t) = f(t) \text{ a.e. } t \\ \eta_\varepsilon(t) \in \gamma^* \partial J(t, \gamma u_\varepsilon(t)) \text{ a.e. } t \in (0, T) \\ u_\varepsilon(0) = u_0, \quad \sqrt{\varepsilon}u'_\varepsilon(0) = u_1. \end{cases}$$

From $H(J)$ (iii) and $H(\gamma)$ we deduce that $\{\eta_\varepsilon\}$ remains in a bounded subset of \mathcal{Z}^* and we may assume, possibly up to a subsequence, that

$$\eta_\varepsilon \rightarrow \eta \text{ weakly in } \mathcal{Z}^* \quad (17)$$

with $\eta \in \mathcal{Z}^*$. Let $F: (0, T) \times Z \rightarrow 2^{\mathcal{Z}^*}$ be defined by $F(t, z) = \gamma^* \partial J(t, \gamma z)$ for $(t, z) \in (0, T) \times Z$. From Proposition 2.1.2 of [2], $H(J)$ and $H(\gamma)$, it follows that the values of F are nonempty, closed and convex subsets of \mathcal{Z}^* , for each $z \in Z$, $F(\cdot, z)$ is measurable, and for a.e. $t \in (0, T)$, $F(t, \cdot)$ is upper semicontinuous from Z into $w\text{-}\mathcal{Z}^*$. Exploiting (16), (17) and the inclusion $\eta_\varepsilon(t) \in \gamma^* \partial J(t, \gamma u_\varepsilon(t))$ for a.e. $t \in (0, T)$, from the convergence theorem of [1], we have

$$\eta(t) \in \gamma^* \partial J(t, \gamma u(t)) \text{ a.e. } t \in (0, T). \quad (18)$$

Let $\mathcal{A}, \mathcal{B}: \mathcal{V} \rightarrow \mathcal{V}^*$ be the Nemitsky operators corresponding to $A(t)$ and B . From $H(A)$ and $H(B)$ it is obvious that \mathcal{A} and \mathcal{B} are linear continuous operators from \mathcal{V} to \mathcal{V}^* and thus also continuous from $w\text{-}\mathcal{V}$ to $w\text{-}\mathcal{V}^*$. Therefore

$$\mathcal{A}u'_\varepsilon \rightarrow \mathcal{A}u', \quad \mathcal{B}u_\varepsilon \rightarrow \mathcal{B}u \text{ both weakly in } \mathcal{V}^*.$$

Hence, using (14) and (17) we pass to the limit in the equation $\varepsilon u''_\varepsilon + \mathcal{A}u'_\varepsilon + \mathcal{B}u_\varepsilon + \eta_\varepsilon = f$ in \mathcal{V}^* and obtain $\mathcal{A}u' + \mathcal{B}u + \eta = f$ in \mathcal{V}^* . This together with (18) implies

$$A(t)u'(t) + Bu(t) + \gamma^* \partial J(t, \gamma u(t)) \ni f(t) \text{ a.e. } t \in (0, T).$$

Finally, from (15), we know that $u_\varepsilon(0) \rightarrow u(0)$ weakly in V , so from $u_\varepsilon(0) = u_0$, we deduce that $u(0) = u_0$. The second initial condition is equivalent to

$$\varepsilon u'_\varepsilon(0) = \sqrt{\varepsilon}u_1. \quad (19)$$

By (13) and (14), we deduce

$$\varepsilon u'_\varepsilon \rightarrow 0 \text{ weakly* in } L^\infty(0, T; H),$$

$$\varepsilon u''_\varepsilon \rightarrow 0 \text{ weakly in } \mathcal{V}^*.$$

Since $L^\infty(0, T; H) \cap \{v : v' \in \mathcal{V}^*\} \subset C(0, T; V^*)$ compactly, we have $\varepsilon u'_\varepsilon \rightarrow 0$ in $C(0, T; V^*)$ and in particular, $\varepsilon u'_\varepsilon(0) \rightarrow 0$ in V^* . Taking the limit in (19), we deduce that the limit initial condition degenerates. We infer that u is a solution to (10). \square

From Theorem 6 we immediately have the existence and regularity result for the quasistatic model (10).

COROLLARY 7: Under hypotheses $H(A)$, $H(B)$, $H(J)$, $H(\gamma)$, (H_1) and for $f \in \mathcal{V}^*$ and $u_0 \in V$, the hemivariational inequality (10) admits at least one solution.

V. QUASISTATIC VISCOELASTIC CONTACT

We study the contact problem (2)–(8) between a viscoelastic body and a foundation. We assume that the volume forces and surface tractions change slowly in time so that the acceleration in the system is negligible. Neglecting the inertial terms in the equation of motion leads to the quasistatic approximation for the process. We show that the quasistatic model can be formulated as a time dependent hemivariational inequality of the form (10) and the abstract result of Theorem 4 is applicable in this case.

The mechanical formulation of the process is following. Recall that the boundary $\Gamma = \partial\Omega$ is supposed to be Lipschitz continuous and Γ consists of three mutually disjoint measurable parts Γ_D , Γ_N and Γ_C such that $m(\Gamma_D) > 0$. We are interested in the resulting process of evolution of the mechanical state on the time interval $(0, T)$. The body is clamped on Γ_D , the volume forces of density f_0 act in Ω and the surface tractions of density f_1 are applied on Γ_N . The body may come in contact with a foundation over a potential contact surface Γ_C . We denote by $u: Q \rightarrow \mathbb{R}^d$ the displacement field, by $\sigma: Q \rightarrow \mathcal{S}_d$ the stress tensor and by $e(u) = \{e_{ij}(u)\}$, $e_{ij}(u) = \frac{1}{2}(u_{i,j} + u_{j,i})$ the strain tensor, where $Q = \Omega \times (0, T)$, $i, j = 1, \dots, d$ and \mathcal{S}_d is the space of symmetric matrices of order d .

We assume a linear viscoelastic model with the constitutive law of the Kelvin-Voigt type

$$\sigma_{ij} = a_{ijkl}e_{kl}(u') + b_{ijkl}e_{kl}(u) \quad \text{in } Q,$$

where $\mathcal{C}(t) = \{a_{ijkl}(t)\}$ and $\mathcal{G} = \{b_{ijkl}\}$, $i, j, k, l = 1, \dots, d$ are the viscosity and the elasticity tensors, respectively. Denote by u_ν and u_τ (σ_ν and σ_τ , respectively) the normal and the tangential components of the displacement u (of the stress field σ , respectively) on Γ , i.e. $u_\nu = u \cdot n$, $u_\tau = u - u_\nu n$ ($\sigma_\nu = (\sigma n) \cdot n$, $\sigma_\tau = \sigma n - \sigma_\nu n$, respectively), where n is the outward normal vector to Γ .

Concerning the contact conditions, we consider the following subdifferential relations $-\sigma_\nu \in \partial j_\nu(x, t, u_\nu)$ and $-\sigma_\tau \in \partial j_\tau(x, t, u_\tau)$. The functions $j_\nu: \Gamma_C \times (0, T) \times \mathbb{R} \rightarrow \mathbb{R}$ and $j_\tau: \Gamma_C \times (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}$ are locally Lipschitz in their last variables and $\partial j_\nu, \partial j_\tau$ denote their Clarke subdifferentials. The initial displacement is denoted by u_0 . The classical formulation of the mechanical problem is as follows: find a displacement field $u: Q \rightarrow \mathbb{R}^d$ such that (2)–(8) are satisfied. For concrete examples of boundary conditions (6) and (7), we refer to [5], [8], [9], [11], [12].

In order to obtain a variational formulation of the problem (2)–(8) we need to define spaces: $V = \{v \in H^1(\Omega; \mathbb{R}^d) :$

$v = 0 \text{ on } \Gamma_D\}$, $Z = H^{1/2}(\Omega; \mathbb{R}^d)$, $H = L^2(\Omega; \mathbb{R}^d)$ and $X = L^2(\Gamma_C; \mathbb{R}^d)$. On V we consider the inner product and the corresponding norm given by $\langle u, v \rangle_V = \langle e(u), e(v) \rangle_{L^2(\Omega; \mathcal{S}_d)}$, $\|v\|_V = \|e(v)\|_{L^2(\Omega; \mathcal{S}_d)}$ for $u, v \in V$. Moreover, let $\gamma \in \mathcal{L}(Z, X)$ be the trace operator, and, for simplicity, $p = 2$.

We also assume that the viscosity and elasticity tensors have the usual properties of ellipticity, symmetry and positivity.

$\underline{H}(\mathcal{C})$: $\mathcal{C}: Q \times \mathcal{S}_d \rightarrow \mathcal{S}_d$ is a viscosity tensor, $\mathcal{C}(t) = \{a_{ijkl}(t)\}$ such that $a_{ijkl} = a_{klij} = a_{ijlk} \in L^\infty(Q)$ and there exists $m_1 > 0$ satisfying $\mathcal{C}(t)\tau \cdot \tau \geq m_1 \|\tau\|_{\mathcal{S}_d}^2$ for all $\tau \in \mathcal{S}_d$, a.e. in Q .

$\underline{H}(\mathcal{G})$: $\mathcal{G}: \Omega \times \mathcal{S}_d \rightarrow \mathcal{S}_d$ is an elasticity tensor, $\mathcal{G} = \{b_{ijkl}\}$ such that $b_{ijkl} = b_{klij} = b_{ijlk} \in L^\infty(\Omega)$ and $\mathcal{G}\tau \cdot \tau \geq 0$ for all $\tau \in \mathcal{S}_d$, a.e. in Ω .

The body forces, surface tractions and initial displacement satisfy

$\underline{H}(f)$: $f_0 \in L^2(Q; \mathbb{R}^d)$, $f_1 \in L^2(0, T; L^2(\Gamma_N; \mathbb{R}^d))$, $u_0 \in V$.

The superpotentials satisfy

$\underline{H}(j_\nu)$: $j_\nu: \Gamma_C \times (0, T) \times \mathbb{R} \rightarrow \mathbb{R}$ is a function such that
(i) $j_\nu(\cdot, \cdot, r)$ is measurable for all $r \in \mathbb{R}$, $j_\nu(\cdot, \cdot, 0) \in L^1(\Gamma_C \times (0, T))$;

(ii) $j_\nu(x, t, \cdot)$ is locally Lipschitz for a.e. $(x, t) \in \Gamma_C \times (0, T)$;

(iii) $|\partial j_\nu(x, t, r)| \leq c_\nu(1 + |r|)$ for a.e. $(x, t) \in \Gamma_C \times (0, T)$, all $r \in \mathbb{R}$ with $c_\nu > 0$.

$\underline{H}(j_\tau)$: $j_\tau: \Gamma_C \times (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a function such that

(i) $j_\tau(\cdot, \cdot, \xi)$ is measurable for all $\xi \in \mathbb{R}^d$, $j_\tau(\cdot, \cdot, 0) \in L^1(\Gamma_C \times (0, T))$;

(ii) $j_\tau(x, t, \cdot)$ is locally Lipschitz for a.e. $(x, t) \in \Gamma_C \times (0, T)$;

(iii) $\|\partial j_\tau(x, t, \xi)\|_{\mathbb{R}^d} \leq c_\tau(1 + \|\xi\|_{\mathbb{R}^d})$ for a.e. $(x, t) \in \Gamma_C \times (0, T)$, all $\xi \in \mathbb{R}^d$ with $c_\tau > 0$.

In the hypotheses $\underline{H}(\mathcal{C})$ and $\underline{H}(\mathcal{G})$, the inner product and the corresponding norm on \mathcal{S}_d are given by $\sigma \cdot \tau = \sigma_{ij}\tau_{ij}$ and $\|\sigma\|_{\mathcal{S}_d} = (\sigma \cdot \sigma)^{1/2}$ for all $\sigma, \tau \in \mathcal{S}_d$. In $\underline{H}(j_\nu)$ and $\underline{H}(j_\tau)$ the subdifferential is taken with respect to the last variables of j_ν and j_τ , respectively.

Next, we define the operators $A(t)$, $B \in \mathcal{L}(V, V^*)$ by

$$\begin{cases} \langle A(t)u, v \rangle = \langle \mathcal{C}(t)e(u), e(v) \rangle_{L^2(\Omega; \mathcal{S}_d)}, \\ \langle Bu, v \rangle = \langle \mathcal{G}e(u), e(v) \rangle_{L^2(\Omega; \mathcal{S}_d)} \end{cases} \quad (20)$$

for $t \in (0, T)$ and $u, v \in V$.

Using the equilibrium equation (2) and the Green formula (assuming the sufficient regularity of the functions involved) we obtain the weak formulation of the problem (2)–(8):

find $u: (0, T) \rightarrow V$ such that $u \in L^\infty(0, T; V)$, $u' \in \mathcal{V}$ and

$$\begin{cases} \int_{\Gamma_C} (j'_\nu(t, u_\nu; v_\nu) + j'_\tau(t, u_\tau; v_\tau)) \, d\Gamma \geq \\ \geq \langle f(t) - A(t)u'(t) - Bu(t), v \rangle \\ \text{a.e. } t \in (0, T), \text{ for all } v \in V \\ u(0) = u_0. \end{cases} \quad (21)$$

Analogously to Lemma 5 of [6], we have

LEMMA 8: Under the hypotheses $H(j_\nu)$ and $H(j_\tau)$, the functional $J: (0, T) \times X \rightarrow \mathbb{R}$ defined by

$$J(t, v) = \int_{\Gamma_C} (j_\nu(x, t, v_\nu(x)) + j_\tau(x, t, v_\tau(x))) d\Gamma \quad (22)$$

for a.e. $t \in (0, T)$, all $v \in X$ satisfies

- (i) $J(\cdot, v)$ is measurable for all $v \in X$ and $J(\cdot, 0) \in L^1(0, T)$;
- (ii) $J(t, \cdot)$ is Lipschitz on bounded subsets of X ;
- (iii) $\|\partial J(t, v)\|_{X^*} \leq \bar{c}(1 + \|v\|_X)$ for all $v \in X$, a.e. $t \in (0, T)$ with $\bar{c} = 2\sqrt{2} \max\{c_\nu, c_\tau\} \max\{1, \sqrt{m(\Gamma_C)}\}$;
- (iv) for all $v, w \in X$, we have

$$J^0(t, v; w) \leq \int_{\Gamma_C} (j_\nu^0(t, v_\nu; w_\nu) + j_\tau^0(t, v_\tau; w_\tau)) d\Gamma, \quad (23)$$

where $J^0(t, v; w)$ denotes the directional derivative of $J(t, \cdot)$ at a point $v \in X$ in the direction $w \in X$. If, in addition,

$$\begin{cases} \text{either } j_\nu(t, \cdot) \text{ or } -j_\nu(t, \cdot) \text{ is regular and} \\ \text{either } j_\tau(t, \cdot) \text{ or } -j_\tau(t, \cdot) \text{ is regular,} \end{cases} \quad (24)$$

then either $J(t, \cdot)$ or $-J(t, \cdot)$ is regular, respectively and (23) holds with equality.

Under our notation we associate with the hemivariational inequality (21) the following inclusion of type (10): find $u \in L^\infty(0, T; V)$ with $u' \in \mathcal{V}$ such that

$$\begin{cases} A(t)u'(t) + Bu(t) + \gamma^* \partial J(t, \gamma u(t)) \ni f(t) \text{ a.e. } t \\ u(0) = u_0. \end{cases} \quad (25)$$

REMARK 9: We notice that if the hypotheses $H(j_\nu)$ and $H(j_\tau)$ hold, then every solution to (25) is a solution to (21). The converse holds provided j_ν and j_τ satisfy the regularity condition (24). These facts follow from the definition of the Clarke subdifferential and Lemma 8 (iv).

The existence and regularity result for the hemivariational inequality (21) reads as follows.

THEOREM 10: If $H(\mathcal{C})$, $H(\mathcal{G})$, $H(f)$, $H(j_\nu)$, $H(j_\tau)$ hold and $m_1 > 2\bar{c}T\beta^2\|\gamma\|\max\{1, \|\gamma\|\}$, where \bar{c} is given in Lemma 8 (iii), $\beta > 0$ is the embedding constant of V into Z and $\|\gamma\| = \|\gamma\|_{\mathcal{L}(Z; X)}$, then problem (21) admits at least one solution $u \in L^\infty(0, T; V)$ with $u' \in \mathcal{V}$.

Proof. It follows from $H(\mathcal{C})$ and $H(\mathcal{G})$ that the operators $A(t)$ and B defined by (20) satisfy $H(A)$ with $\alpha = m_1$ and $H(B)$, respectively. It is a consequence of Lemma 8 (i) – (iii) that the functional J given by (22) satisfies $H(J)$. Also $H(\gamma)$ follows easily by the properties of the trace operator. The conclusion follows from Corollary 7 and Remark 9. \square

We conclude this section with short comments on multivalued boundary conditions (6) and (7) which are met in solid mechanics. The condition (6) is a generalization of a normal compliance condition to the nonmonotone setting. Let the function $j_\nu: \mathbb{R} \rightarrow \mathbb{R}$ be defined by $j_\nu(r) = \int_0^r p_\nu(s) ds$, where the function $p_\nu \in L^\infty_{loc}(\mathbb{R})$ is such that $|p_\nu(s)| \leq$

$p_1(1 + |s|)$ for $s \in \mathbb{R}$ with $p_1 > 0$ and $\lim_{r \rightarrow s^\pm} p_\nu(r)$ exist for all $s \in \mathbb{R}$. It is well known (see e.g. [9], [12]) that $\partial j_\nu(r) = \widehat{p}_\nu(r)$ for $r \in \mathbb{R}$, where the multivalued function $\widehat{p}_\nu: \mathbb{R} \rightarrow 2^{\mathbb{R}}$ is obtain by filling in the gaps. In this case $H(j_\nu)$ holds and (6) has the form $-\sigma_\nu \in \widehat{p}_\nu(u_\nu)$ on $\Gamma_C \times (0, T)$.

In the friction condition (7), if $j_\tau = 0$, we obtain the frictionless contact. If $j_\tau(x, t, \xi) = S(x, t)\|\xi\|$, where $S \in L^\infty(\Gamma_C \times (0, T))$ and $S > 0$ a.e., then we get a version of the static Tresca friction law where the friction bound depends on time. For other examples of functions j_ν and j_τ , we refer to [7]. Other two- and three-dimensional nonconvex superpotential laws are detailed in Section 4.6.1 of [9].

VI. CONCLUSION

In the paper we presented a method to study quasistatic hemivariational inequalities. The approach is natural and based on the asymptotic behavior as the mass coefficient in the inertial term of the evolution problem tends to zero. To our best knowledge the method is new and has not been considered in the literature. We also point out that the results of this paper can be applicable to a large class of quasistatic models of contact phenomena. It would be also interesting to extend the result of the paper to a class of problems with nonlinear viscosity and elasticity operators.

REFERENCES

- [1] J.P. Aubin and A. Cellina, *Differential Inclusions. Set-Valued Maps and Viability Theory*, Springer-Verlag, Berlin, New York, Tokyo (1984).
- [2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, Interscience, New York (1983).
- [3] Z. Denkowski, S. Migórski and N.S. Papageorgiou, *An Introduction to Nonlinear Analysis: Theory*, Kluwer Academic/Plenum Publishers, Boston, Dordrecht, London, New York (2003).
- [4] Z. Denkowski, S. Migórski and N.S. Papageorgiou, *An Introduction to Nonlinear Analysis: Applications*, Kluwer Academic/Plenum Publishers, Boston, Dordrecht, London, New York (2003).
- [5] W. Han and M. Sofonea, *Quasistatic Contact Problems in Viscoelasticity and Viscoplasticity*, American Mathematical Society, International Press, 2002.
- [6] S. Migórski, Dynamic hemivariational inequality modeling viscoelastic contact problem with normal damped response and friction, *Appl. Anal.*, 84 (2005), 669–699.
- [7] S. Migórski and A. Ochal, A unified approach to dynamic contact problems in viscoelasticity, *J. Elasticity*, 83 (2006), 247–275.
- [8] S. Migórski, A. Ochal and M. Sofonea, Integrodifferential hemivariational inequalities with applications to viscoelastic frictional contact, *Math. Models Methods in Applied Sci.*, 18 (2008), 271–290.
- [9] Z. Naniewicz and P. D. Panagiotopoulos, *Mathematical Theory of Hemivariational Inequalities and Applications*, Marcel Dekker, Inc., New York, Basel, Hong Kong (1995).
- [10] A. Ochal, Existence results for evolution hemivariational inequalities of second order, *Nonlinear Analysis*, 60 (2005), 1369–1391.
- [11] P. D. Panagiotopoulos, *Inequality Problems in Mechanics and Applications. Convex and Nonconvex Energy Functions*, Birkhäuser, Basel (1985).
- [12] P. D. Panagiotopoulos, *Hemivariational Inequalities, Applications in Mechanics and Engineering*, Springer-Verlag, Berlin (1993).
- [13] M. Shillor, M. Sofonea and J.J. Telega, *Models and Analysis of Quasistatic Contact*, Springer, Berlin (2004).
- [14] M. Sofonea, W. Han and M. Shillor, *Analysis and Approximation of Contact Problems with Adhesion or Damage*, Pure and Applied Mathematics, vol. 276, Chapman and Hall/CRC, Boca Raton (2006).

On an isotropic differential inclusion

Ana Cristina Barroso, Gisella Croce and Ana Margarida Ribeiro

Abstract—Differential inclusions arise in successful models proposed to describe the microstructures of elastic crystals. In this paper we are interested in the existence of Lipschitz maps $u : \Omega \rightarrow \mathbb{R}^2$ satisfying the inclusion

$$\begin{cases} Du \in E, & \text{a.e. in } \Omega \\ u = \varphi, & \text{on } \partial\Omega \end{cases}$$

where Ω is an open bounded subset of \mathbb{R}^2 and E is a compact subset of $\mathbb{R}^{2 \times 2}$, which is isotropic, that is to say, invariant under rotations. We will show an existence result under suitable hypotheses on the boundary datum φ .

Index Terms—Differential inclusion, isotropic set, singular values, rank one convexity.

I. INTRODUCTION

In the last twenty years successful models for studying the behaviour of crystal lattices undergoing solid-solid phase transitions have been studied. In such models it is assumed that the elements of crystal lattices have certain preferable affine deformations; this is true for example for martensite or for quartz crystals (see [1], [11]). This physical problem motivates the mathematical question of the existence of solutions to Dirichlet problems related to differential inclusions such as $Du \in E$ a.e. in Ω , where Ω is a domain of \mathbb{R}^n and $E \subset \mathbb{R}^{n \times n}$ is a compact set.

Two abstract theories to establish the existence of solutions of general differential inclusion problems are due to Dacorogna and Marcellini (see [5], [7]), whose result is based on Baire's category theorem, and Müller and Šverák [12], [13], who use ideas of convex integration by Gromov. In these two theories special notions of convexity are used. More precisely, the rank one convex hull of the set E , plays an important role. We say that a set $E \subseteq \mathbb{R}^{n \times n}$ is rank one convex if

$$A, B \in E, \text{rk}(A - B) = 1 \Rightarrow tA + (1 - t)B \in E, \forall t \in [0, 1].$$

Given a set $E \subseteq \mathbb{R}^{n \times n}$ its rank one convex hull, denoted by $\text{Rco } E$, is the smallest rank one convex set that contains E . We point out that we are following the notation used by Dacorogna and Marcellini in [7]; the rank one convex hull is denoted by $lc(E)$ by Müller and Šverák in [12]. The following characterization of $\text{Rco } E$ holds

$$\text{Rco } E = \bigcup_{i=0}^{\infty} \text{R}_i \text{co } E,$$

A.C. Barroso is with the Maths Dep. and CMAF, University of Lisbon, Lisbon, PORTUGAL (e-mail: barroso@ptmat.fc.ul.pt)

G. Croce is with the LMAH, Le Havre University, Le Havre, FRANCE (e-mail: gisella.croce@univ-lehavre.fr)

A.M. Ribeiro is with the Maths Dep. and CMA, FCT-Universidade Nova de Lisboa, Caparica, PORTUGAL (e-mail: amfr@fct.unl.pt)

where $\text{R}_0 \text{co } E = E$ and

$$\text{R}_{i+1} \text{co } E =$$

$$\{tA + (1 - t)B, t \in [0, 1], A, B \in \text{R}_i \text{co } E, \text{rk}(A - B) = 1\}.$$

Provided certain approximation properties hold, if the gradient of the boundary datum φ belongs to the interior of $\text{Rco } E$, then there exists a solution $u \in \varphi + W_0^{1,\infty}(\Omega, \mathbb{R}^n)$ to $Du \in E$ a.e. in Ω . However, the approximation properties are different in each of the two theories.

Using these abstract theorems various interesting problems related to the existence of microstructures have been solved, such as the two well problem, where $E = \mathcal{SO}(2)A \cup \mathcal{SO}(2)B$, where A and B are two fixed $\mathbb{R}^{2 \times 2}$ matrices (see [6], [7], [9], [11], [12]).

In this article we study the case where the set E is an arbitrary $\mathbb{R}^{2 \times 2}$ isotropic set, that is, invariant under rotations. More precisely, we assume that E is a compact subset of $\mathbb{R}^{2 \times 2}$ such that $AEB \subseteq E$ for every A, B in the orthogonal group $\mathcal{O}(2)$. Let Ω be an open bounded subset of \mathbb{R}^2 . We investigate the existence of weakly differentiable maps $u : \Omega \rightarrow \mathbb{R}^2$ that satisfy

$$\begin{cases} Du \in E, & \text{a.e. in } \Omega \\ u = \varphi, & \text{on } \partial\Omega. \end{cases} \quad (1)$$

Since E is isotropic it can be written as

$$E = \{\xi \in \mathbb{R}^{2 \times 2} : (\lambda_1(\xi), \lambda_2(\xi)) \in K\}, \quad (2)$$

for some compact set $K \subset \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq y\}$, where we have denoted by $0 \leq \lambda_1(\xi) \leq \lambda_2(\xi)$ the singular values of the matrix ξ , that is, the eigenvalues of the matrix $\sqrt{\xi \xi^t}$, which are

$$\begin{aligned} \lambda_1(\xi) &= \frac{1}{2} \left[\sqrt{\|\xi\|^2 + 2|\det(\xi)|} - \sqrt{\|\xi\|^2 - 2|\det(\xi)|} \right] \\ \lambda_2(\xi) &= \frac{1}{2} \left[\sqrt{\|\xi\|^2 + 2|\det(\xi)|} + \sqrt{\|\xi\|^2 - 2|\det(\xi)|} \right]. \end{aligned}$$

Thanks to the properties of the singular values (see [10]), problem (1) can be rewritten in the following equivalent way:

$$\begin{cases} \|Du(x)\|^2 = a^2 + b^2 & \text{a.e. in } \Omega, (a, b) \in K, \\ |\det Du(x)| = ab & \text{a.e. in } \Omega, (a, b) \in K, \\ u(x) = \varphi(x) & x \in \partial\Omega. \end{cases}$$

In the case where K consists of a unique point these two equations are the vectorial eikonal equation and the equation of prescribed absolute value of the Jacobian determinant.

The main result of our article is the following

Theorem 1.1: Let $E := \{\xi \in \mathbb{R}^{2 \times 2} : (\lambda_1(\xi), \lambda_2(\xi)) \in K\}$ where $K \subset \{(x, y) \in \mathbb{R}^2 : 0 < x \leq y\}$ is a compact set. Let $\Omega \subset \mathbb{R}^2$ be a bounded open set and let $\varphi \in C_{\text{piec}}^1(\bar{\Omega}, \mathbb{R}^2)$

be such that $D\varphi \in \text{int Rco}E$ in Ω . Then there exists a map $u \in \varphi + W_0^{1,\infty}(\Omega, \mathbb{R}^2)$ such that $Du \in E$ a.e. in Ω .

This result was first obtained by Croce in [4] using the theory developed by Dacorogna and Marcellini and a refinement due to Dacorogna and Pisante [8]. In this article we treat the same problem using the theory by Müller and Šverák which leads to different technical difficulties. We point out that in the case where K consists of a unique point and $K \subset \mathbb{R}^n$, $n \geq 2$ the same existence result was obtained by Dacorogna and Marcellini in [7].

We will use the following characterisation of the rank one convex hull of E due to Croce [3], [4]. Letting

$$f_\theta(x, y) := xy + \theta(y - x), \quad x > 0, y > 0, \theta \geq 0 \quad (3)$$

the following result holds.

Theorem 1.2: Let K be a compact set satisfying

$$K \subset \{(x, y) \in \mathbb{R}^2 : 0 < x \leq y\} \quad (4)$$

and let

$$E = \{\xi \in \mathbb{R}^{2 \times 2} : (\lambda_1(\xi), \lambda_2(\xi)) \in K\}. \quad (5)$$

Then $\text{Rco} E$ is the set of $\mathbb{R}^{2 \times 2}$ matrices ξ such that

$$f_\theta(\lambda_1(\xi), \lambda_2(\xi)) \leq \max_{(a,b) \in K} f_\theta(a, b), \quad \forall \theta \in [0, \max_{(a,b) \in K} b].$$

Moreover, $\text{int Rco} E$ is the set of $\mathbb{R}^{2 \times 2}$ matrices ξ such that

$$f_\theta(\lambda_1(\xi), \lambda_2(\xi)) < \max_{(a,b) \in K} f_\theta(a, b), \quad \forall \theta \in [0, \max_{(a,b) \in K} b].$$

II. IN-APPROXIMATION

To show theorem 1.1 we will use an existence result due to Müller and Šverák [12] which requires the following in-approximation property.

Definition 2.1: (In-approximation) Let E be a compact subset of $\mathbb{R}^{m \times n}$. We say that a sequence of sets $\{U_i\}$ is an in-approximation of E if

- 1) the sets U_i are open and contained in a fixed ball;
- 2) $U_i \subseteq \text{Rco}U_{i+1}$;
- 3) if $\xi_n \in U_n$ and $\xi_n \rightarrow \xi$, as $n \rightarrow \infty$, then $\xi \in E$.

In this section we will show that the set E , defined by (5) and (4), admits an in-approximation. Since a characterization of the rank one convex hull of an open isotropic set is not available, we will construct closed sets V_n from which we will obtain the open sets U_n of the in-approximation.

Definition 2.2: Let $\varepsilon_n = \frac{1}{n}$ and let r_n be a decreasing sequence such that $0 \leq r_n < \varepsilon_n$. For $(a, b) \in K$ we define the sets,

$$R_{(a,b)}^n = \{(x, y) \in \mathbb{R}^2 : a + \varepsilon_n - r_n \leq x \leq a + \varepsilon_n, \frac{ab - \varepsilon_n}{a + \varepsilon_n} - r_n \leq y \leq \frac{ab - \varepsilon_n}{a + \varepsilon_n}\}$$

and $V_n := \{\xi \in \mathbb{R}^{2 \times 2} : (\lambda_1(\xi), \lambda_2(\xi)) \in K_n\}$, where $K_n =$

$$\bigcup_{(a,b) \in K} R_{(a,b)}^n.$$

Proposition 2.3: The function $f_\theta(x, y)$ defined in (3) satisfies the following properties:

- i) f_θ is strictly increasing in y , for every $x > 0$ and $\theta \geq 0$;

- ii) f_θ is strictly increasing in x , for every $y > \theta$ and is strictly decreasing in x , for every $y < \theta$;

- iii) $f_\theta(\cdot, \theta)$ is constant, for every $\theta \geq 0$;

- iv) setting

$$\begin{aligned} \alpha_n^{(a,b)}(\theta) &= f_\theta\left(a + \varepsilon_n - r_n, \frac{ab - \varepsilon_n}{a + \varepsilon_n}\right) \\ \beta_n^{(a,b)}(\theta) &= f_\theta\left(a + \varepsilon_n, \frac{ab - \varepsilon_n}{a + \varepsilon_n}\right) \end{aligned}$$

one has

$$\max_{(x,y) \in R_{(a,b)}^n} f_\theta(x, y) = \max\left\{\alpha_n^{(a,b)}(\theta), \beta_n^{(a,b)}(\theta)\right\}$$

$$= \begin{cases} \beta_n^{(a,b)}(\theta), & \theta \in [0, \max_{(x,y) \in R_{(a,b)}^n} y] \\ \alpha_n^{(a,b)}(\theta), & \theta \geq \max_{(x,y) \in R_{(a,b)}^n} y; \end{cases}$$

- v) for every $\theta \in [\max_{(x,y) \in R_{(a,b)}^n} y, \max_{(x,y) \in R_{(a,b)}^{n+1}} y]$ the following inequality holds:

$$\alpha_n^{(a,b)}(\theta) < \beta_{n+1}^{(a,b)}(\theta);$$

for every $\theta \geq 0$ the following inequality holds:

$$\beta_n^{(a,b)}(\theta) < \beta_{n+1}^{(a,b)}(\theta);$$

- vi) assume that $\max_{(x,y) \in R_{(a,b)}^{n+1}} y < \max_{(x,y) \in R_{(a',b')}^{n+1}} y$, then for every

$$\theta \in [\max_{(x,y) \in R_{(a,b)}^{n+1}} y, \max_{(x,y) \in R_{(a',b')}^{n+1}} y]$$

$$\text{we have } \alpha_n^{(a,b)}(\theta) < \alpha_{n+1}^{(a',b')}(\theta);$$

- vii) for every $\theta \in [0, \max_{(x,y) \in R_{(a,b)}^{n+1}} y]$ the following inequality holds:

$$\max_{(x,y) \in R_{(a,b)}^n} f_\theta(x, y) < \max_{(x,y) \in R_{(a,b)}^{n+1}} f_\theta(x, y).$$

Proof: The first three properties are clear and the fourth one follows from *i)*, *ii)* and *iii)*. The second inequality in *v)* follows immediately from the fact that

$$\frac{ab - \varepsilon_n}{a + \varepsilon_n} < \frac{ab - \varepsilon_{n+1}}{a + \varepsilon_{n+1}}. \quad (6)$$

Due to the linearity in θ , it suffices to show the remaining inequalities in *v)* and *vi)* for θ belonging to the boundaries of the respective intervals. This is achieved using (6), *i)* and *iii)*. Finally, *vii)* is a consequence of *iv)* and *v)*. ■

Lemma 2.4: Let $b_M = \max_{(a,b) \in K} b$ and $a_M = \max_{(a,b_M) \in K} a$.

Then

$$\max_{(x,y) \in K_n} y = \frac{a_M b_M - \varepsilon_n}{a_M + \varepsilon_n},$$

for all sufficiently large n .

Proof: If $(a, b) \in K$ satisfies $a \leq a_M$ then it is easy to see that

$$\frac{ab - \varepsilon_n}{a + \varepsilon_n} \leq \frac{a_M b_M - \varepsilon_n}{a_M + \varepsilon_n}.$$

It remains to show the above inequality for points $(a, b) \in K$ such that $a \geq a_M$ and $b < b_M$. We argue by contradiction

and assume there exists a sequence $(a_n, b_n) \in K$ and $\varepsilon_{n'}$ a subsequence of ε_n such that

$$\frac{a_M b_M - \varepsilon_{n'}}{a_M + \varepsilon_{n'}} \leq \frac{a_n b_n - \varepsilon_{n'}}{a_n + \varepsilon_{n'}}.$$

Since K is compact and $(a_n, b_n) \in K$, up to a subsequence $(a_n, b_n) \rightarrow (a, b) \in K$, so passing to the limit in the above inequality we obtain $b \geq b_M$, contradicting $b < b_M$. ■

We will now prove the following proposition.

Proposition 2.5: The sets V_n introduced in Definition 2.2 satisfy $V_n \subseteq \text{int Rco} V_{n+1}$.

Proof: Due to the compactness of K and to the definition of the sets $R_{(a,b)}^n$, standard arguments show that the set K_n is compact. Therefore, by Theorem 1.2, it suffices to prove that for every $\theta \in [0, \max_{(x,y) \in K^{n+1}} y]$

$$\max_{(x,y) \in K_n} f_\theta(x, y) < \max_{(x,y) \in K_{n+1}} f_\theta(x, y).$$

Let $\theta \in [0, \max_{(x,y) \in K^{n+1}} y]$. Then there exists $(a, b) \in K$ (depending on θ) such that

$$\max_{(x,y) \in K_n} f_\theta(x, y) = \max_{(x,y) \in R_{(a,b)}^n} f_\theta(x, y).$$

Recall that, by Lemma 2.4,

$$\max_{(x,y) \in K^{n+1}} y = \frac{a_M b_M - \varepsilon_{n+1}}{a_M + \varepsilon_{n+1}} \geq \frac{ab - \varepsilon_{n+1}}{a + \varepsilon_{n+1}}.$$

If $0 \leq \theta \leq \frac{ab - \varepsilon_{n+1}}{a + \varepsilon_{n+1}}$ the result follows by property *vii*) of Proposition 2.3.

If $\theta \in \left[\frac{ab - \varepsilon_{n+1}}{a + \varepsilon_{n+1}}, \frac{a_M b_M - \varepsilon_{n+1}}{a_M + \varepsilon_{n+1}} \right]$ we have, by *iv*) and *vi*) of Proposition 2.3,

$$\begin{aligned} \max_{(x,y) \in K_n} f_\theta(x, y) &= \max_{(x,y) \in R_{(a,b)}^n} f_\theta(x, y) \\ &= \alpha_n^{(a,b)}(\theta) \\ &< \alpha_{n+1}^{(a_M, b_M)}(\theta) \\ &\leq \max_{(x,y) \in R_{(a_M, b_M)}^{n+1}} f_\theta(x, y) \\ &\leq \max_{(x,y) \in K_{n+1}} f_\theta(x, y). \end{aligned}$$

Theorem 2.6: Let E be given by (5) and (4). Then E admits an in-approximation.

Proof: Let V_n be the sets considered in the previous proposition and define the sets $U_n = \text{int} V_n$.

Step 1) U_n are open by definition and it is clear that the sequence $\{U_n\}$ is uniformly bounded.

Step 2) We will prove the second condition of the definition of in-approximation. Due to Proposition 2.5, we have $U_n \subset V_n \subseteq \text{int Rco} V_{n+1}$. We will show that, for every n ,

$$\text{int Rco} V_n \subseteq \text{Rco} U_n, \quad (7)$$

this will imply that $U_n \subset \text{Rco} U_{n+1}$, as required.

Let $\xi \in \text{int Rco} V_n$. We will prove that $\xi \in \text{Rco} \tilde{V}_n$, where \tilde{V}_n is the set defined by

$$\tilde{V}_n = \{\xi \in \mathbb{R}^{2 \times 2} : (\lambda_1(\xi), \lambda_2(\xi)) \in \tilde{K}_n\},$$

for a certain compact set $\tilde{K}_n \subset \text{int} K_n$. By continuity of the function $\xi \rightarrow (\lambda_1(\xi), \lambda_2(\xi))$ this will entail that $\tilde{V}_n \subseteq \text{int} V_n$, therefore

$$\xi \in \text{Rco} \tilde{V}_n \subseteq \text{Rco} \text{int} V_n = \text{Rco} U_n,$$

and we will have proved (7). For simplicity of notation we set $(\lambda_1(\xi), \lambda_2(\xi)) = (x, y)$. Our aim is thus to find a compact set $\tilde{K}_n \subset \text{int} K_n$ such that, for every $\theta \in [0, \max_{(a,b) \in \tilde{K}_n} b]$, the following inequality holds

$$f_\theta(x, y) \leq \max_{(a,b) \in \tilde{K}_n} f_\theta(a, b). \quad (8)$$

For $\lambda > 0$ define $\tilde{K}_n^\lambda = \bigcup_{(a,b) \in K} R_{(a,b)}^{n,\lambda}$ where

$$R_{(a,b)}^{n,\lambda} = \{(x, y) \in \mathbb{R}^2 : a + \varepsilon_n - r_n + \lambda \leq x \leq a + \varepsilon_n - \lambda, \frac{ab - \varepsilon_n}{a + \varepsilon_n} - r_n + \lambda \leq y \leq \frac{ab - \varepsilon_n}{a + \varepsilon_n} - \lambda\}.$$

It follows that $R_{(a,b)}^{n,\lambda} \subset \text{int} R_{(a,b)}^n \subset \text{int} K_n$ and so $\tilde{K}_n^\lambda \subset \text{int} K_n$. Since $\xi \in \text{int Rco} V_n$, we have that $f_\theta(x, y) < \max_{(a,b) \in K_n} f_\theta(a, b)$, so to show (8) it suffices to prove that, as $\lambda \rightarrow 0^+$,

$$\max_{(a,b) \in \tilde{K}_n^\lambda} f_\theta(a, b) \rightarrow \max_{(a,b) \in K_n} f_\theta(a, b), \quad (9)$$

uniformly with respect to θ . Notice that, as in the case of K_n ,

$$\max_{(x,y) \in \tilde{K}_n^\lambda} f_\theta(x, y) = \sup_{(a,b) \in K} \max \left\{ \alpha_{n,\lambda}^{(a,b)}(\theta), \beta_{n,\lambda}^{(a,b)}(\theta) \right\}$$

where

$$\begin{aligned} \alpha_{n,\lambda}^{(a,b)}(\theta) &= f_\theta \left(a + \varepsilon_n - r_n + \lambda, \frac{ab - \varepsilon_n}{a + \varepsilon_n} - \lambda \right) \\ \beta_{n,\lambda}^{(a,b)}(\theta) &= f_\theta \left(a + \varepsilon_n - \lambda, \frac{ab - \varepsilon_n}{a + \varepsilon_n} - \lambda \right). \end{aligned}$$

Hence (9) will follow from the fact that, as $\lambda \rightarrow 0^+$,

$$\begin{aligned} \sup_{(a,b) \in K} \alpha_{n,\lambda}^{(a,b)}(\theta) &\rightarrow \sup_{(a,b) \in K} \alpha_n^{(a,b)}(\theta) \\ \sup_{(a,b) \in K} \beta_{n,\lambda}^{(a,b)}(\theta) &\rightarrow \sup_{(a,b) \in K} \beta_n^{(a,b)}(\theta), \end{aligned} \quad (10)$$

uniformly in θ . As

$$\begin{aligned} |\alpha_{n,\lambda}^{(a,b)}(\theta) - \alpha_n^{(a,b)}(\theta)| &\leq \lambda \left(\max_{(a,b) \in K} a + 1 \right) + \lambda \frac{ab + \varepsilon_n}{a + \varepsilon_n} \\ &\quad + \lambda^2 + 2\lambda\theta \\ &\leq \lambda \left(\max_{(a,b) \in K} a + 1 \right) + \lambda^2 + 2\lambda\theta \\ &\quad + \lambda \max \left\{ 1, \max_{(a,b) \in K} b \right\} \end{aligned}$$

and this last expression tends to 0, uniformly with respect to θ and to (a, b) , we conclude the first statement of (10). A similar argument yields the second one.

Step 3) Given $(x_n, y_n) \in K_n$ there exists $(a_n, b_n) \in K$ such that $(x_n, y_n) \in R_{(a_n, b_n)}^n$. As K is compact, up to a subsequence, $(a_n, b_n) \rightarrow (a, b) \in K$, so, by the inequalities that define $R_{(a_n, b_n)}^n$, $(x_n, y_n) \rightarrow (a, b)$.

Let us now show the third condition of the definition of in-approximation. Assume that $\xi_n \in U_n$ and that $\xi_n \rightarrow \xi$. Since $\xi_n \in V_n$, $(\lambda_1(\xi_n), \lambda_2(\xi_n)) \in K_n$ so, by the above reasoning, $(\lambda_1(\xi_n), \lambda_2(\xi_n))$ converges to a point $(a, b) \in K$. On the other hand, by continuity, $\lambda_i(\xi_n) \rightarrow \lambda_i(\xi)$, $i = 1, 2$, and therefore $(\lambda_1(\xi), \lambda_2(\xi)) = (a, b) \in K$. Thus $\xi \in E$. ■

III. EXISTENCE THEOREM

We are going to prove Theorem 1.1. We will assume that the boundary datum φ is $C^1_{piec}(\bar{\Omega}, \mathbb{R}^2)$, that is to say, $\varphi \in W^{1,\infty}(\Omega, \mathbb{R}^2)$, there exist open sets $\omega_i \subset \Omega$ such that $\varphi \in C^1(\bar{\omega}_i, \mathbb{R}^2)$ and $\Omega \setminus \bigcup_i \omega_i$ is a set of Lebesgue measure zero.

We will use the following abstract theorem of Müller and Šverák [12]) to prove Theorem 1.1.

Theorem 3.1: Let $\Omega \subset \mathbb{R}^n$ be an open, bounded set and let E be a compact set which admits an in-approximation by the open sets U_i . Let $\varphi : \Omega \rightarrow \mathbb{R}^m$ be a C^1 function such that $D\varphi \in U_1$ in Ω . Then there exists $u \in \varphi + W^{1,\infty}_0(\Omega, \mathbb{R}^m)$ such that $Du \in E$ a.e. in Ω .

We begin by considering the case where φ is an affine function and to this effect we will need the following proposition.

Proposition 3.2: Let E be the set defined by (5) and (4) and let $\xi \in \text{int Rco } E$. Then there exists an in-approximation sequence U_n for E such that $\xi \in U_1$.

Proof: Consider the sequence of sets V_n defined in the previous section. We will show that there exists $N = N(\xi) \in \mathbb{N}$ such that

$$\xi \in \text{int Rco } V_N. \quad (11)$$

For simplicity of notation set $(\lambda_1(\xi), \lambda_2(\xi)) = (x, y)$. We must show that if

$$f_\theta(x, y) < \max_{(a,b) \in K} f_\theta(a, b), \quad \forall \theta \in [0, \max_{(a,b) \in K} b]$$

then there exists $N = N(x, y)$ such that, letting K_n be the sequence of sets in Definition 2.2,

$$f_\theta(x, y) < \max_{(a,b) \in K_n} f_\theta(a, b), \quad \forall \theta \in [0, \max_{(a,b) \in K_n} b]. \quad (12)$$

Since, by construction of K_n , $\max_{(a,b) \in K_n} b < \max_{(a,b) \in K} b$, it suffices to prove that

$$\max_{(a,b) \in K_n} f_\theta(a, b) \rightarrow \max_{(a,b) \in K} f_\theta(a, b), \quad n \rightarrow +\infty, \quad (13)$$

uniformly with respect to $\theta \in [0, \max_{(a,b) \in K} b]$. By Proposition 2.3, *iv*)

$$\max_{(a,b) \in K_n} f_\theta(a, b) = \sup_{(a,b) \in K} \max\{\alpha_n^{(a,b)}(\theta), \beta_n^{(a,b)}(\theta)\}$$

and

$$\begin{aligned} & \left| \sup_{(a,b) \in K} \max\{\alpha_n^{(a,b)}(\theta), \beta_n^{(a,b)}(\theta)\} - \max_{(a,b) \in K} f_\theta(a, b) \right| \\ & \leq \sup_{(a,b) \in K} |\max\{\alpha_n^{(a,b)}(\theta), \beta_n^{(a,b)}(\theta)\} - f_\theta(a, b)|. \end{aligned}$$

Therefore we must show that, as $n \rightarrow +\infty$,

$$|\alpha_n^{(a,b)}(\theta) - f_\theta(a, b)| \rightarrow 0, \quad |\beta_n^{(a,b)}(\theta) - f_\theta(a, b)| \rightarrow 0,$$

uniformly with respect to θ and to (a, b) . We start with the first limit. Letting

$$m_n = \frac{ab - \varepsilon_n}{a + \varepsilon_n} - a - \varepsilon_n + r_n, \quad q_n = (a + \varepsilon_n - r_n) \frac{ab - \varepsilon_n}{a + \varepsilon_n}$$

we have $\alpha_n^{(a,b)}(\theta) = m_n\theta + q_n$. Notice that $q_n - ab \rightarrow 0$ and $m_n - b + a \rightarrow 0$ uniformly with respect to (a, b) . This implies the result. The same reasoning applies to the second limit.

To complete the proof we notice that, letting $U_n = \text{int } V_n$, for every fixed $N \in \mathbb{N}$, the sequence

$$\text{Rco } U_N, U_{N+1}, U_{N+2}, \dots$$

is an in-approximation of E . Indeed the rank one convex hull of an open set U_N is open and rank one convex. Since, by construction, $U_N \subseteq \text{Rco } U_{N+1}$ and $\text{Rco } U_N$ is the smallest rank one convex set that contains U_N we conclude that $\text{Rco } U_N \subseteq \text{Rco } U_{N+1}$. Moreover, if $\xi \in \text{int Rco } E$ then $\xi \in \text{Rco } U_N$, by (11) and inclusion (7). ■

Theorem 3.3: Let Ω be an open, bounded subset of \mathbb{R}^2 and let E be the set defined by (5) and (4). Let $\xi \in \mathbb{R}^{2 \times 2}$ be such that $\xi \in \text{int Rco } E$ and let $\varphi : \Omega \rightarrow \mathbb{R}^2$ satisfy $D\varphi = \xi$ in Ω . Then there exists $u \in \varphi + W^{1,\infty}_0(\Omega, \mathbb{R}^2)$ such that $Du \in E$.

The proof of Theorem 3.3 follows immediately from the previous proposition and from Theorem 3.1. To obtain our existence result in the general case we will once again make use of Proposition 3.2 together with the following result, proved by Dacorogna and Marcellini in [7] (Corollary 10.15).

Theorem 3.4: Let Ω be an open subset of \mathbb{R}^n and A be an open subset of $\mathbb{R}^{m \times n}$. Let $\varphi \in C^1(\Omega, \mathbb{R}^m) \cap W^{1,\infty}(\Omega, \mathbb{R}^m)$ be such that

$$D\varphi(x) \in A, \quad \forall x \in \Omega.$$

Then there exists a function $v \in W^{1,\infty}(\Omega, \mathbb{R}^m)$ such that v is piecewise affine in Ω , $v = \varphi$ on $\partial\Omega$ and $Dv \in A$ a.e. in Ω .

We will now prove Theorem 1.1.

Proof: Assume first that $\varphi \in C^1(\bar{\Omega}, \mathbb{R}^2)$. We define the open set A as the set consisting of $\mathbb{R}^{2 \times 2}$ matrices ξ such that

$$f_\theta(\lambda_1(\xi), \lambda_2(\xi)) < \max_{(a,b) \in K} f_\theta(a, b), \quad \theta \in [0, \max_{(a,b) \in K} b].$$

We apply the previous theorem to φ and A , in order to obtain a map $v \in W^{1,\infty}(\Omega, \mathbb{R}^2)$ such that $v = \varphi$ on $\partial\Omega$, $Dv = c_i$ in Ω_i for some constant $c_i \in A$ and $\bigcup_i \Omega_i = \Omega$. Due to Theorem 3.3 we can solve the problem

$$\begin{cases} Du \in E, & \text{a.e. in } \Omega_i \\ u(x) = v(x), & x \in \partial\Omega_i \end{cases}$$

in each set Ω_i . Denoting by u_i the solution in Ω_i , the map defined by $u = u_i$ in Ω_i belongs to $\varphi + W^{1,\infty}_0(\Omega, \mathbb{R}^2)$ and satisfies $Du \in E$.

Now suppose that $\varphi \in C^1_{piec}(\bar{\Omega}, \mathbb{R}^2)$. This means that there exist open sets $\omega_i \subset \Omega$ such that $\varphi \in C^1(\bar{\omega}_i, \mathbb{R}^2)$ and $\Omega \setminus \bigcup_i \omega_i$ is a set of Lebesgue measure zero. By the first case, for each i , there exists $w_i \in \varphi + W^{1,\infty}_0(\omega_i, \mathbb{R}^2)$ such that $Dw_i \in E$ a.e. in ω_i . Thus, the function u defined as w_i in ω_i belongs to $\varphi + W^{1,\infty}_0(\Omega, \mathbb{R}^2)$ and satisfies $Du \in E$, a.e. in Ω . ■

We conclude this article by pointing out that Theorem 1.1 is not far from being optimal in the case where the boundary datum φ is affine. To explain this, we need some further notions of convexity given in [7].

Definition 3.5: A function $f : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R} \cup \{+\infty\}$ is polyconvex if there exists $g : \mathbb{R}^5 \rightarrow \mathbb{R} \cup \{+\infty\}$ convex such that $f(A) = g(A, \det(A))$.

A measurable function $f : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}$ is quasiconvex if

$$f(A) \leq \frac{1}{|\Omega|} \int_{\Omega} f(A + D\psi) dx$$

for every bounded domain Ω of \mathbb{R}^2 , for every $A \in \mathbb{R}^{2 \times 2}$ and for every $\psi \in W_0^{1,\infty}(\Omega, \mathbb{R}^2)$ ($|\Omega|$ stands for the Lebesgue measure of Ω).

A function $f : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R} \cup \{+\infty\}$ is rank one convex if $f(tA + (1-t)B) \leq tf(A) + (1-t)f(B)$ whenever $t \in [0, 1]$ and $\text{rk}(A - B) = 1$.

It is well known that, for $f : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}$,

$$f \text{ polyconvex} \Rightarrow f \text{ quasiconvex} \Rightarrow f \text{ rank one convex.}$$

Definition 3.6: A set $E \subseteq \mathbb{R}^{2 \times 2}$ is polyconvex if for all $t_i \geq 0$ with $\sum_{i=1}^5 t_i = 1$ and all $A_i \in E$ with

$$\sum_{i=1}^5 t_i \det A_i = \det \left(\sum_{i=1}^5 t_i A_i \right)$$

then $\sum_{i=1}^5 t_i A_i \in E$.

The polyconvex hull of a given set E is defined as the smallest polyconvex set that contains E .

Let $E \subset \mathbb{R}^{2 \times 2}$. Let \mathcal{P} be the set of polyconvex functions $f : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}$ such that $f|_E \leq 0$. We recall the following characterization of the closure of the polyconvex hull of E

$$\overline{\text{Pco}} E = \{\xi \in \mathbb{R}^{2 \times 2} : f(\xi) \leq 0, \forall f \in \mathcal{P}\}.$$

Now, suppose that u is a solution of

$$\begin{cases} Du \in E, & \text{a.e. in } \Omega \\ u = u_{\xi_0}, & \text{on } \partial\Omega \end{cases}$$

where u_{ξ_0} is an affine function with $Du_{\xi_0} = \xi_0$. Then there exists a map $\psi \in W_0^{1,\infty}(\Omega, \mathbb{R}^2)$ such that $u = u_{\xi_0} + \psi$. Let $f \in \mathcal{P}$. Then f is also quasiconvex and thus

$$f(\xi_0) \leq \frac{1}{|\Omega|} \int_{\Omega} f(\xi_0 + D\psi) dx = \frac{1}{|\Omega|} \int_{\Omega} f(Du) dx \leq 0$$

since $f|_E \leq 0$. This implies that $\xi_0 \in \overline{\text{Pco}} E$. In the case where E is an isotropic compact subset of $\mathbb{R}^{2 \times 2}$ is has been shown in [4] and [2] that $\text{Rco } E = \overline{\text{Rco}} E = \overline{\text{Pco}} E$. Therefore $\xi_0 \in \text{Rco } E$.

ACKNOWLEDGMENT

The research of Ana Cristina Barroso was partially supported by Fundação para a Ciência e Tecnologia, Financiamento Base 2008-ISFL/1/209. Ana Margarida Ribeiro was partially supported by Fundação para a Ciência e Tecnologia, Financiamento Base 2008-ISFL/1/297. Part of this work was completed during a visit of Gisella Croce to CMA, FCT-UNL, whose hospitality is gratefully acknowledged.

REFERENCES

- [1] Ball, J. M. and James, R. D., *Fine phase mixtures as minimizers of energy*, Arch. Rational Mech. Anal., **100** (1987), 13-52.
- [2] Cardaliaguet, P. and Tahraoui, R., *Equivalence between rank-one convexity and polyconvexity for isotropic sets of $\mathbb{R}^{2 \times 2}$, I*, Nonlinear Anal. **50** (2002), 1179-1199.
- [3] Croce, G., *Sur quelques inclusions différentielles de premier et de deuxième ordre*, PhD thesis, EPFL (2004).
- [4] Croce, G., *A differential inclusion: the case of an isotropic set*, ESAIM Control Optim. Calc. Var. **11** (2005), 122-138.
- [5] Dacorogna, B. and Marcellini, P., *General existence theorems for Hamilton-Jacobi equations in the scalar and vectorial cases*, Acta Math. **178** (1997), no. 1, 1-37.
- [6] Dacorogna, B. and Marcellini, P., *Cauchy-Dirichlet problem for first order nonlinear systems*, J. Funct. Anal. **152** (1998), no. 2, 404-446.
- [7] Dacorogna, B. and Marcellini, P., *Implicit partial differential equations*, Birkhäuser, Boston, 1999.
- [8] Dacorogna, B. and Pisante, G., *A general existence theorem for differential inclusions in the vector valued case*, Port. Math. **62** (2005), no. 4, 421-436.
- [9] Dolzmann, G. and Müller, S., *Microstructures with finite surface energy: the two-well problem*, Arch. Rational Mech. Anal. **132** (1995), no. 2, 101-141.
- [10] Horn, R.A. and Johnson, Ch. R., *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1991.
- [11] Müller, S., *Variational models for microstructure and phase transitions*, in Calculus of variations and geometric evolution problems (Cetraro, 1996), 85-210, Lecture Notes in Math., 1713, Springer, Berlin, 1999.
- [12] Müller, S. and Šverák, V., *Attainment results for the two-well problem by convex integration*, in Geometric analysis and the calculus of variations, 239-251, Internat. Press, Cambridge, MA, 1996.
- [13] Müller, S. and Šverák, V., *Convex integration for Lipschitz mappings and counterexamples to regularity*, Ann. of Math. (2), **157** (2003), no. 3, 715-742.

On some transmission problems in Hölder spaces

Ahmed MEDEGHRI

We consider some transmission problems written in the form of abstract elliptic equations in Hölder spaces completing in this way the work in L_p cases. Our approach makes use the concept of impedance operator which leads to obtain direct and simplified problems. We then use the Dunford calculus and some techniques in order to prove existence, uniqueness and maximal regularities results.

Coefficients de singularités géométriques pour des problèmes d'évolution

Mohand MOUSSAOUI

Il est bien connu que les problèmes posés dans des ouverts à coins présentent des singularités. Celles-ci se propagent dans le temps avec des coefficients dont on analysera la régularité.

Sturm-Liouville problems for a complete abstract second order differential equation of elliptic type in UMD spaces

Stéphane MAINGOT

We give some new results concerning a complete abstract differential equation of second order with general Robin boundary conditions. The study is developed in UMD spaces and uses the celebrated Dore-Venni Theorem. Existence, uniqueness and maximal regularity of the strict solution are proved: in fact, we furnish an explicit representation formula of the solution, by using Klein's reduction order method.

A general decay result in a viscoelastic Timoshenko system

Salim MESSAOUDI

The issue of stabilization of Timoshenko systems has attracted a great deal of researchers and several results concerning the uniform decay of solutions have been established. In this talk, we establish a generalized stability result for a wider class of relaxation functions.

On some transmission problems with boundary Dirichlet conditions

Fatimetou MINT AGHRABATT

In this communication, we will give some new results on a family of transmission problems between a thin layer and a fixed body. In particular existence, unicity and maximal regularity of solutions are proved by using the techniques of the abstract differential equations theory.

Complete abstract differential equations of elliptic type on the half-line : application of Dore-Venni and Da Prato-Grisvard sum theory in L_p -spaces

Amine ELTAIEF

We will present here some new results on complete abstract second order differential equations of elliptic type set in \mathbb{R}_+ . In the framework of UMD spaces, we use the celebrated Dore-Venni Theorem to prove existence and uniqueness for the strict solution. We will use also the Da Prato-Grisvard Sum Theory to furnish results when the space is not supposed to be UMD.

On some abstract fourth differential equation with transmission and boundary conditions

Hassan DIARAMOUNA SIDIBE

In this work we consider a transmission problem for the bilaplacian operator set in junction of two rectangular bodies: a fixed body and a thin layer. The study is performed in L_p -spaces. Then we prove that there exists a unique solution having a maximal regularity iff some compatibility conditions are verified. We finish by study completely the limit problem when the thickness of the thin layer tends to zero.

Chaotic Dynamics, Control and Applications of Complex Systems

Organizing Committee

- Miguel A. F. SANJUÁN
University Juan Carlos, Madrid, Spain
miguel.sanjuan@urjc.es

Description

The main goal of this special session is to join together physicists, mathematicians and other scientists interested in new developments of chaotic dynamics in its broader sense, the different available methods of controlling chaos and their applications in complex systems arising in many fields of engineering and sciences, and in particular to life sciences such as the analysis of the dynamics of neurons and genetic regulation networks.

The topics included, but not limited to, in this session are:

- New developments of chaotic dynamics
- Novel methods of controlling chaotic and complex dynamics
- Hamiltonian and dissipative chaotic systems
- Fractal structures in phase space
- Chaotic dynamics of neuronal and genetic models
- Control and synchronization in neuronal and genetic networks

Contents

Finding a realistic guaranteed-chaotic physical model is not easy	
James A. YORKE	181
New developments on partial control of chaotic systems	
Samuel ZAMBRANO , Miguel A. F. SANJUÁN	182
Riddled basins in complex physical and biological systems	
Ricardo Luiz VIANA	183
Synchronization of time varying networks	
R.E. AMRITKAR	184
A simple way of calculating the topological entropy for interval maps	
José AMIGÓ , Rui DILÃO	185
Chimera states in non-locally coupled phase oscillators with propagation delays	
Abhijit SEN	186
Phase control and synchronization in excitable systems	
Jesús M. SEOANE	187
Avoiding escapes in open dynamical systems using phase control	
Miguel A. F. SANJUÁN	188
Targeting synchronized response in chaotic oscillators	
Syamal Kumar DANA	189
Synchronization in chains of oscillators	
Ioan GROSU , M. HASLER , A. BIRZU	190
Simple discrete 3-dimensional stirring models	
Judy KENNEDY	191
How to compute safe sets with escape time sets	
Juan SABUCO , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	192
Point-vortex interaction in an oscillatory deformation field: Hamiltonian dynamics, harmonic resonance and transition to chaos	
Xavier PERROT , Xavier CARTON	193
Applying ordinal patterns to spatially extended systems	
José AMIGÓ , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	194
Control and synchronization of linearly coupled inertia ratchets with co-existing attractors	
U. E. VINCENT , I. O. OLUSOLA , D. MAYER	195
Control of noisy transient chaos in a discrete time continuous analog circuit	
Alexandre WAGEMAKERS , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	196
A Hindmarsh-Rose type silicon neuron	
Takashi KOHNO , Kazuyuki AIHARA	197
Estimation of the control parameter of a map through the analysis of its order patterns	
David ARROYO , Gonzalo ALVAREZ , José AMIGÓ	198
A pulsed-therapy of heterogeneous tumor model	
Abdelkader LAKMECHE	199
Partial control of the Duffing oscillator with fractal basins	
Mattia COCCOLO , Samuel ZAMBRANO , Miguel A. F. SANJUÁN	200
Anti-phase, in-phase synchronization, bifurcation and robustness of nonlinear interacting oscillators	
Rui DILÃO	201
A new type of orbits in the five body problem	
Arsen DZHANOEV , Alexander LOSKUTOV , Miguel A. F. SANJUÁN	202
Dynamical consequences of mesoscopic organization in complex systems: dynamical stability & synchronization in modular and hierarchical networks	
Sitabhra SINHA	203

**Finding a realistic guaranteed-chaotic physical model is not
easy**

James A Yorke

*Institute for Physical Sciences and Technology
University of Maryland
College Park, MD 20742
USA*

New developments on partial control of chaotic systems

Samuel Zambrano and Miguel A. F. Sanjuán

*Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

A chaotic saddle is a zero-measure repelling set where the dynamics is chaotic. This set typically appears due to the existence of a horseshoe in phase space, and it is responsible for the appearance of transient chaos. Sometimes it is desirable to sustain transient chaos, i.e. to keep trajectories close to the saddle. Thus, we recently proposed a technique called partial control [1,2] that allows one to keep the trajectories of a dynamical system close to the saddle even in presence of an environmental noise stronger than the applied control. In this talk we present new results concerning this control strategy that improve its applicability. We show that this technique can be applied even if we have just access to one of the system's parameters, and that the perturbation needed is typically smaller than with other control techniques. We also show that this “parametric partial control” can be applied even if we have access to just one of the system's variables. We also discuss the steps given towards a generalization of this technique to higher-dimensional dynamical systems.

[1] Samuel Zambrano, Miguel A. F. Sanjuán, and James A. Yorke, *Phys. Rev. E*, **77** 055201(R) (2008).

[2] Samuel Zambrano and Miguel A. F. Sanjuán, *Phys. Rev. E*, **79**, 026217 (2009).

Riddled basins in complex physical and biological systems ^{*}

Ricardo Luiz Viana [†]

Departamento de Física, Universidade Federal do Paraná,
Caixa Postal 19044, 81531-990, Curitiba, Paraná, Brazil

Complex systems have typically more than one attractor, either periodic or chaotic, and their basin structure ultimately determines the final-state predictability. When certain symmetries exist in the phase space, their basins of attraction may be riddled, which means that they are so densely intertwined that it may be virtually impossible to determine the final state, given a finite uncertainty in the determination of the initial conditions. Riddling occurs in a variety of complex systems of physical and biological interest ¹. We review the mathematical conditions for riddling to occur, and present two illustrative examples of this phenomenon: coupled Lorenz-like piecewise-linear maps ² and a deterministic model for competitive indeterminacy in populations of flour beetles ³

^{*}This work was partially supported by CNPq, CAPES, FAPESP, and Fundação Araucária (Brazilian government agencies).

[†]e-mail: viana@fisica.ufpr.br

¹M. A. F. Sanjuán, J. Aguirre, and R. L. Viana, *Fractal structures in nonlinear dynamics*, Rev. Mod. Phys. **81**, 333 (2009).

²M. C. Vergés, R. F. Pereira, S. R. Lopes, R. L. Viana, and T. Kapitaniak, *Riddling and chaotic synchronization of coupled piecewise-linear Lorenz maps*, Physica A **388**, 2515 (2009).

³R. F. Pereira, S. Camargo, S. E. de S. Pinto, S. R. Lopes, and R. L. Viana, *Periodic-orbit analysis and scaling laws of intermingled basins of attraction in an ecological dynamical system*, Phys. Rev. E **78**, 056214 (2008).

Synchronization of time varying Networks

R.E.Amritkar

*Physical Research Laboratory, Navarangpura, Ahmedabad 380009, India
e-mail: amritkar@prl.res.in*

Synchronization of dynamical networks has been investigated by many researchers with static connectivity link between the nodes. However, many networks with dynamical nodes change their connectivity with time. Synchronization properties of coupled dynamical systems on time-varying networks and also a comparison with time-average networks for a general class of coupling matrices, commutating as well as non-commutating cases has been discussed in this work. Some interesting relations of critical coupling constant for synchronization and switching times are derived which are verified by numerical simulations and real experiments in electronic circuits.

A simple way of calculating the topological entropy for interval maps

José Amigó¹ and Rui Dilão²

1) Operations Research Center, Miguel Hernández University,
Elche (Spain)

jm.amigo@umh.es

2) NonLinear Dynamics Group, IST

rui@sd.ist.utl.pt; ruidilao@gmail.com; <https://sd.ist.utl.pt>

Topological entropy as introduced in 1965 by Adler, Konheim and McAndrew, is an invariant of topological conjugacy for self-maps of an interval. Milnor and Thurston, in their classical paper on kneading invariants for maps of the interval, conjectured that the topological entropy of the logistic quadratic family of maps of the interval is a monotonically increasing function of a control parameter. Using techniques of complex analysis, in 1984 A. Douady proved the conjecture of Milnor for the quadratic map of an interval. However, for general families of maps of an interval, in general it is not known how to calculate the topological entropy neither to numerically estimate it efficiently. Here, with techniques developed in 1982 by Dias de Deus, Dilão and Taborda Duarte, we describe an algorithm that enables to determine the number of monotonous pieces of iterates of maps of an interval, enabling to calculate straightforwardly the topological entropy. As classical bifurcations are associated with the lack of topological conjugacy, we use topological entropy in order to predict the reversion of bifurcations and the reversion of the Sharkovskii ordering in families of maps of an interval.

Chimera states in non-locally coupled phase oscillators with propagation delays

Abhijit Sen

*Institute for Plasma Research, Bhat, Gandhinagar 382 428, India
e-mail: senabhijit@gmail.com*

Chimera states, where phase-locked and incoherent activity can simultaneously exist at different spatial locations are a novel collective mode of non-locally coupled systems and have potential applications in a number of physical, chemical and biological systems. Unlike other collective states such as phase-locked states or traveling waves, chimera states have not been investigated a great deal and there are many open questions regarding them. One such issue pertains to the effect of propagation delays on their existence and stability. We investigate this question on a system of coupled phase oscillators that includes distance dependent propagation delays. Our numerical simulations, carried out for a large number of coupled oscillators (ranging from 64 to 256), provide the first evidence of the existence of chimera states in a time-delayed system. Time delay breaks up the spatial pattern into several clusters of coherent regions interspersed with incoherent regions and the number and distribution of these clusters is a sensitive function of the propagation delay. A wide range of system parameters and delay values are explored to delineate the existence and stability regions of the chimera state. Our numerical results are complemented by solutions of appropriate self-consistency conditions that are derived analytically for the system.

Phase control and synchronization in excitable systems

Jesús M. Seoane

*Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

In this work we study excitable systems paying attention in some aspects concerning to their control and synchronization. Excitable systems are relevant in neuronal dynamics and therefore this method might have important applications. We use as prototype model the periodically driven FitzHugh-Nagumo (FHN) model, which displays both spiking and non-spiking behaviours in chaotic or periodic regimes. The phase control technique [1] consists of applying a harmonic perturbation with a suitable phase ϕ that we adjust in search of different behaviours of the FHN dynamics. We compare our numerical results with experimental measurements performed on an electronic circuit and find good agreement between them [2]. We also study the phenomenon of synchronization in uncoupled excitable systems due to common noise. We use as prototype model two identical FHN in presence of both, white and colored noise. We obtain, numerically and experimentally, a better synchronization insofar we increase the strength of that common noise [3]. We expect our work might be useful for a better understanding of excitable systems and synchronization phenomena in neuronal dynamics. This is joint work with S. Zambrano, Inés P. Mariño, Miguel A. F. Sanjuán (Spain) and S. Euzzor, A. Geltrude, K. Al Naimee, R. Meucci and F. T. Arecchi (Italy).

[1] S. Zambrano, E. Allaria, S. Brugioni, I. Leyva, R. Meucci, M.A.F. Sanjuán, and F. Arecchi, *Chaos* **16**, 013111 (2006).

[2] S. Zambrano, J. M. Seoane, Inés P. Mariño, Miguel A. F. Sanjuán, S. Euzzor, R. Meucci, and F.T. Arecchi, *New Journal of Physics* **9**, 073030 (2008).

[3] S. Zambrano, Inés P. Mariño, J. M. Seoane, Miguel A. F. Sanjuán, K. Al Naimee, A. Geltrude, S. Euzzor, R. Meucci, and F.T. Arecchi, *Phys. Rev. E* (To be submitted).

Avoiding escapes in open dynamical systems using phase control

Miguel A.F. Sanjuán

Nonlinear Dynamics, Chaos and Complex Systems Group

Departamento de Física

Universidad Rey Juan Carlos

Tulipán s/n,

28933 Móstoles, Madrid, Spain

A trajectory of an open dynamical system has the possibility to escape from a region in phase space where typically an initial condition is taken. In this talk, we analyze the problem of avoiding escapes in open dynamical systems by using a control method developed by the author and his collaborators which is called phase control [1,2,3]. For this purpose, we use as a prototype model the Helmholtz oscillator, which is the simplest nonlinear oscillator with escapes when the energy is typically above a threshold value. For some parameter values, this oscillator presents a critical value of the forcing for which all particles escape from its single well. By using the phase control technique, weakly changing the shape of the potential via a periodic perturbation of suitable phase ϕ , we avoid the escapes in different regions of the phase space. We provide numerical evidence, heuristic arguments, and an experimental implementation in an electronic circuit of this phenomenon. The ideas developed here are suitable to be applied for avoiding escapes in more complicated physical situations. This is joint work with J.M. Seoane and S. Zambrano (Spain) and S. Euzzor, R. Meucci and F. T. Arecchi (Italy).

References:

- [1] S. Zambrano, E. Allaria, S. Brugioni, I. Leyva, R. Meucci, M.A.F. Sanjuán, and F. Arecchi, *Chaos* 16, 013111 (2006).
- [2] S. Zambrano, I. P. Mariño, F. Salvadori, R. Meucci, M.A.F. Sanjuán, and F.T. Arecchi, *Phys. Rev. E* 74, 016202 (2006).
- [3] J. M. Seoane, S. Zambrano, S. Euzzor, R. Meucci, F.T. Arecchi, and M.A.F. Sanjuán, *Phys. Rev. E* 78, 016205 (2008).

Targeting synchronized response in chaotic oscillators

Syamal Kumar Dana

*Indian Institute of Chemical Biology
Jadavpur, Kolkata 700032, India
e-mail: skdana@iicb.res.in; sdana_ecsu@yahoo.com*

Defining an appropriate coupling function and thereby targeting a desired synchronized response in chaotic oscillators and its controlling is important for real applications. A general approach is described here to derive such a coupling function for any chaotic oscillator. For unidirectionally coupled chaotic oscillators, the coupling function is easy to define and then to realize complete synchronization, antisynchronization or generalized synchronization and even to induce an amplitude death as a response state. Defining such a coupling for mutual synchronization is also possible as shown in numerical simulations. Controlling such synchronized states is another important issue which is addressed here by inserting a control parameter in the definition of the coupling. The coupling is basically nonlinear and its complexity depends upon the order of nonlinearity that exists in the original system. However, a physical realization of the coupling function is not difficult as shown in real experiments using electronic circuits. The controllability of the synchronized states has potential application in digital encoding and others. The coupling is applicable for identical as well as mismatched chaotic systems where mismatch in parameters is a practical reality. In the process, we find an interesting route to synchronization with parameter mismatch that follows a new scaling law.

This approach of defining a coupling function is extended to induce a hybrid or mixed type synchronization where separate state variables can attain different form of synchronized states in two coupled chaotic oscillators: two similar state variables can be at complete synchronized state, another pair of state variables can attain antisynchronization state, even a third response can cease to oscillate. The results are further extended to induce synchronization and antisynchronization in dynamical networks.

SYNCHRONIZATION IN CHAINS OF OSCILLATORS

I.Grosu¹, M.Hasler² and A. Birzu³

1 Faculty of Bioengineering, University of Medicine and Pharmacy "Gr.T.Popa", Iasi, Romania

2 LANOS, EPFL, Lausanne, Switzerland

3 Adrian Birzu, Faculty of Chemistry, University "Al.I.Cuza", Iasi, Romania

Synchronization is a fascinating phenomenon in nature and a useful one in science and engineering. Using the OPCL(Open-Plus-Closed-Loop) method of control [1] we developed a method of synchronization both in master-slave (unidirectional) [2,3,6] and mutual (bidirectional) coupling [4,5]. These results are extended from two oscillators to several oscillators in a chain.

Let's consider a chain of N oscillators in a chain ($i=1,2,\dots,N$). For the sake of concreteness we work with FitzHugh-Nagumo model [7] :

$$\frac{du}{dt} = u - u^3/3 - w + I, \quad \frac{dw}{dt} = \phi(u+a-bw) \quad (1)$$

with $I=0, a=0.7, b=0.8, \phi=0.08$

We see that the model has one nonlinearity in the first equation. The proposed coupling for the oscillator "i" in the chain is :

$$\frac{du_i}{dt} = u_i - u_i^3/3 - w_i + I + (p-1+s_i^2)(u_i - s_i), \quad \frac{dw_i}{dt} = \phi(u_i + a - bw_i)$$

where s_i are calculated as certain averages among the neighbors of the oscillator "i" and $p < 0$. For unidirectional coupling we have :

$$s_1 = u_1, s_2 = (u_1 + u_2)/2, \dots, s_i = (u_{i-1} + u_i)/2, \dots, s_N = (u_{N-1} + u_N)/2$$

For bidirectional coupling we have :

$$s_1 = (u_1 + u_2)/2, s_2 = (u_1 + u_2 + u_3)/3, \dots, s_i = (u_{i-1} + u_i + u_{i+1})/3, \dots, s_N = (u_{N-1} + u_N)/2$$

In the following we present numerical results for $N=4$ oscillators and $N=128$ oscillators.

We see that the synchronization is achieved in both cases. We hope to use these results in investigating the transmission of information along the chains [8].

References

- [1] E. Atlee Jackson, I. Grosu, "An Open-Plus-Closed-Loop (OPCL) Control of Complex Dynamic Systems", *Physica D*, **85**, 1-9, 1995
- [2] I.Grosu, "Robust Synchronization", *Phys. Rev. E* 56,3709-3712,1997
- [3] A.I.Lerescu et al "Collection of master-slave synchronized chaotic systems", *Chaos, Solitons and Fractals* 22, 599-604, 2004
- [4] A.I.Lerescu, S.Oancea, I.Grosu, "Collection of Mutually Synchronized Chaotic Systems" *Physics Letters A*, 352, 222-226, 2006
- [5] I.Grosu, "General Coupling for Synchronization of 3 Identical Oscillators", *Int. J. of Bifurcation and Chaos* 17(10), 3519-22, 2007
- [6] I.Grosu, E. Padmanaban, Prodyot K.Roy and Syamal K.Dana, "Designing coupling for synchronization and amplification of chaos" *Phys. Rev. Lett.* 100, 234102 (2008)
- [7] E. De Lange, These No. 3617(2006), LANOS, EPFL, Lausanne, Switzerland
- [8] J. Pahle, A.K.Green, C.J.Dixon, U.Kummer "Information transfer in signaling pathways" *BMC Bioinformatics*, 9, 139, 2008.

Simple discrete 3-dimensional stirring models

Judy Kennedy

*Dept. of Mathematics
PO Box 10047
Lamar University
Beaumont, TX 77710
USA*

We investigate some simple 3-dimensional stirring models. The models are idealized and volume-preserving, and we believe they mimic stirring in tanks. We study them from both rigorous and computational viewpoints. This is joint work with Barry Peratt.

How to compute safe sets with escape time sets

Juan Sabuco, Samuel Zambrano and Miguel A.F. Sanjuán

*Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

The partial control technique [1,2] allows one to keep the trajectory of a dynamical system with a horseshoe inside a square where there is a horseshoe-like mapping (and thus close to the nonattractive chaotic set that is due to this kind of mapping) in presence of environmental noise. The main advantage of this technique is that by making use of certain zero-measure *safe sets*, this goal can be achieved even if the control applied is smaller than the effect of noise in the system. Here we explore the relation between these safe sets and the sets of points with different escape times that can be found in the square, the *escape time sets*. We show that the safe sets are a subset of those escape time sets. Furthermore, we give the criteria that allow one to select certain points of these escape time sets, the nonzero measure *extended safe sets*, that can play the role of the original safe sets in the partial control technique. We also analyze the advantages and the drawbacks of using these new “extended safe sets” instead of the original ones.

[1] Samuel Zambrano, Miguel A. F. Sanjuán, and James A. Yorke, Phys. Rev. E, 77 055201(R) (2008).

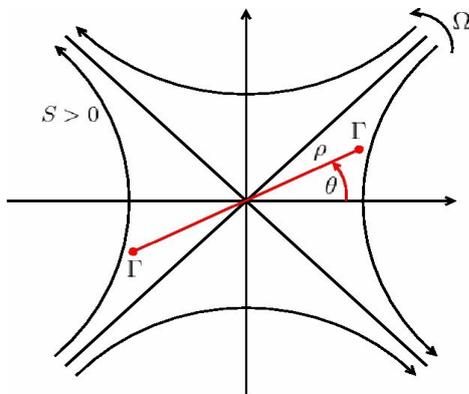
[2] Samuel Zambrano and Miguel A. F. Sanjuán, Phys. Rev. E, 79, 026217 (2009).

Point-vortex interaction in an oscillatory deformation field: Hamiltonian dynamics, harmonic resonance and transition to chaos

Xavier Perrot and Xavier Carton
LPO, UFR Sciences, UBO,
6 Av. Le Gorgeu, 29200 Brest, France
xcarton@univ-brest.fr

We study the Hamiltonian system of two point vortices, embedded in external strain and rotation (see figure 1). This external deformation field mimics the influence of neighboring vortices or currents in complex flows. When the external field is stationary, the equilibria of the two vortices, symmetric with respect to the center of the plane, are determined. The stability analysis indicates that two saddle points lie at the crossing of separatrices, which bound streamfunction lobes having neutral centers.

When the external field varies periodically with time, resonance becomes possible between the forcing and the oscillation of vortices around the neutral centers. A multiple time-scale expansion provides the slow-time evolution equation for these vortices, which, for weak periodic deformation, oscillate within their original (steady) trajectory. These analytical results accurately compare with numerical integration of the complete equations of motion. As the periodic deformation field increases, this vortex oscillation migrates out of the original trajectories, towards the location of the separatrices. With a periodic external field, these separatrices have given way to heteroclinic trajectories with multiple self-intersections, as shown by the calculation of the Melnikov function.



Chaos appears in vortex trajectories as they enter the aperiodic domain around the heteroclinic curves. In fact, this chaotic domain progressively fills out the plane, replacing KAM tori and cantori, as the periodic deformation field reaches finite amplitude. The appearance of windows of periodicity is illustrated (see figure 2).

Figure 1 ; Position of the two vortices and external flow

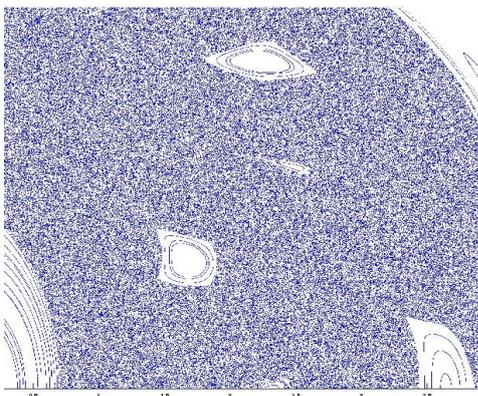


Figure 2 : Islands of periodicity inside the chaotic domain (Poincaré section of the flow at the forcing frequency)

Applying ordinal patterns to spatially extended systems

José M. Amigó¹, Samuel Zambrano² and Miguel A.F. Sanjuán²

¹*Universidad Miguel Hernández,
Elche, Spain*

²*Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

Permutation entropy was introduced in [1] as a complexity measure of time series. Permutation entropy replaces the probabilities of length L symbol blocks in the definition of Shannon's entropy by the probabilities of length- L ordinal patterns, a digest of the ups and downs of L consecutive elements of the time series. Since then permutation entropy and ordinal patterns have found a number of other interesting applications. One important challenge now is to expand the field of applications from time series and 'time' dependent dynamical systems, to spatially extended systems, so that one can also apply this approach to the study of space-time dynamics and space-time chaos. In order to tackle the viability of this program, the authors have studied two simple systems: *Cellular automata* (CA) and *coupled map lattices* (CML) [2], both in one space dimension. The formal similarity between such CA and CML is evident. If, furthermore, one discretizes the continuous dynamics of a CML, thus going over to the symbolic dynamics of the oscillators, then the similarity is almost complete. This allows eventually a unified treatment of both space-time systems. In our presentation we shall apply tools developed in the analysis of time series analysis via ordinal patterns [3], to (i) the classification of CA (ii) the phenomenology of CML, and (iii) the topological entropy calculation of both. The results are satisfactory and support the capabilities of our ordinal pattern-based approach in the study of spatially extended systems.

[1] C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* 88 ,174102. (2002).

[2] J.P. Crutchfield and K. Kaneko. Phenomenology of spatiotemporal chaos. In: *Directions in Chaos* (World Scinetific, Singapore). 1987.

[3] J.M. Amigó, S. Zambrano and M.A.F. Sanjuán. Combinatorial detection of determinism in noisy deterministic time series. *Europhys. Lett.* 83, 60005 (2008).

Control and Synchronization of linearly coupled Inertia Ratchets with co-existing attractors

U. E. Vincent^{1,2}, I. O. Olusola³ and D. Mayer¹

¹ *Institute of Theoretical Physics, Technical University of Clausthal Arnold-Sommerfeld Str. 6, 38678 Clausthal-Zellerfeld, Germany*

² *Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*

³ *Department of Physics, University of Agriculture, P. M. B. 2240 Abeokuta, Nigeria*

ABSTRACT

In the recent times, there has been renewed interest in transport phenomena and in particular directed transports of nonlinear non-equilibrium dynamical systems modelled by ratchet systems due their varieties of applications in several natural systems such as asymmetric crystals, semiconductor surfaces under light radiation, vortexes in Josephson junction arrays, micro-fluidic channels, and others. However, many outstanding problems regarding transport in ratchet systems remains un-resolved. One fundamental question is ‘How does the ratchet transport mechanism behave if two or more ratchets interact via a specific coupling?’ Recently, some attempts have been made to proffer possible answers through systematic investigations of the synchronization dynamics based on varieties of coupled ratchets. Notwithstanding, the transport properties of coupled ratchets in the multi-stable stable states are still far from well-understood. In this paper, we consider the dynamics of two elastically coupled inertia ratchets in a perturbed asymmetric potential. We investigate the coupled ratchets in the bi-stable states, where (i) two non- identical attractors, corresponding to a binary mixture of non-identical particles; and (ii) two identical attractors, corresponding to a binary mixture of identical particles co-exist in phase space - the dynamics being more complex unlike the mono-stable state - transporting currents in opposite directions. We show that the particle-particle interactions could lead to transports reversal in either direction as the strength of the interaction between the ratchets is progressively increased up to the fully synchronized state where optimal and enhanced transports could arise. Using Lyapunov stability theory and the linear matrix inequalities, we obtain sufficient criteria for full synchrony and hence, enhanced transport.

Acknowledgement: UEV is a Newton Fellow and also a Fellow of the Alexander von Humboldt Foundation. His work is supported the Alexander von Humboldt Foundations and the Royal Society of London.

Control of noisy transient chaos in a discrete time continuous analog circuit

Alexandre Wagemakers, Samuel Zambrano and Miguel A.F. Sanjuán

*Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

We present the analog circuit implementation of a novel method for the control of an unstable chaotic transient dynamics in presence of noise. The circuit simulates the dynamics of a discrete time tent map where the trajectories diverge to infinity for all the initial conditions. It is known that the presence of an unstable chaotic saddle in phase space causes all trajectories to diverge after a chaotic transient. A theoretical control of such class of systems in the presence of noise has been proposed in [1] and later generalized for higher dimensions in [2, 3]. With such a strategy the system can be maintained on a chaotic transient even when the noise applied exceeds the control. We present the analog circuit implementation of the control method applied to the discrete time tent map circuit. These encouraging results validates the theory and opens new perspectives for the application of the control technique in higher dimensions and continuous time dynamics.

[1] Jacobo Aguirre, Francesco d'Ovidio, and Miguel A. F. Sanjuán. Controlling chaotic transients: Yorke's Game of Survival. *Phys. Rev. E*, 69:016203, 2004.

[2] Samuel Zambrano, Miguel A. F. Sanjuán, and James A. Yorke. Partial Control of Chaotic Systems. *Phys. Rev. E*, 77:055201(R), 2008.

[3] Samuel Zambrano and Miguel A. F. Sanjuán. Exploring Partial Control of Chaotic Systems. *Phys. Rev. E*, 79:026217, 2009.

A Hindmarsh-Rose type silicon neuron

Takashi Kohno[†] and Kazuyuki Aihara[†]

[†] Institute of Industrial Science, The University of Tokyo, Meguro-ku, Tokyo 153-8505

Abstract

Silicon neuron is electrical circuit designed to reproduce electrophysiological functions of a neuronal cell. It is utilized not only as a tool for some studies on theoretical neuroscience that require real-time emulation of neural network but also as an element of neuromorphic hardware. Most of the silicon neuron circuits are based on either phenomenological or conductance-based models. The former abstracts some of various behaviors of a neuronal cell that seem to be important in information processing in nerve system. It does not take their mechanisms into account and may lack some critical properties of the neuronal cell. Because the model is highly abstracted, we can implement it with very simple circuit. The latter describes the dynamics of ionic channels on the membrane of a neuronal cell. It requires hard elaboration and large device resources to be implemented by electrical circuit not only because it is a set of complex nonlinear differential equations but also because there tends to be incompatibility between the characteristics of electrical device and lipid and protein.

We proposed a new type of silicon neuron [1] that is based on the results of mathematical analyses on conductance-based neuron models. They have been studied from the viewpoint of nonlinear dynamics to elucidate the essence of the mechanism lying behind various behaviors of neuronal cells. The analytical techniques including dimension reduction and phase-plane and bifurcation analyses successfully revealed the essence of various neuronal phenomena such as overshoot, threshold, refractoriness, and Hodgkin's classification. Our silicon neuron is designed to reproduce mathematical structures in phase plane and bifurcations of equilibria and limit cycles. We can use any curves that are easily realized by electrical devices to construct such mathematical structures instead of the curves described in the equations of conductance-based neuron models. This allows us to simplify circuitry maintaining dynamics in neuron models.

Based on previous works, we designed a burst silicon neuron circuit. It is three-dimensional ordinary differential equations of membrane potential, recovery variable, and very slow negative-feedback variable. It has the same topological structures in the phase plane and the bifurcation diagram as the square-wave burster such as Hindmarsh-Rose model [2] and an extended Morris-Lecar model [3]. The first two variables in our system compose a basic excitable system, which has bistability between an equilibrium and a limit cy-

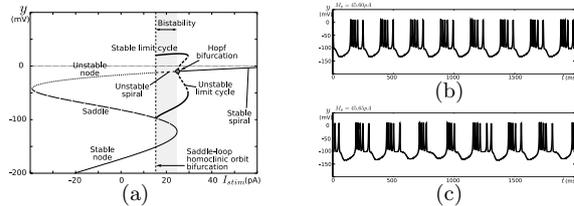


Figure 1: (a) Bifurcation diagram of the basic excitable system in our silicon neuron model. Examples of time waveform of membrane potential for (b) regular bursting and (c) chaotic bursting.

cle as shown in their bifurcation diagram (see Fig. 1 (a)). The last (very slow negative-feedback) variable makes the system to transfer back and forth between these two stable states, which leads to burst firing in the same way as the square-wave burster. In numerical simulation of our model, we observed tonic, chaotic, burst, and chaotic burst firings dependent on the scale of the last variable (see Fig. 1 (b) and (c)), which are quite similar to the other square-wave burster models. The right-hand side of our model are composed of the characteristic functions of differential pair circuits. These functions are modified hyperbolic tangent because our circuit is designed for metal-oxide semiconductor field-effect transistor (MOSFET) process that is driven in subthreshold operating condition. MOSFET consumes very low power in this condition and our circuit is estimated to consume about a few micro watts. In HSpice simulation of a layout mask data, we observed very similar patterns except for chaotic firing. We have already completed fabrication of our circuit by TSMC $.35\mu$ CMOS process and planning circuit experiments.

Keywords: Silicon neuron, Hindmarsh-Rose model, Burst neuron

References

- [1] Takashi Kohno and Kazuyuki Aihara, "A Design Method for Analog and Digital Silicon Neurons - Mathematical-Model-Based Method-," AIP Conference Proceedings, Vol. 1028, pp. 113–128, 2008.
- [2] J. Hindmarsh and R. Rose, "A model of neuronal bursting using three coupled first order differential equations," Proceedings of the Royal Society of London, Vol. B221, pp. 87–102, 1984.
- [3] J. Rinzel and B. Ermentrout, "Analysis of Neural Excitability and Oscillations," in "Methods in Neural Modeling", ed. C. Koch and I. Segev, pp. 251–291, MIT Press, 1998.

Estimation of the control parameter of a map through the analysis of its order patterns

David Arroyo,^{1,*} Gonzalo Alvarez,¹ and José María Amigó²

¹*Instituto de Física Aplicada, Consejo Superior de Investigaciones Científicas, Serrano 144—28006 Madrid, Spain*

²*Centro de Investigación Operativa, Universidad Miguel Hernández,
Avda. de la Universidad s/n, 03202 Elche, Spain*

The study of the order patterns associated to a map is very useful to establish how far its orbits are to be considered as source of white noise [1]. In this work we show that order patterns can also be used to infer an estimation of the parameter controlling the dynamics of certain maps.

Given a closed interval $I \subset \mathbb{R}$ and a map $f : I \rightarrow I$, the *orbit* of (the initial condition) $x \in I$ is defined as the set $\mathcal{O}_f(x) = \{f^n(x) : n \in \mathbb{N}_0\}$, where $\mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, \dots\}$, $f^0(x) = x$ and $f^n(x) = f(f^{n-1}(x))$. Orbits are used to define *order L -patterns* (or order patterns of length L), which are permutations of the elements $\{0, 1, \dots, L-1\}$, $L \geq 2$. We write $\pi = [\pi_0, \pi_1, \dots, \pi_{L-1}]$ for the permutation $0 \mapsto \pi_0, \dots, L-1 \mapsto \pi_{L-1}$.

Definition 1 (Order pattern). The point $x \in I$ is said to define (or realize) the order L -pattern $\pi = \pi(x) = [\pi_0, \pi_1, \dots, \pi_{L-1}]$ if

$$f^{\pi_0}(x) < f^{\pi_1}(x) < \dots < f^{\pi_{L-1}}(x). \quad (1)$$

Alternatively, x is said to be of type π . The set of all possible order patterns of length L is denoted by \mathcal{S}_L .

In general, if f_λ is a family of self-maps of the closed interval $I \subset \mathbb{R}$ parameterized by $\lambda \in J \subset \mathbb{R}$, and the set P_π is defined as

$$P_\pi = \{x \in I : x \text{ is of type } \pi\}, \quad (2)$$

where $\pi \in \mathcal{S}_L$, then P_π depends on f_λ and, consequently, on λ . Moreover, we will assume that f_λ is ergodic for $J \subset \mathbb{R}$ so as the orbits of f_λ can be used to build up statistics independently from the value of the initial condition. According to Birkhoff's ergodic theorem [2, p.34], if f_λ is ergodic with respect to the invariant measure μ , then the orbit of $x \in I$ visits the set P_π with relative frequency $\mu(P_\pi)$, for almost all x with respect to μ . As a result, it is possible to study the dependence of P_π on λ by counting and normalizing the occurrences of π in sliding windows of width L along $\mathcal{O}_{f_\lambda}(x)$, x being a 'typical' initial condition. Since we are primarily interested in the relation between the probabilities $\mu(P_\pi)$ (or relative frequencies) of order patterns $\pi \in \mathcal{S}_L$ and the control parameter λ of the map considered, we will refer to it as the λ -distribution function (in short: λ -DF) of π , considering their relation to the probability distribution functions (we fix π instead of fixing λ). If a map f_λ leads to λ -DF being a one-to-one or a few-to-one relation to the control parameter, then it is possible to get an estimation of λ . This is the case of the skew tent map defined as

$$f_\lambda(x) = \begin{cases} x/\lambda, & \text{if } 0 \leq x < \lambda, \\ (1-x)/(1-\lambda), & \text{if } \lambda \leq x \leq 1. \end{cases} \quad (3)$$

In this work we show that the rate of occurrences of the order pattern $[0, 1, \dots, L-1]$ in an orbit of the skew tent map leads to an estimation of λ . The methodology used for the skew tent map can be applied to any map with the ergodic property, a continuous invariant measure, and a bijective or quasi bijective λ -DF.

¹ J. M. Amigó, S. Zambrano and M. A. F. Sanjuán, "True and false forbidden patterns in deterministic and random dynamics," *Europhysics Letters* **79**, 50001–p1, –p5 (2007).

² P. Walters, *An Introduction to Ergodic Theory*, vol. 79 of *Graduate Texts in Mathematics* (Springer-Verlag, New York, 1982).

*Electronic address: david.arroyo@iec.csic.es

A pulsed-therapy of heterogeneous tumor model

Abdelkader Lakmeche

Département de mathématiques,
Faculté des sciences,
Université Djillali LIABES, BP. 89,
22000 Sidi Bel Abbès, Algeria

Abstract

One of the principal causes of the failure of chemotherapeutic treatment of cancer is its resistant development. There are, in general, two types of resistance, acquired resistance, which comes from cellular mutations, and induced resistance coming from the chemotherapeutic use. The two types of resistant tumors cells are physically completely different, and hence differently modeled. In this work, we are interested in induced resistance. We examine a nonlinear impulsive mathematical model, describing the dynamics of a heterogeneous tumor, constituted by two compartments, the sensitive cells and the drug resistance cells. We consider the case of use of several drugs with instantaneous effects described by impulses. We take into account the interactions between sensitive and drug resistance cells drug, described by terms contained in the nonlinear terms of the differential equation system. We are interested by the stability of the disease. The stability of the trivial solution correspond to the eradication of the disease, and loss of stability with bifurcation of nontrivial solutions, correspond to persistence of the disease.

References

- [1] A. Lakmeche and O. Arino, Nonlinear mathematical model of pulsed-therapy of heterogeneous tumors, *Nonlinear Analysis, Real World Applications* 2 (2001) 455-465.
- [2] A. Lakmeche and O. Arino, Bifurcations of nontrivial periodic solutions of impulsive differential equations arising from chemotherapeutic treatment, *Dynamics Cont. Discrete Impl. Systems* 7 (2000) 2, 265-288.
- [3] J. C. Panetta, A mathematical model of drug resistant: Heterogeneous tumors, *Math. Biosc.* 147 (1998) 41-61.

Partial control of the Duffing Oscillator with fractal basins

Mattia Coccolo, Samuel Zambrano and Miguel A.F. Sanjuán

*Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

Recently a technique has been proposed to keep the trajectories of a dynamical system with transient chaos close to a chaotic saddle and far from coexisting attractors. This technique is the partial control technique [1], and it can be used when there is a horseshoe-like map in phase space. On the other hand, the existence of fractal basins of attraction in a dynamical system can be related in some cases to the existence of this type of mapping [2]. Taking the Duffing oscillator as a paradigm [3], we show here how the partial control technique can be implemented on a continuous-time dynamical system with different coexisting attractors whose basins are fractalized. It is also shown how this technique allows one to keep the trajectories far from those attractors, even in presence of a noise stronger than the applied control.

[1] Samuel Zambrano, Miguel A. F. Sanjuán, and James A. Yorke, *Phys. Rev. E*, 77 055201(R) (2008).

[2] Jacobo Aguirre, Ricardo L. Viana, and Miguel A. F. Sanjuán. *Fractal Structures in Nonlinear Dynamics. Rev. Mod. Phys.*, 81(1):333-386 (2009).

[3] Jacobo Aguirre and Miguel A. F. Sanjuán. *Unpredictable behavior in the Duffing oscillator: Wada basins. Physica D*, 171:4151 (2002).

Anti-phase, in-phase synchronization, bifurcation and robustness of nonlinear interacting oscillators

Rui Dilão

NonLinear Dynamics Group, IST

rui@sd.ist.utl.pt; ruidilao@gmail.com; <https://sd.ist.utl.pt>

We propose and analyze a new interaction mechanism between oscillators leading to exact anti-phase and in-phase synchronization of pendulum clocks, and we determine a sufficient condition for the existence of an exact anti-phase synchronous state. We show that exact anti-phase and in-phase synchronous states can coexist in phase space, and the periods of the synchronous states are different from the eigen-periods of the individual oscillators. We analyze the robustness of the system when the parameters of the individual pendulum clocks are varied, and we show numerically that exact anti-phase and in-phase synchronous states exist in systems of coupled oscillators with different individual parameters.

A new type of orbits in the five body problem

Arsen Dzhanoev¹, Alexander Loskutov¹ and Miguel A.F. Sanjuán²

¹ *Moscow State University, Physics Faculty,
Leninskie Gory, Moscow, 119992, Russia*

² *Nonlinear Dynamics, Chaos and Complex Systems Group
Departamento de Física
Universidad Rey Juan Carlos
Tulipán s/n,
28933 Móstoles, Madrid, Spain*

On the basis of the three-body problem a new type of orbit in the five body problem is constructed. It is analytically shown that along with the well known chaotic and regular orbits in the five-body problem there also exists a qualitatively different type of orbit which we call “*stabilized*”. To show that we consider the Sitnikov problem [1] that consists of two equal masses M (called primaries) moving in circular or elliptic orbits about their common center of mass and a third, test mass μ moving along the straight line passing through the centre of mass normal to the orbital plane of the primaries.

Using the Melnikov method the existence of transverse homoclinic orbits could be shown. For the Sitnikov problem it was proved [2] that for all but a finite number of values of the eccentricity e the system is non-integrable, i. e. chaotic. We consider only small values of e . Hence due to the KAM-theory [3], since our system has $3/2$ degrees of freedom the invariant tori bound the phase space and chaotic motion is finite and takes place in a small vicinity of a separatrix layer.

The main objective of our work is to show through analytical and numerical methods the existence of the stabilized orbits in this special case of five-body problem. In general, this is related to the stabilization and control of unstable and chaotic behaviour of dynamical systems by external forces [4, 5]. Recently [5] for a case that is useful for many physical applications we have shown that the dynamics of the system with the split separatrices (chaos) can be regularized by a series of “kicks”.

In summary, on the basis of the elliptic Sitnikov problem we constructed a configuration of five bodies which we called the extended Sitnikov problem and analytically showed that in this configuration along with chaotic and regular orbits a new type of orbit (stabilized) could be realized.

[1] K. A. Sitnikov, Dokl. Akad. Nauk USSR 133, No2, 303 (1960).

[2] H. Dankowicz and Ph. Holmes, J. Differential Equations 116, 468 (1995).

[3] V. I. Arnold, Usp. Math. Nauk 18, 91 (1963).

[4] V. V. Alexeev and A. Loskutov, Sov. Phys.-Dokl. 32, 270 (1987).

[5] A. R. Dzhanoev, A. Loskutov H. Cao and M.A.F.Sanjuán: DCDS-B 7, 275 (2007).

Dynamical Consequences of Mesoscopic Organization in Complex Systems: Dynamical Stability & Synchronization in Modular and Hierarchical Networks

Sitabhra Sinha

*Institute of Mathematical Sciences, Chennai, India
e-mail: sitabhra@imsc.res.in*

Mesoscopic organization, e.g., into modules and hierarchical structures, is ubiquitous among complex networks, occurring in systems as diverse as cellular networks involved in metabolism and signaling, to cortico-cortical networks, human society, food webs and the internet. We show that many of the empirically observed "small-world" networks seen, e.g., in the brain, may owe their properties to a modular topological structure. We demonstrate that their dynamics (including spin-ordering, diffusion and synchronization behaviors) may differ significantly from the conventional models of small-world networks. We also investigate the dynamical implications of hierarchical ordering, and, the existence of modules, by looking at how the stability of arbitrary equilibria and synchronized states changes as a function of increasing hierarchy and modularity. Further, we explore the reasons behind why so many networks evolve such mesoscopic organization by noting that such systems are often subject to multiple structural and functional constraints. We show that hierarchy and modularity can emerge as a result of simultaneously satisfying several conflicting constraints.

State Space of a Rotating Rotor

V.M. Sokol, Israeli Independent Academy for Development of Sciences

Abstract

The method of continuous complex systemic measurement of dynamic parameters of a rotating rotor (taking into account nonlinearity of its characteristics) is offered. The method provides identification (calculation) of instant positions of a symmetry axis and a rotation axis of rotor by results of measurements of sensors system, and also the subsequent systemic measurement and calculation of mechanical parameters of a rotor in discrete time points during each rotor revolution and during all work cycle [1, 2]. At that, instant magnitudes of such parameters, as instant magnitude of angular speed, amplitude of radial and angular oscillations, damping coefficient, mass eccentricity and unbalance parameters, the inertia moment and inertia tensor of a rotor, the mechanical characteristic (dependence of the rotating moment on angular speed), resistance to rotation, rotor diameter and non-round form of its shaft, and also parameters of vibration of a rotor [3 - 9] may be identified in real time on the basis of the fundamental equations of a rotor movement.

Aggregate of the measured instant magnitudes of the enumerated parameters of a rotor may be considered as a multidimensional vector describing dot mappings of a rotor state [10].

Multiple recurrence of measurement and calculation of rotor parameters during each revolution and during all work cycle allow to receive sequence of dot mappings and to identify state space of a rotor. The analytical description of movement of a rotor within the differential equations may be received with the help of Poincare sections.

Identification of dot mappings and state spaces of a rotor is a basis for carrying out (in real time) the bifurcation analyses and detection of rare attractors.

Index Terms: measurement of mechanical parameters, dot mappings, bifurcation analysis, rare attractors.

References

- [1]. V.M. Sokol, "To a Question on Identification of Stochastic Mechanical Parameters of a Rotor," *Proc. of Institute for Advanced Studies*, Issue 5, Arad: Publishing IASA, 2005, pp. 49 – 55.
- [2]. V.M. Sokol, "Complex System of Continuous Measurement of Dynamic Parameters of Rotors," *Proc. of the International Scientific Conf. "The Modern Achievements of Science and Education"*, Khmelnytsky (Ukraine) – Netania (Israel), 2007, pp. 96-101.
- [3]. V.M. Sokol, "Mathematical Identification Methods of Dynamic Parameters and Characteristics of Rotors Systems," *Proc. of Institute for Advanced Studies*, Issue 6, Arad: Publishing IASA, 2006, pp. 19 – 34.
- [4]. V.M. Sokol, "Determination of Rotors Moment of Inertia During its Balancing," *Automation and modern technologies*, 1995, 7, pp. 22-24.
- [5]. V.M. Sokol, "Inertia Moment and Oscillation of a Statically Unbalanced Rotor," *Proc. of Institute for Advanced Studies*, Issue 3, Arad: Publishing IASA, 2003, pp. 7-31.
- [6]. V.M. Sokol, "Inertia Tensor and Oscillations of Dynamically Unbalanced Rotor," *Proc. of Institute for Advanced Studies*, Issue 4, Arad: Publishing IASA, 2004, pp. 13 – 20.
- [7]. V.M. Sokol, "Non-contact Dynamometry of Rotating Rotors (Mathematical Model)," *Proc. of VI International Conf. on Improvement of Quality, Reliability and Long Usage of Technical Systems and Technological Processes*, Hurghada (Egypt), 2007, pp. 61-65.
- [8]. V.M. Sokol, "Parameters Measurement of the Unbalance of Rotating Rotors During Their Operation," *Proc. of the 2-th International Scientific Conf. "The Modern Achievements of Science and Education"*, Netania (Israel), 2008, pp. 6-8.
- [9]. V.M. Sokol, "Measurement Method of the Rotor Vibration," *Academic transactions of Israeli Independent Academy for Development of Sciences*, 2009, in press.
- [10]. V.M. Sokol, "Method of Dot Mappings in Measurement of Dynamic Parameters of Rotary Systems," *Journal of Vibroengineering*, 2008, 10(4), pp. 451- 455.

Vladimir M. Sokol was born in Vinnitsa (Ukraine) in 1938. He is a graduate of physical faculty of Vinnitsa pedagogical university. He has defended a doctor's degree (Ph D) at Khmelnytsky national university. Dynamics of rotors systems and also continuous systemic parameters measurement of rotating rotors are the areas of his scientific interests.

He was the head of science at the special design office "Mechatronics" and (later) head of science of a researches centre "Pallada" at Vinnitsa technical university. He is leading scientist at Institute for Advanced Studies in Arad after migration into Israel at 2000. He is the author more than 120 scientific publications.

Dr Sokol is the academician of the Ukrainian Technological Academy (since 1997) and the Israeli Independent Academy for Development of Sciences (since 2008).

Chaos and Synchronization of Fractional-Order Systems

Organizing Committee

- Zaid M. ODIBAT
University of Le Havre, France
z.odibat@gmail.com

Description

The development of models based on fractional-order differential systems has recently gained popularity in the investigation of dynamical systems. Fractional derivatives provide an excellent instrument for the description of memory and hereditary properties of various materials and processes. The real objects of objects of the fractional-order systems are that we have more degrees of freedom in the model and that a “memory” is included in the model. Recently, the chaotic dynamics of fractional-order systems began to attract much attention in recent years. It has been shown that the fractional-order systems, as generalizations of many well-known systems, can also behave chaotically, such as the fractional Chen, Chua, Rossler, Lorenz and Arneodo systems. Furthermore, recent work based on numerical simulations show that chaotic fractional-order systems can also be synchronized. The main topics of interest for this session include:

- Chaos in fractional order systems
- Synchronization of fractional order systems
- Control in fractional order systems
- Fractional order systems modeling
- Numerical methods for fractional order systems

Contents

A note on chaos control and synchronization of fractional order systems Zaid M. ODIBAT	208
A special kind of synchronization of different chaotic discrete-time systems Yong CHEN , Yiliang JIN , Xin LI	213
Non-standard discretization of fractional differential equations G. Hussian ERJAEI , Modi ALNASR , Shaher MOMANI	218
Complexified dynamical systems from real fractional actionlike with time-dependent fractional exponent on multi-fractals sets Ahmad Rami EL-NABULSI	223
Designing modified projective synchronization for fractional order chaotic systems Luo RUNZI , Deng SHUCHENG , Wei ZHENGMIN	228
On some stability conditions and hyperchaos synchronization in the new fractional order hyperchaotic Chen system Ahmed E. MATOUK	232
Backstepping control of fractional-order chaotic systems E. NASERI , A. RANJBAR , S.H. HOSSEINIA , Shaher MOMANI	237
Control of Genesio-Tesi and Chen chaotic systems using a fractional-order controller M. MAHMOUDIAN , A. RANJBAR , E. NASERI , S.H. HOSSEINIA , Shaher MOMANI	242
Synchronization of chaotic fractional-order Couillet system via ASMC M. SHAHIRI T. , A. RANJBAR , R. GHADERI , S.H. HOSSEINIA , Shaher MOMANI	247
Dynamical behaviors, linear feedback control and synchronization of the fractional order Liu system Ahmed E. MATOUK	253
Chaotic synchronization of fractional-order Chua's system with time-varying delays Shangbo ZHOU , Xiaoran LIN , Hua LI	258
Solution of the fractional diffusion equation with absorbent term and external force Subir DAS	264
Solution of vibration equation by homotopy analysis method Ms. S. CHAKRABORTY	268

A Note on Chaos Control and Synchronization of Fractional Order Systems

Zaid M. Odibat*

Faculty of Science and Technology, University of Le Havre
76058 Le Havre Cedex, France
z.odibat@gmail.

Abstract—The dynamics of fractional order systems have attracted a great deal of attentions in recent years. In this paper, we study the chaos synchronization of two identical fractional order systems with a suitable feedback controller applied to the response system. Based on the stability results of linear fractional order systems, sufficient conditions for chaos synchronization are derived. The numerical results show that fractional order chaotic systems can be synchronized.

Index Terms—Chaos synchronization, Fractional order system, Caputo fractional derivative, Stability, Control.

I. INTRODUCTION

THE development of models based on fractional order differential systems has recently gained popularity in the investigation of dynamical systems. Fractional derivatives provide an excellent instrument for the description of memory and hereditary properties of various materials and processes. The main reason for using the integer-order models was the absence of solution methods for fractional differential equations. The advantages of the real objects of the fractional order systems are that we have more degrees of freedom in the model and that a “memory” is included in the model.

Recently, the chaotic dynamics of fractional order systems has been investigated and studied in mathematical and physical communities in the last few decades. The research efforts have been devoted to the chaos control and chaos synchronization problems in many dynamical fractional order systems. It has been shown that the fractional order systems, as generalizations of many well-known systems, can also behave chaotically, such as the fractional Duffing system [1], the fractional Chua system [2,3], the fractional Rössler system [4], the fractional Chen system [5-7], the fractional Lorenz system [8], the fractional Arneodo’s system [9] and the fractional Lü system [10]. In [3-6] it has been shown that some fractional-order systems can produce chaotic attractors with order less than 3.

Recent studies show that chaotic fractional order systems can also be synchronized. In many literatures, synchronization among fractional order systems is only investigated through numerical simulations. A simple method for chaos synchronization of fractional order systems based on the stability criteria of linear differential systems is presented in [11-13].

*On sabbatical leave from Prince Abdullah Bin Ghazi Faculty of Science and IT, Al-Balqa’ Applied University, Salt-Jordan

Numerical algorithms for chaos synchronization of fractional-order systems based on Laplace transform theory are presented in [14-17]. In the present paper, using the master-slave synchronization scheme and based on the stability results of linear fractional order systems, we study the synchronization of two coupled fractional order chaotic systems.

There are several definitions of a fractional derivative of order $\alpha > 0$ [18-20]. In this work, Caputo fractional derivative is considered. The fractional differential operator in the sense of Caputo [21] is defined as,

$$D^\alpha f(t) = J^{m-\alpha} D^m f(t). \quad (1)$$

Here D^m is the usual integer differential operator of order m , $m - 1 < \alpha \leq m$, and J^μ is the Riemann-Liouville integral operator of order $\mu > 0$, defined by

$$J^\mu f(t) = \frac{1}{\Gamma(\mu)} \int_0^t (t - \tau)^{\mu-1} f(\tau) d\tau, \quad t > 0. \quad (2)$$

Caputo’s definition has the advantage of dealing properly with initial value problems in which the initial conditions are given in terms of the field variables and their integer order which is the case in most physical processes.

II. STABILITY ANALYSIS

Stability analysis of fractional order systems, which is of main interest in control theory, has been thoroughly investigated where necessary and sufficient conditions have been derived [22-25] (see also references therein). In this section, we recall the main stability results. For this object, we consider the following n dimensional fractional order system,

$$\begin{cases} \frac{d^{\alpha_1} x_1}{dt^{\alpha_1}} = f_1(x_1, x_2, \dots, x_n), \\ \frac{d^{\alpha_2} x_2}{dt^{\alpha_2}} = f_2(x_1, x_2, \dots, x_n), \\ \vdots \\ \frac{d^{\alpha_n} x_n}{dt^{\alpha_n}} = f_n(x_1, x_2, \dots, x_n), \end{cases} \quad (3)$$

where α_i is a rational number between 0 and 1 and $\frac{d^{\alpha_i}}{dt^{\alpha_i}}$ is the Caputo fractional derivative of order α_i , for $i = 1, 2, \dots, n$. Assume that $\alpha_i = k_i/m_i$, $(k_i, m_i) = 1$, $k_i, m_i \in \mathbb{N}$, for $i = 1, 2, \dots, n$. Let m be the least common multiple of the denominators m_i ’s of α_i ’s.

First, if the system (3) is a linear system, that is $[f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})]^T = [a_{ij}]_{i,j=1}^n \mathbf{x} = A\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$, then we have the following results:

- If $\alpha = \alpha_1 = \alpha_2 = \dots = \alpha_n$, then the fractional order system (3) is asymptotically stable iff $|\arg(\text{spec}(A))| > \alpha\pi/2$. In this case the components of the state decay towards 0 like $t^{-\alpha}$ [22].
- If α_i 's are rational numbers between 0 and 1, then the system (3) is asymptotically stable if all roots λ of the equation $\det(\text{diag}(\lambda^{m\alpha_1}, \lambda^{m\alpha_2}, \dots, \lambda^{m\alpha_n}) - A) = 0$ satisfy $|\arg(\lambda)| > \gamma\pi/2$, where $\gamma = 1/m$ [23].

Second, if function f_i has second continuous partial derivatives in a ball centered at an equilibrium point $\mathbf{x}^* = (x_1, x_2, \dots, x_n)$, that is $f_i(x_1, x_2, \dots, x_n) = 0$, for $i = 1, 2, \dots, n$, then we have the following results:

- If $\alpha = \alpha_1 = \alpha_2 = \dots = \alpha_n$, then the equilibrium point \mathbf{x}^* of system (3) is asymptotically stable if $|\arg(\text{spec}(J|_{\mathbf{x}^*}))| > \alpha\pi/2$, where the matrix J is the Jacobian matrix of the system (3) that is defined as $J = [\frac{\partial f_i}{\partial x_j}]_{i,j=1}^n$ [24].
- If α_i 's are rational numbers between 0 and 1, then the equilibrium point \mathbf{x}^* of system (3) is asymptotically stable if all roots λ of the equation $\det(\text{diag}(\lambda^{m\alpha_1}, \lambda^{m\alpha_2}, \dots, \lambda^{m\alpha_n}) - J|_{\mathbf{x}^*}) = 0$ satisfy $|\arg(\lambda)| > \gamma\pi/2$, where $\gamma = 1/m$ [25].

The previous stability results play an important role in studying the existence of chaotic attractors and the synchronization between fractional order systems.

III. SYNCHRONIZATION OF FRACTIONAL ORDER SYSTEMS

Chaotic attractors for fractional order 3D systems has been demonstrated based on numerical simulation results. In [26], based on stability results, a necessary condition for fractional order systems to exhibit chaotic attractors similar to its integer order counterpart is presented. This necessary condition is equivalent to,

$$\min_i \{|\arg(\lambda_i)|\} \leq \gamma\pi/2, \quad (4)$$

where $\gamma = 1/m$ and λ_i 's are the roots of the equation $\det(\text{diag}(\lambda^{m\alpha_1}, \lambda^{m\alpha_2}, \lambda^{m\alpha_3}) - J|_{\mathbf{x}^*}) = 0$, for every equilibrium point \mathbf{x}^* . Otherwise, one of these equilibrium points becomes asymptotically stable and attracts the nearby trajectories.

Now, based on the stability results of fractional differential equations, we briefly discuss the issue of controlling fractional order chaotic systems to realize synchronization with error feedback control. In order to observe synchronization behavior, we construct the master (drive) system and the slave (response) system as,

$$M : \begin{cases} \frac{d^{\alpha_1} x_m}{dt^{\alpha_1}} = f_1(x_m, y_m, z_m) + g_1(x_m, y_m, z_m), \\ \frac{d^{\alpha_2} y_m}{dt^{\alpha_2}} = f_2(x_m, y_m, z_m) + g_2(x_m, y_m, z_m), \\ \frac{d^{\alpha_3} z_m}{dt^{\alpha_3}} = f_3(x_m, y_m, z_m) + g_3(x_m, y_m, z_m), \end{cases} \quad (5)$$

$$S : \begin{cases} \frac{d^{\alpha_1} x_s}{dt^{\alpha_1}} = f_1(x_s, y_s, z_s) + g_1(x_m, y_m, z_m) + u_1(t), \\ \frac{d^{\alpha_2} y_s}{dt^{\alpha_2}} = f_2(x_s, y_s, z_s) + g_2(x_m, y_m, z_m) + u_2(t), \\ \frac{d^{\alpha_3} z_s}{dt^{\alpha_3}} = f_3(x_s, y_s, z_s) + g_3(x_m, y_m, z_m) + u_3(t), \end{cases} \quad (6)$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ indicates the fractional orders, $\frac{d^{\alpha_i}}{dt^{\alpha_i}}$ is the fractional differential operator in Caputo sense, $0 < \alpha_i \leq 1$, f_i is a linear function and g_i is a nonlinear function for $i = 1, 2, 3$. Subscripts m and s stand for the master system and slave system, respectively, $\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t)]^T$ is the controller to be designed such that these two chaotic systems can achieve synchronization. Defining the synchronization error as $e_1(t) = x_s(t) - x_m(t)$, $e_2(t) = y_s(t) - y_m(t)$, $e_3(t) = z_s(t) - z_m(t)$, then we get the error system,

$$\begin{cases} \frac{d^{\alpha_1} e_1}{dt^{\alpha_1}} = f_1(e_1, e_2, e_3) + u_1(t), \\ \frac{d^{\alpha_2} e_2}{dt^{\alpha_2}} = f_2(e_1, e_2, e_3) + u_2(t), \\ \frac{d^{\alpha_3} e_3}{dt^{\alpha_3}} = f_3(e_1, e_2, e_3) + u_3(t). \end{cases} \quad (7)$$

Our aim is to determine the controller $\mathbf{u}(t)$,

$$\mathbf{u}(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{pmatrix} = \begin{pmatrix} k_1 e_1 \\ k_2 e_2 \\ k_3 e_3 \end{pmatrix}, \quad (8)$$

where $k_1, k_2, k_3 \in \mathbb{R}$, such that the drive system (5) and the response system (6) are synchronized ($\|\mathbf{e}(t)\| \rightarrow 0$, as $t \rightarrow +\infty$).

An idea of synchronization is to use a controller to make the output of the slave system copy in some manner the master system one. Obviously, the synchronization between (5) and (6) is equivalent to the asymptotical stability of the zero solution (zero equilibrium point) to error system (7).

According to the stability results, the drive system (5) and the response system (6) are synchronized if all roots λ of the characteristic equation $\det(\text{diag}(\lambda^{r_1}, \lambda^{r_2}, \lambda^{r_3}) - J) = 0$ satisfy $|\arg(\lambda)| > \gamma\pi/2$, where $\gamma = 1/m$, $r_i = m\alpha_i$, $i = 1, 2, 3$, and J is the Jacobian matrix for the error system (7) at the origin.

To demonstrate this technique, some examples of synchronization for two identical fractional order chaotic systems are discussed in the following sections.

IV. FRACTIONAL CHEN SYSTEM

Chen and Ueta [27] introduced, in 1999, the chaotic Chen system which is similar but not topologically equivalent to the Lorenz system. The fractional version of Chen system reads as,

$$\begin{cases} \frac{d^{\alpha_1} x}{dt^{\alpha_1}} = a(y - x), \\ \frac{d^{\alpha_2} y}{dt^{\alpha_2}} = (c - a)x - xz + cy, \\ \frac{d^{\alpha_3} z}{dt^{\alpha_3}} = xy - bz, \end{cases} \quad (9)$$

where $0 < \alpha_1, \alpha_2, \alpha_3 \leq 1$. Integer order Chen system displays chaotic attractors, for example, when $(a, b, c) = (35, 3, 28)$. Simulations are performed to obtain chaotic behavior of the fractional order Chen system for different fractional orders α when $(a, b, c) = (35, 3, 28)$. For example, chaotic attractors are found in [12] when $\alpha = (0.95, 0.95, 0.95)$. In [6] chaotic behaviors are found when $\alpha = (0.9, 0.9, 0.9)$. Moreover, in [14] and [26], it is found that for the parameters $\alpha = (0.985, 0.99, 0.98)$ and $\alpha = (0.8, 1, 0.9)$, respectively, the fractional order Chen system can display chaotic attractors. According to our approach, the master and the slave fractional order Chen systems are described as,

$$M : \begin{cases} \frac{d^{\alpha_1} x_m}{dt^{\alpha_1}} = a(y_m - x_m), \\ \frac{d^{\alpha_2} y_m}{dt^{\alpha_2}} = (c - a)x_m - x_m z_m + c y_m, \\ \frac{d^{\alpha_3} z_m}{dt^{\alpha_3}} = x_m y_m - b z_m, \end{cases} \quad (10)$$

$$S : \begin{cases} \frac{d^{\alpha_1} x_s}{dt^{\alpha_1}} = a(y_s - x_s) + u_1(t), \\ \frac{d^{\alpha_2} y_s}{dt^{\alpha_2}} = (c - a)x_s - x_m z_m + c y_s + u_2(t), \\ \frac{d^{\alpha_3} z_s}{dt^{\alpha_3}} = x_m y_m - b z_s + u_3(t), \end{cases} \quad (11)$$

where $u_i(t) = k_i e_i$, $i = 1, 2, 3$. Then the error system gives,

$$\begin{cases} \frac{d^{\alpha_1} e_1}{dt^{\alpha_1}} = a(e_2 - e_1) + k_1 e_1, \\ \frac{d^{\alpha_2} e_2}{dt^{\alpha_2}} = (c - a)e_1 + c e_2 + k_2 e_2, \\ \frac{d^{\alpha_3} e_3}{dt^{\alpha_3}} = -b e_3 + k_3 e_3. \end{cases} \quad (12)$$

The Jacobian matrix for the error system (12) is

$$J = \begin{pmatrix} -a + k_1 & a & 0 \\ c - a & c + k_2 & 0 \\ 0 & 0 & -b + k_3 \end{pmatrix}, \quad (13)$$

and so, the characteristic equation $\det(\text{diag}(\lambda^{\alpha_1}, \lambda^{\alpha_2}, \lambda^{\alpha_3}) - J) = 0$ can be written as,

$$((\lambda^{\alpha_1} + a - k_1)(\lambda^{\alpha_2} - c - k_2) + a(a - c))(\lambda^{\alpha_3} + b - k_3) = 0. \quad (14)$$

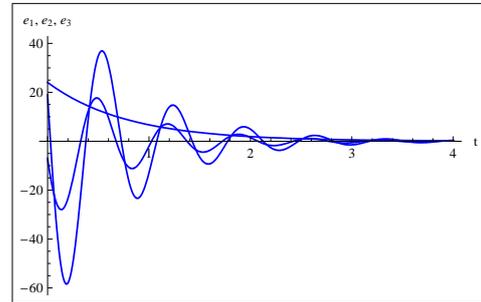
Now, in case of $(a, b, c) = (35, 3, 28)$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, the systems (9) and (10) are synchronized if k_1, k_2 and k_3 satisfy the control laws,

$$\begin{cases} k_1 + k_2 - 7 \mp w < 0, & \text{if } A = w \in \mathbb{R} \\ \left| \frac{w}{k_1 + k_2 - 7} \right| < \tan(\alpha\pi/2), & \text{if } A = iw \in i\mathbb{R} \end{cases} \quad (15)$$

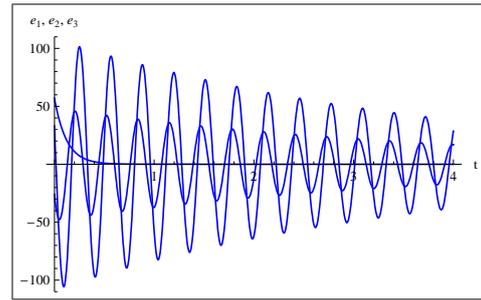
$$k_3 < 3, \quad (16)$$

where $A = ((7 - k_1 - k_2)^2 - 4(245 - (35 - k_1)(28 + k_2)))^{1/2}$.

For example, when $(a, b, c) = (35, 3, 28)$, taking $\alpha = (0.95, 0.95, 0.95)$ and $(k_1, k_2, k_3) = (20, -15, 2)$ or $\alpha = (0.9, 0.9, 0.9)$ and $(k_1, k_2, k_3) = (25, -18, -2)$, then in both cases the controller $\mathbf{u}(t) = (k_1, k_2, k_3)\mathbf{e}(t)$ satisfies the control



(a) $\alpha = (0.95, 0.95, 0.95)$ and $(k_1, k_2, k_3) = (20, -15, 2)$



(b) $\alpha = (0.9, 0.9, 0.9)$ and $(k_1, k_2, k_3) = (25, -18, -2)$

Fig. 1. Synchronization of the fractional order Chen system (9).

laws (15) and (16). Therefore, the drive system (10) and the response system (11) are synchronized. The error functions evolution, in these cases, is shown in Fig. 1. The eigenvalues for the Jacobian matrix in the first case are $\lambda_1 = -1 + 7i$, $\lambda_2 = -1 - 7i$ and $\lambda_3 = -1$, and in the second case are $\lambda_1 = 12.04159i$, $\lambda_2 = -12.04159i$ and $\lambda_3 = -5$.

From Fig. 1, it is obvious that the components of the error system (12) decay towards zero as $t \rightarrow +\infty$. So, we can numerically conclude that the designed controller, $\mathbf{u}(t) = (20, -15, 2)\mathbf{e}(t)$ in the first case or $\mathbf{u}(t) = (25, -18, -2)\mathbf{e}(t)$ in the second case, can effectively control the chaotic fractional order Chen system to achieve synchronization between the drive system (10) and the response system (11).

Also, in case of $(a, b, c) = (35, 3, 28)$ and $\alpha = (0.985, 0.99, 0.98)$, the systems (9) and (10) are synchronized if the roots of the equation,

$$((\lambda^{197} + 35 - k_1)(\lambda^{198} - 28 - k_2) + 245)(\lambda^{196} + 3 - k_3) = 0, \quad (17)$$

lie in the region $|\arg(\lambda)| > \pi/400$. For example, if we take $(k_1, k_2, k_3) = (35, -28, 2)$, then we can show that all roots of Eq. (17), that becomes $(\lambda^{395} + 245)(\lambda^{196} + 1) = 0$, lie in the region $|\arg(\lambda)| > \pi/400$. Therefore, the drive system (10) and the response system (11) are synchronized. The error functions evolution, in this case, is shown in Fig. 2. It is clear, from Fig. 2, that the components of the error system (12) decay towards zero as $t \rightarrow +\infty$. So, we can numerically conclude that the designed controller, $\mathbf{u}(t) = (35, -28, 2)\mathbf{e}(t)$, can effectively control the chaotic fractional order Chen system to achieve synchronization between the drive system (10) and the response system (11).

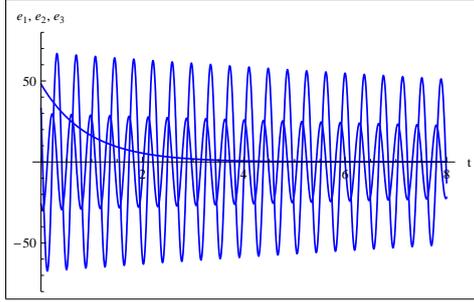


Fig. 2. Synchronization of the fractional order Chen system (9), when $\alpha = (0.985, 0.99, 0.98)$ and $(k_1, k_2, k_3) = (35, -28, 2)$.

V. FRACTIONAL CHUA SYSTEM

Now, we consider the Chua system, which is found by Chua [28], [29] in 1986. The fractional version of Chua system reads as,

$$\begin{cases} \frac{d^{\alpha_1} x}{dt^{\alpha_1}} = a(y - x - f(x)), \\ \frac{d^{\alpha_2} y}{dt^{\alpha_2}} = x - y + z, \\ \frac{d^{\alpha_3} z}{dt^{\alpha_3}} = -by - cz, \end{cases} \quad (18)$$

where $f(x) = m_1 x + \frac{1}{2}(m_0 - m_1)(|x + 1| - |x - 1|)$, $0 < \alpha_1, \alpha_2, \alpha_3 \leq 1$. Simulations are performed to obtain chaotic behavior of fractional order Chua system for different fractional orders α . For example, chaotic attractors are found in [12] when $\alpha = (0.98, 0.98, 0.94)$ and $(a, b, c, m_0, m_1) = (10.1911, 10.3035, 0.1631, -1.1126, -0.8692)$. Moreover, in [6] it is found that for the parameters $\alpha = (0.93, 0.99, 0.92)$ and $(a, b, c, m_0, m_1) = (10.725, 10.593, 0.268, -1.1726, -0.7872)$, the fractional order Chua system can display chaotic attractors.

According to our approach, the master and the slave fractional order Chua systems are described as,

$$M : \begin{cases} \frac{d^{\alpha_1} x_m}{dt^{\alpha_1}} = a(y_m - x_m - f(x_m)), \\ \frac{d^{\alpha_2} y_m}{dt^{\alpha_2}} = x_m - y_m + z_m, \\ \frac{d^{\alpha_3} z_m}{dt^{\alpha_3}} = -by_m - cz_m, \end{cases} \quad (19)$$

$$S : \begin{cases} \frac{d^{\alpha_1} x_s}{dt^{\alpha_1}} = a(y_s - x_s - f(x_m)) + u_1(t), \\ \frac{d^{\alpha_2} y_s}{dt^{\alpha_2}} = x_s - y_s + z_s + u_2(t), \\ \frac{d^{\alpha_3} z_s}{dt^{\alpha_3}} = -by_s - cz_s + u_3(t), \end{cases} \quad (20)$$

where $u_i(t) = k_i e_i$, $i = 1, 2, 3$. Then the error system gives,

$$\begin{cases} \frac{d^{\alpha_1} e_1}{dt^{\alpha_1}} = a(e_2 - e_1) + k_1 e_1, \\ \frac{d^{\alpha_2} e_2}{dt^{\alpha_2}} = e_1 - e_2 + e_3 + k_2 e_2, \\ \frac{d^{\alpha_3} e_3}{dt^{\alpha_3}} = -be_2 - ce_3 + k_3 e_3. \end{cases} \quad (21)$$

The Jacobian matrix for the error system (21) is

$$J = \begin{pmatrix} -a + k_1 & a & 0 \\ 1 & -1 + k_2 & 1 \\ 0 & -b & -c + k_3 \end{pmatrix}, \quad (22)$$

and so, the characteristic equation $\det(\text{diag}(\lambda^{r_1}, \lambda^{r_2}, \lambda^{r_3}) - J) = 0$ can be written as,

$$\begin{aligned} &(\lambda^{r_1} + a - k_1)((\lambda^{r_2} + 1 - k_2)(\lambda^{r_3} + c - k_3) + b) \\ &\quad - a(\lambda^{r_3} + c - k_3) = 0. \end{aligned} \quad (23)$$

Now, in case of $\alpha = (0.98, 0.98, 0.94)$ and $(a, b, c, m_0, m_1) = (10.1911, 10.3035, 0.1631, -1.1126, -0.8692)$, the systems (18) and (19) are synchronized if all roots of the equation,

$$\begin{aligned} &(\lambda^{49} + 10.1911 - k_1)((\lambda^{49} + 1 - k_2)(\lambda^{47} + 0.1631 - k_3) + 10.3035) \\ &\quad - 10.1911(\lambda^{47} + 0.1631 - k_3) = 0. \end{aligned} \quad (24)$$

lie in the region $|\arg(\lambda)| > \pi/100$. For example, if we take $(k_1, k_2, k_3) = (2, 2, -3)$ then all roots of Eq. (24) lie in the region $|\arg(\lambda)| > \pi/100$. We use Mathematica to verify that the 145 roots of Eq. (24) lie in the region $|\arg(\lambda)| > \pi/100$. Therefore, the drive system (19) and the response system (20) are synchronized. The error functions evolution, in this case, is shown in Fig. 3. From Fig. 3, for the given parameters, we can conclude that the components of the error system (21) decay towards zero as $t \rightarrow +\infty$. So, we can numerically conclude that the designed controller $\mathbf{u}(t) = (2, 2, -3)\mathbf{e}(t)$ can effectively control the chaotic fractional order Chua system to achieve synchronization between the systems (19) and (20).

Also, in case of $\alpha = (0.93, 0.99, 0.92)$ and $(a, b, c, m_0, m_1) = (10.725, 10.593, 0.268, -1.1726, -0.7872)$, the systems (18) and (19) are synchronized if all roots of the equation,

$$\begin{aligned} &(\lambda^{93} + 10.725 - k_1)((\lambda^{99} + 1 - k_2)(\lambda^{92} + 0.268 - k_3) + 10.593) \\ &\quad - 10.725(\lambda^{92} + 0.268 - k_3) = 0. \end{aligned} \quad (25)$$

lie in the region $|\arg(\lambda)| > \pi/200$. For example, if we take $(k_1, k_2, k_3) = (5, -3, 1)$ then all roots of Eq. (25), using Mathematica, lie in the region $|\arg(\lambda)| > \pi/200$. Therefore, the drive system (19) and the response system (20) are synchronized. The error functions evolution, in this case, is shown in Fig. 4. From Fig. 4, for the given parameters, we can conclude that the components of the error system (21) decay towards zero as $t \rightarrow +\infty$. So, we can numerically conclude that the designed controller, $\mathbf{u}(t) = (5, -3, 1)\mathbf{e}(t)$, can effectively control the chaotic fractional order Chua system to achieve synchronization between the systems (19) and (20).

If $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$, then the systems (19) and (20) are synchronized if k_1, k_2 and k_3 satisfy the control law that the three roots of Eq. (23), when $r_1 = r_2 = r_3 = 1$, lie in the region $|\arg(\lambda)| > \alpha\pi/2$.

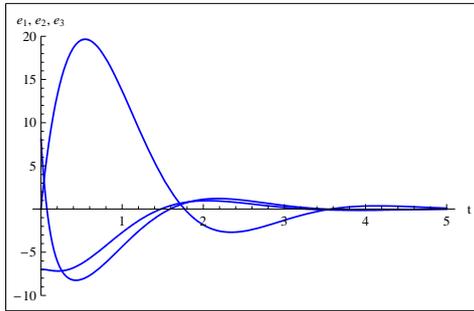


Fig. 3. Synchronization of the fractional order Chua system (18), when $(a, b, c) = (10.1911, 10.3035, 0.1631)$, $\alpha = (0.98, 0.98, 0.94)$ and $(k_1, k_2, k_3) = (2, 2, -3)$.

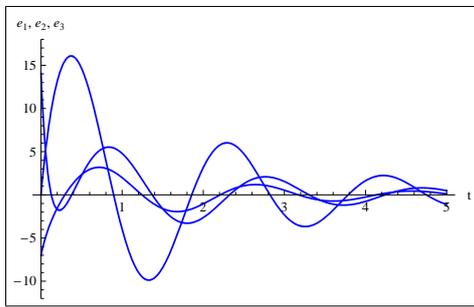


Fig. 4. Synchronization of the fractional order Chua system (18), when $(a, b, c) = (10.725, 10.593, 0.268)$ $\alpha = (0.93, 0.99, 0.92)$ and $(k_1, k_2, k_3) = (5, -3, 1)$.

VI. CONCLUSION

In this paper, we study the master-slave synchronization, based on the stability results of linear fractional order systems, of coupled fractional order chaotic systems. We establish theoretical results for chaos control and chaos synchronization of fractional order systems, such as the fractional extension of Chen and Chua systems. The designed controller that applied to the response system affects the system dynamics to realize synchronization. The controller is designed such that the components of the error system decay towards zero as the time variable, t , tends to infinity. The numerical simulations show the effectiveness and the feasibility of the proposed scheme.

REFERENCES

- [1] Z. M. Ge and C. Y. Ou, "Chaos in a fractional order modified Duffing system," *Chaos, Solitons & Fractals*, vol. 34, no. 2, pp. 262-291, Oct. 2007.
- [2] I. Petráš, "A note on the fractional-order Chua's system," *Chaos, Solitons & Fractals*, vol. 38, no. 1, pp. 140-147, Oct. 2008.
- [3] T. Hartley, C. Lorenzo and H. Qammer H, "Chaos in a fractional order Chua's system," *IEEE Trans CAS-I*, vol. 42, pp. 485-490, Aug. 1995.
- [4] C. Li and G. Chen, "Chaos and hyperchaos in fractional order Rössler equations," *Physica A*, vol. 341, pp.55-61, 2004.
- [5] C. Li C and G. Peng, "Chaos in Chen's system with a fractional order," *Chaos, Solitons & Fractals* vol. 22, no. 2, pp. 443-450, Oct. 2004.
- [6] C. Li and G. Chen, "Chaos in the fractional order Chen system and its control," *Chaos, Solitons & Fractals*, vol. 22, no. 3, pp. 549-554, Nov. 2004.

- [7] J. G. Lu and G. Chen, "A note on the fractional-order Chen system," *Chaos, Solitons & Fractals* vol. 27, no. 3, pp. 685-688, Feb. 2006.
- [8] I. Grigorenko and E. Grigorenko, "Chaotic dynamics of the fractional Lorenz system," *Phys. Rev. Lett.*, vol. 91, no. 3, pp. 034101-034104, July 2003.
- [9] A. Arneodo, P. Couillet, E. Spiegel and C. Tresser C, "Asymptotic chaos," *Physica D*, vol. 14, no. 3, pp. 327-347, March 1985.
- [10] W. H. Deng and C. P. Li, "Chaos synchronization of the fractional Lü system," *Physica A*, vol. 353, pp. 61-72, Aug. 2005.
- [11] L. J. Sheu, H. K. Chen, J. H. Chen and L. M. Tam, "Chaos in a new system with fractional order," *Chaos, Solitons & Fractals*, vol. 31, no. 5, pp. 1203-1212, March 2007.
- [12] J. Yan and C. Li, "On chaos synchronization of fractional differential equations," *Chaos, Solitons & Fractals*, vol. 32, no. 2, pp. 725-735, April 2007.
- [13] C. Li and J. Yan, "The synchronization of three fractional differential systems," *Chaos, Solitons & Fractals*, vol. 32, no. 2, pp. 751-757, April 2007.
- [14] J. Wang, X. Xiong and Y. Zhang, "Extending synchronization scheme to chaotic fractional-order Chen systems," *Physica A*, vol. 370, no. 2, pp. 279-285, Oct. 2006.
- [15] C. P. Li, W. H. Deng and D. Xu, "Chaos synchronization of the Chua system with a fractional order," *Physica A*, vol. 360, no. 2, pp. 171-185, Feb. 2006.
- [16] H. Zhu, S. Zhou and J. Zhang, "Chaos and synchronization of the fractional-order Chua's system," *Chaos, Solitons & Fractals*, in press.
- [17] Y. Yu and H. Li, "The synchronization of fractional-order Rössler hyperchaotic systems," *Physica A*, vol. 387, no. 5, pp. 1393-1403, Feb. 2008.
- [18] I. Podlubny, *Fractional Differential Equations*, New York, Academic Press, 1999.
- [19] R. Gorenflo and F. Mainardi, Fractional calculus: Integral and differential equations of fractional order. in *Fractals and Fractional Calculus* , (eds.: Carpinteri and Mainardi), New York, 1997.
- [20] K. B. Oldham and J. Spanier, *The Fractional Calculus*, New York, Academic Press, 1974.
- [21] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent, Part II," *J. Roy. Astr. Soc.*, vol. 13, pp. 529-539, May 1967.
- [22] D. Matignon, "Stability results of fractional differential equations with applications to control processing," Proc. *IMACS, IEEE-SMC*, Lille, France, pp. 963-968, 1996.
- [23] W. Deng, C. Li and J. Lü, "Stability analysis of linear fractional differential system with multiple time delays," *Nonlinear Dyn.*, vol. 48, no. 4, pp. 409-416, June 2007.
- [24] E. Ahmed, A. El-Sayed and H. El-Saka, "Equilibrium points, stability and numerical solutions of fractional-order predator-prey and rabies models," *J. Math. Anal. Appl.*, vol. 325, no. 1, pp. 542-553, Jan. 2007.
- [25] M. S. Tavazoei and M. Haeri, "A note on the stability of fractional order systems," *Math. Comput. Simulat.*, vol. 79, no. 5, pp. 1566-1576, Jan. 2009.
- [26] M. S. Tavazoei and M. Haeri, "Chaotic attractors in incommensurate fractional order systems," *Physica D*, vol. 237 pp. 2628-2637, March 2008.
- [27] C. Chen and T. Ueta, "Yet another chaotic attractor," *I. J. Bifurc. Chaos*, vol. 9, no. 7, pp. 1465-1466, July 1999.
- [28] L. O. Chua and G-N. Lin, "Canonical realization of Chua's circuit family," *IEEE transactions on Circuits and Systems*, vol. 37, no. 7, pp. 885-902, July 1990.
- [29] L. O. Chua, I. Makoto, L. Kocarev and K. Eckert, "Chaos synchronization in Chua's circuit," *J. Circuits, Systems and Computers*, vol. 3, no. 1, pp. 93-108, March 1993.



Zaid M. Odibat received his Ph.D from the University of Jordan (Jordan) in 2002. Since 2002 he has been at Al-Balqa Applied University. He left Al-Balqa Applied University to spend one year at the University of Le Havre (2008-2009). His research interests focus on Numerical analysis, Ordinary and partial differential equations, Fractional calculus, Numerical solution of fractional nonlinear systems, Control theory, Optimization techniques.
Email: z.odibat@gmail.com

A special kind of synchronization of different chaotic discrete-time systems

Yong Chen, Yiliang Jin, and Xin Li,

Abstract—This paper presents a special kind of the generalized synchronization of different discrete-time systems, proved by Lyapunov asymptotical stability theorem. With the aid of symbolic-numeric computation, we use the scheme to realize the synchronization between 3D Rössler discrete-time system and Hénon-like discrete-time system (The goal system for synchronization is a function), as well as between 3D discrete-time hyperchaotic system and Hénon-like map (The goal system for synchronization is the Rössler system) via three scalar controllers. This discrete-time chaotic system synchronization developed may be applied to the design of secure communication.

Index Terms—discrete-time chaotic system, symbolic-numeric computation, the generalized synchronization, backstepping design.

I. INTRODUCTION

Since Lorenz [1] presented the well-known Lorenz chaotic system and Rössler [2] first introduced the Rössler hyperchaotic system, many chaotic systems have been reported in nonlinear field. In particular, since the pioneering works [3–6], Chaos (hyperchaos) synchronization have played significant roles because of its potential applications in secure communication. Up to now, many types of chaos synchronization have been presented, such as complete synchronization, partial synchronization, phase synchronization, lag synchronization, projective synchronization, generalized synchronization, etc [7–14]. For a particular chaotic system, the drive system, together with an identical or a different system, the response system, our goal is to synchronize them via coupling or other method. Amongst all kinds of chaos synchronization, the generalized synchronization is proposed in the continuous-time systems [15–20]. It means there exists a functional relationship between the states of the drive system and those of the response system. A special kind of generalized synchronization $y = x + F$ is studied, where x, y are the state vectors of the drive system and the response system respectively, F is a given vector function, which may take various forms. When $F = 0$, it reduces to a complete synchronization. Many powerful methods have been reported to investigate some types of chaos (hyperchaotic) synchronization in continuous-time systems. In fact, many mathematical models of neural networks, biological process, physical process and chemical process, were defined using discrete-time dynamical systems [21, 22]. Recently, more and more attentions were paid to the

chaos (hyperchaos) control and synchronization in discrete-time dynamical systems [23–28].

In this paper, we will use the scheme of the discrete-time dynamical systems generalized synchronization and the backstepping design to realize our goal. A systematic and automatic algorithm is set up to achieve successfully synchronization. The numeric computation between 3D Rössler discrete-time system and Hénon-like discrete-time system (The goal system for synchronization is a function), as well as between 3D discrete-time hyperchaotic system due to Wang [29] and Hénon-like map (The goal system for synchronization is the Rössler system) are used to verify the effectiveness of the proposed scheme.

The rest of paper is arranged as follows: In section A, we first introduce generalized synchronization in discrete-time systems; In section B, we give the numeric results of the generalized synchronization between 3D Rössler discrete-time system and Hénon-like discrete-time system; In section C, we investigate the scheme between the 3D discrete-time hyperchaotic system and Hénon-like map; Finally, some conclusions and discussions are given.

A. Generalized synchronization of discrete-time chaotic systems

Firstly we introduce the generalized synchronization in discrete-time systems, and then we use Lyapunov stability theory to realize our scheme.

Definition: For two discrete-time (chaotic or hyperchaotic) dynamical systems (i) $x(k+1) = M(x(k))$ and (ii) $y(k+1) = H(y(k)) + u(x(k), y(k))$, where $(x(k), y(k)) \in R^{m+m}$, $k \in Z/Z^-$, and $u(x(k), y(k)) \in R^m$, let (iii) $E(k) = (E_1(k), E_2(k), \dots, E_m(k)) = (x_1(k) - y_1(k) + F_1(k), x_2(k) - y_2(k) + F_2(k), \dots, x_m(k) - y_m(k) + F_m(k))$ be boundary vector functions, if there exists proper controllers $u(x(k), y(k)) = (u_1(x(k), y(k)), \dots, u_m(x(k), y(k)))^T$ such that $\lim_{k \rightarrow \infty} (E(k)) = 0$, we say that there exist **generalized synchronization** between these two discrete-time chaotic systems.

Remark: It is necessary to point out that the controller u depends on the synchronization method chosen. When $E_i(k) = 0$, ($i = 1, \dots, m$), $u = M(x) - H(y) + F$ is the situation when all the error functions equal to zero and the corresponding controller is trivial situation. For $E_i(k) = 0$, we need only to solve the equations

$E(k) = (E_1(k), E_2(k), \dots, E_m(k)) = (x_1(k) - y_1(k) + F_1(k), x_2(k) - y_2(k) + F_2(k), \dots, x_m(k) - y_m(k) + F_m(k)) = (0, 0, \dots, 0)$ to get the trivial controller "u". So here we just

Y. Chen is with Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, 200062 China e-mail: (ychen@sei.ecnu.edu.cn

Y. Jin and X. Li are with Nonlinear Science Center and Department of Mathematics of Ningbo University.

consider the general condition $\lim_{k \rightarrow \infty} (E(k)) = 0$.

Based on the Lyapunov stability theory, for the error discrete-time (iii) generated by drive system (i) and response system (ii), let $L(E_1(k), E_2(k), \dots, E_m(k))|_{E_i(k) \equiv 0 (i=1,2,\dots,m)} = 0$, if $\Delta L(k) = L(k+1) - L(k) \leq 0$, with the equality holding if and only if $E_i(k) \equiv 0 (i=1,2,\dots,m)$, it is said that systems (i) and (ii) are generalized synchronized.

In this letter based on the backstepping design method, we would like to present a systematic, generalized and constructive scheme to seek the controllers such that 3D Rössler discrete-time system and Hénon-like discrete-time system, as well as between 3D discrete-time hyperchaotic system and Hénon-like map via three scalar controllers are generalized synchronized.

B. Numeric results of the generalized synchronization between 3D Rössler discrete-time system and Hénon-like discrete-time system

In the following, using the backstepping method and based on Lyapunov stability theory, generalized synchronization of 3D Rössler discrete-time system and Hénon-like discrete-time system is realized step by step.

Consider Rössler discrete-time system

$$\begin{cases} x_1(k+1) = 3.8x_1(k)(1-x_1(k)) - 0.05(x_3(k) + 0.35) \\ \quad * (1-2x_2(k)), \\ x_2(k+1) = 3.78x_2(k)(1-x_2(k)) + 0.2x_3(k), \\ x_3(k+1) = 0.1(1-1.9x_1(k))(x_3(k) + 0.35) \\ \quad * (1-2x_2(k) - 1), \end{cases} \quad (1)$$

and Hénon-like system with controllers $u(x, y)$

$$\begin{cases} y_1(k+1) = -by_2(k) + u_1(x, y), \\ y_2(k+1) = 1 + y_3(k) - ay_2(k)^2 + u_2(x, y), \\ y_3(k+1) = by_2(k) + y_1(k) + u_3(x, y), \end{cases} \quad (2)$$

as the drive system and response system, respectively.

Firstly we give out the figures (Fig.1(a) and Fig.1(b)) of the two systems with initial values $[x_1(0) = 0.1, x_2(0) = 0.2, x_3(0) = -0.5]$ and $[y_1(0) = 0.2, y_2(0) = 0.7, y_3(0) = 0.06]$, respectively. Here $a = 1.07, b = 0.3$.

(I): Let the error states be

$$\begin{cases} E_1(k) = x_1(k) - y_1(k) + \tanh(x_1(k)), \\ E_2(k) = x_2(k) - y_2(k) + \tanh(x_2(k)), \\ E_3(k) = x_3(k) - y_3(k) + \tanh(x_3(k)), \end{cases} \quad (3)$$

Based on the backstepping design method and Lyapunov stability theory, we can get the controllers. Here we omit the concrete process. Finally, with the aid of symbolic computation, from

$$E_1(k) = x_1(k) - y_1(k) + \tanh(x_1(k)), \quad (4)$$

$$E_2(k) = E_1(k+1) - c_{11}E_1(k), \quad (5)$$

$$E_3(k) = E_2(k+1) - c_{21}E_1(k) - c_{22}E_2(k), \quad (6)$$

$$E_3(k+1) - c_{31}E_1(k) - c_{32}E_2(k) - c_{33}E_3(k) = 0, \quad (7)$$

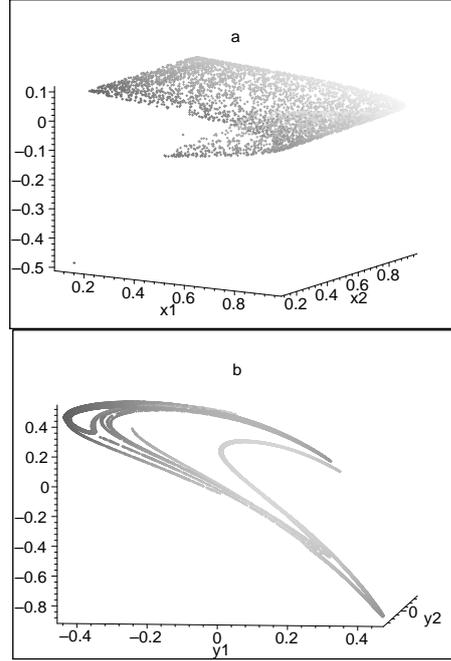


Fig. 1. (a) Rössler discrete-time attractor; (b) Hénon-like discrete-time attractor

Let the Lyapunov function be $L(k) = |E_1(k)| + d_1|E_2(k)| + d_2|E_3(k)|$, $d_2 > d_1 > 1$. Then from (4), (5), (6) and (7), we obtain the derivative of the Lyapunov function $L(k)$

$$\begin{aligned} \Delta L(k) &= L(k+1) - L(k) \\ &\leq (d_2|c_{31}| + d_1|c_{21}| + |c_{11}| - 1)|E_1(k)| + (d_2|c_{32}| \\ &\quad + d_1(|c_{22}| - 1) + 1)|E_2(k)| \\ &\quad + (d_2|c_{33}| + d_1 - d_2)|E_3(k)|. \end{aligned}$$

If we set these constants $c_{11}, c_{21}, c_{22}, c_{31}, c_{32}, c_{33}$ satisfy

$$\begin{aligned} d_1|c_{21}| + d_2|c_{31}| + |c_{11}| &< 1, \\ d_1|c_{22}| + d_2|c_{32}| &< d_1 - 1, \\ |c_{33}| &< \frac{d_2 - d_1}{d_2}, \end{aligned}$$

then $\Delta L(k)$ is negative definite which denotes that the resulting close-loop discrete-time system

$$\begin{pmatrix} E_1(k+1) \\ E_2(k+1) \\ E_3(k+1) \end{pmatrix} = \begin{pmatrix} c_{11} & 1 & 0 \\ c_{21} & c_{22} & 1 \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} E_1(k) \\ E_2(k) \\ E_3(k) \end{pmatrix}$$

is globally asymptotically stable and $\lim_{k \rightarrow +\infty} E_i(k) = 0$, that is to say, Rössler 3D discrete-time hyperchaotic system (1) and the Hénon-like map (2) are function projective synchronized.

Then with the aid of symbolic computation, from the above

equations (1) and (2) we obtained the controllers

$$\left\{ \begin{array}{l} u_1(x, y) = 3.8x_1(k) - 3.8x_1(k)^2 - 0.05x_3(k) \\ \quad + 0.1x_3(k)x_2(k) - 0.0175 - 0.965x_2(k) \\ \quad + by_2(k) - \tanh(-3.8x_1(k) + 3.8x_1(k)^2 \\ \quad + 0.05x_3(k) - 0.1x_3(k)x_2(k) + 0.0175 \\ \quad - 0.035x_2(k)) - c_{11}x_1(k) + c_{11}y_1(k) \\ \quad - c_{11} \tanh(x_1(k)) + y_2(k) - \tanh(x_2(k)), \\ u_2(x, y) = 3.78x_2(k) - 3.78x_2(k)^2 - 0.8x_3(k) \\ \quad - 1 + ay_2(k)^2 + \tanh(3.78x_2(k) - 3.78x_2(k)^2 \\ \quad + 0.2x_3(k)) - c_{21}x_1(k) + c_{21}y_1(k) \\ \quad - c_{21} \tanh(x_1(k)) - c_{22}x_2(k) \\ \quad + c_{22}y_2(k) - c_{22} \tanh(x_2(k)) - \tanh(x_3(k)), \\ u_3(x, y) = 0.1x_3(k) - 0.2x_3(k)x_2(k) - 0.065 \\ \quad - 0.07x_2(k) - 0.19x_1(k)x_3(k) \\ \quad + 0.38x_1(k)x_3(k)x_2(k) + 0.1235x_1(k) \\ \quad + 0.133x_1(k)x_2(k) - by_2(k) - y_1(k) \\ \quad + \tanh(0.0005(-10 + 19x_1(k))(-20x_3(k) \\ \quad + 40x_3(k)x_2(k) + 13 + 14x_2(k))) \\ \quad - c_{31}x_1(k) + c_{31}y_1(k) - c_{31} \tanh(x_1(k)) \\ \quad - c_{32}x_2(k) + c_{32}y_2(k) \\ \quad - c_{32} \tanh(x_2(k)) - c_{33}x_3(k) \\ \quad + c_{33}y_3(k) - c_{33} \tanh(x_3(k)). \end{array} \right. \quad (8)$$

In the following we use numerical simulations to verify the effectiveness of the obtained controllers $u(x, y)$. Here take $c_{11} = 0.3, c_{21} = 0.02, c_{22} = 0.4, c_{31} = 0.05, c_{32} = 0.1, c_{33} = -0.2, d_1 = 4, d_2 = 6$, and the initial values of system (1) and (2) respectively. The graphs of the error states are shown in Fig.2 (a)-(c), and the attractors of the two systems with controllers are displayed in Fig.3.

C. Generalized synchronization of 3D discrete-time hyperchaotic system and Hénon-like map

In this section, we would like to realize the generalized synchronization of 3D discrete-time hyperchaotic system

$$\left\{ \begin{array}{l} x_1(k+1) = 0.5\delta x_2(k) + (-2.3\delta + 1)x_1(k), \\ x_2(k+1) = 0.2\delta x_3(k) - 1.9\delta x_1(k) + x_2(k), \\ x_3(k+1) = 2\delta - 0.6\delta x_2(k)x_3(k) + (-1.9\delta + 1)x_3(k), \end{array} \right. \quad (9)$$

and the 3D Hénon-like discrete-time map

$$\left\{ \begin{array}{l} y_1(k+1) = -by_2(k) + u_1(x, y, z), \\ y_2(k+1) = 1 + y_3(k) - ay_2(k)^2 + u_2(x, y, z), \\ y_3(k+1) = by_2(k) + y_1(k) + u_3(x, y, z), \end{array} \right. \quad (10)$$

as the drive system and response system, respectively.

Firstly we give out the figures (Fig.4) of the discrete-time hyperchaotic system due to Wang with initial values $[x_1(0) = 0.05, x_2(0) = 0.03, x_3(0) = 0.02], [y_1(0) = 0.2, y_2(0) = 0.7, y_3(0) = 0.06]$, and $\delta = 1$.

The goal system for synchronization is the Rössler system

$$\left\{ \begin{array}{l} z_1(k+1) = 3.8z_1(k)(1 - z_1(k)) - 0.05(z_3(k) + 0.35) \\ \quad * (1 - 2z_2(k)), \\ z_2(k+1) = 3.78z_2(k)(1 - z_2(k)) + 0.2z_3(k), \\ z_3(k+1) = 0.1(1 - 1.9z_1(k))(z_3(k) + 0.35) \\ \quad * (1 - 2z_2(k)) - 1, \end{array} \right. \quad (11)$$

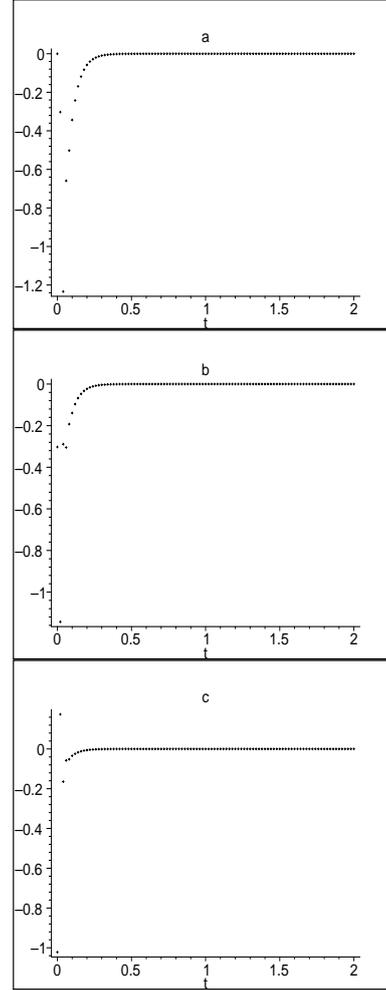


Fig. 2. the orbits of the error states: (a) the orbit of e_1 ; (b) the orbit of e_2 ; (c) the orbit of e_3 .

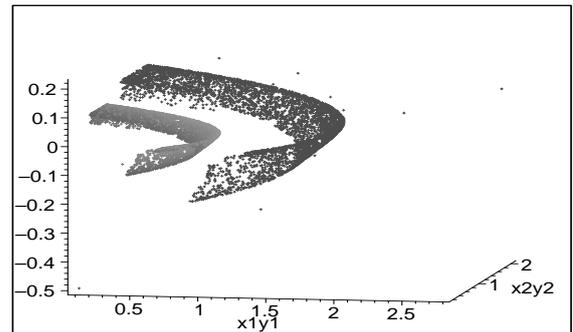


Fig. 3. the two attractors after being synchronized (the dark one is the response system with the controllers, and the other is the drive system).

with the initial value $[z_1(0) = 0.1, z_2(0) = 0.2, z_3(0) = -0.5]$.

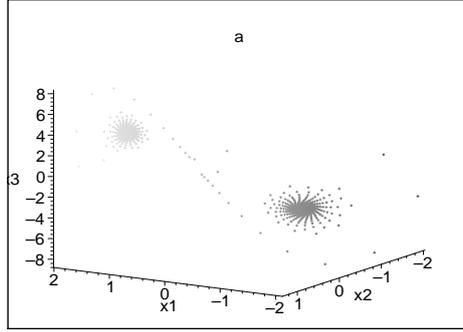


Fig. 4. the discrete-time hyperchaotic system due to Wang attractor

(I): Let the error states be

$$\begin{cases} E_1(k) = x_1(k) - y_1(k)/2 + \frac{1}{32} \tanh(x_1(k))^2 \tanh(z_1(k))^2, \\ E_2(k) = x_2(k) - y_2(k)/2 + \frac{1}{32} \tanh(x_2(k))^2 \tanh(z_2(k))^2, \\ E_3(k) = x_3(k) - y_3(k)/2 + \frac{1}{32} \tanh(x_3(k))^2 \tanh(z_3(k))^2, \end{cases} \quad (12)$$

Based on the backstepping design method and Lyapunov stability theory, we can get the controllers. Here we omit the concrete process. Finally, with the aid of symbolic computation, from

$$E_1(k) = x_1(k) - y_1(k)/2 + \frac{1}{32} \tanh(x_1(k))^2 \tanh(z_1(k))^2, \quad (13)$$

$$E_2(k) = E_1(k+1) - c_{11}E_1(k), \quad (14)$$

$$E_3(k) = E_2(k+1) - c_{21}E_1(k) - c_{22}E_2(k), \quad (15)$$

$$E_3(k+1) - c_{31}E_1(k) - c_{32}E_2(k) - c_{33}E_3(k) = 0, \quad (16)$$

Let the Lyapunov function be $L(k) = |E_1(k)| + d_1|E_2(k)| + d_2|E_3(k)|$, $d_2 > d_1 > 1$. Then from (13), (14), (15) and (16), we obtain the derivative of the Lyapunov function $L(k)$

$$\begin{aligned} \Delta L(k) &= L(k+1) - L(k) \\ &\leq (d_2|c_{31}| + d_1|c_{21}| + |c_{11}| - 1)|E_1(k)| \\ &\quad + (d_2|c_{32}| + d_1(|c_{22}| - 1) + 1)|E_2(k)| \\ &\quad + (d_2|c_{33}| + d_1 - d_2)|E_3(k)|. \end{aligned}$$

If we set these constants $c_{11}, c_{21}, c_{22}, c_{31}, c_{32}, c_{33}$ satisfy

$$\begin{aligned} d_1|c_{21}| + d_2|c_{31}| + |c_{11}| &< 1, \\ d_1|c_{22}| + d_2|c_{32}| &< d_1 - 1, \\ |c_{33}| &< \frac{d_2 - d_1}{d_2}, \end{aligned}$$

then $\Delta L(k)$ is negative definite which denotes that the resulting close-loop discrete-time system

$$\begin{pmatrix} E_1(k+1) \\ E_2(k+1) \\ E_3(k+1) \end{pmatrix} = \begin{pmatrix} c_{11} & 1 & 0 \\ c_{21} & c_{22} & 1 \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} E_1(k) \\ E_2(k) \\ E_3(k) \end{pmatrix}$$

is globally asymptotically stable and $\lim_{k \rightarrow +\infty} E_i(k) = 0$, that is to say, the discrete-time hyperchaotic system (9) and the Hénon-like map (10) are function projective synchronized.

Then with the aid of symbolic computation, from the above equations (9) and (10) we obtained the controllers

$$\begin{cases} u_1(x, y) = \delta x_2(k) - 4.6\delta x_1(k) \\ \quad + 2x_1(k) + by_2(k) \\ \quad + 0.0625 \tanh(0.5\delta x_2(k)) \\ \quad - 2.3\delta x_1(k) + x_1(k))^2 \\ \quad \tanh(-3.8z_1(k) + 3.8z_1(k))^2 \\ \quad + 0.05z_3(k) - 0.1z_3(k)z_2(k) + 0.0175 \\ \quad - 0.035z_2(k))^2 - 2c_{11}x_1(k) + c_{11}y_1(k) - 0.0625c_{11} \\ \quad \tanh(x_1(k))^2 \tanh(z_1(k))^2 - 2x_2(k) + y_2(k) \\ \quad - 0.0625 \tanh(x_2(k))^2 \tanh(z_2(k))^2, \\ u_2(x, y) = 0.4\delta x_3(k) - 3.8\delta x_1(k) + 2x_2(k) \\ \quad - 1 + ay_2(k)^2 + 0.0625 \tanh(0.2\delta x_3(k)) \\ \quad - 1.9\delta x_1(k) + x_2(k))^2 \tanh(3.78z_2(k)) \\ \quad - 3.78z_2(k))^2 + 0.2z_3(k))^2 - 2c_{21}x_1(k) \\ \quad + c_{21}y_1(k) - 0.0625c_{21} \tanh(x_1(k))^2 \tanh(z_1(k))^2 \\ \quad - 2c_{22}x_2(k) + c_{22}y_2(k) \\ \quad - 0.0625c_{22} \tanh(x_2(k))^2 \tanh(z_2(k))^2 \\ \quad - 2x_3(k) - 0.0625 \tanh(x_3(k))^2 \tanh(z_3(k))^2, \\ u_3(x, y) = 4\delta - 1.2\delta x_2(k)x_3(k) - 3.8\delta x_3(k) \\ \quad + 2x_3(k) - by_2(k) - y_1(k) \\ \quad + 0.0625 \tanh(-2\delta + 0.6\delta x_2(k)x_3(k)) \\ \quad + 1.9\delta x_3(k) - x_3(k))^2 \\ \quad \tanh(0.0005(-10 + 19z_1(k))(-20z_3(k) \\ \quad + 40z_3(k)z_2(k) + 13 + 14z_2(k)))^2 \\ \quad - 2c_{31}x_1(k) + c_{31}y_1(k) - 0.0625c_{31} \\ \quad \tanh(x_1(k))^2 \tanh(z_1(k))^2 - 2c_{32}x_2(k) \\ \quad + c_{32}y_2(k) - 0.0625c_{32} \tanh(x_2(k))^2 \\ \quad \tanh(z_2(k))^2 - 2c_{33}x_3(k) + c_{33}y_3(k) \\ \quad - 0.0625c_{33} \tanh(x_3(k))^2 \tanh(z_3(k))^2. \end{cases} \quad (17)$$

In the following we use numerical simulations to verify the effectiveness of the obtained controllers $u(x, y, z)$. Here take $c_{11} = 0.3, c_{21} = 0.02, c_{22} = 0.4, c_{31} = 0.05, c_{32} = 0.1, c_{33} = -0.2, d_1 = 4, d_2 = 6$, and the initial values of system (9), (10) and (11) respectively. The graphs of the error state are shown in Fig.5 (a)-(c), and the attractors of the two systems with controllers are displayed in Fig.6.

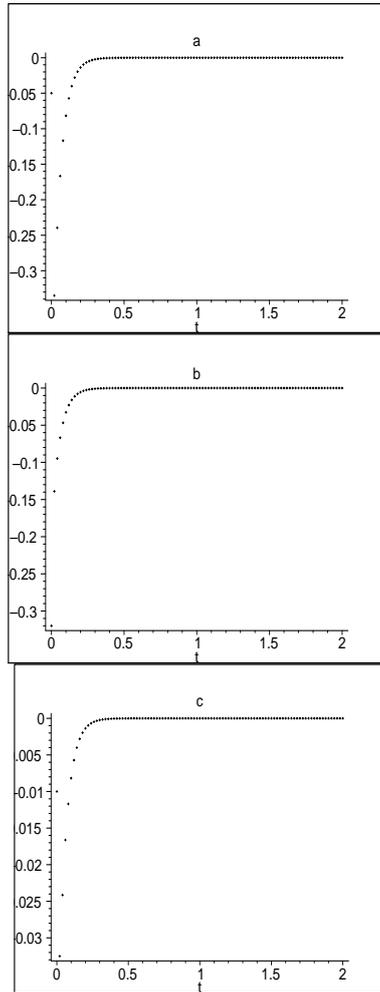


Fig. 5. the orbits of the error states: (a) the orbit of e_1 ; (b) the orbit of e_2 ; (c) the orbit of e_3 .

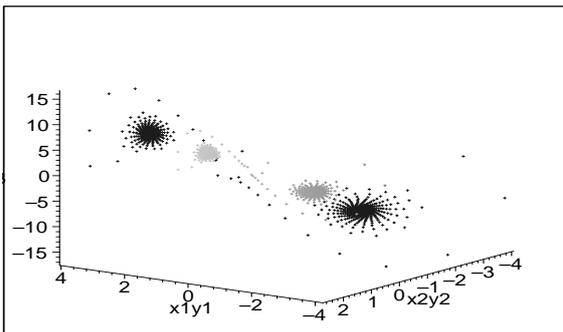


Fig. 6. the two attractors after being synchronized (the dark one is the response system with the controllers, and the other is the drive system).

II. CONCLUSION

In summary, based on backstepping method and Lyapunov stability theory, a systematic and automatic scheme is devel-

oped investigate the generalized synchronization between the discrete-time drive systems and response systems: between the 3D Rössler discrete-time system and Hénon-like discrete-time system (The goal system for synchronization is a function), as well as between 3D discrete-time hyperchaotic system and Hénon-like map (The goal system for synchronization is the Rössler system) via three scalar controllers. Numerical simulations show the effectiveness of the proposed scheme. Some interesting figures are drawn to show the generalized synchronization between different discrete-time system. In addition, the scheme can be also applied to investigate the tracking problem in the discrete-time systems and to generate automatically the scalar controller in computer with the aid of symbolic-numeric computation.

ACKNOWLEDGMENT

The work is supported the National Natural Science Foundation of China under Grant Nos 10735030 and 90718041, Shanghai Leading Academic Discipline Project (No B412), Programme for Changjiang Scholars and Innovative Research Team in University (IRT0734) and K. C. Wong Magna Fund in Ningbo University).

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] E. N. Lorenz, *J. Atmos. Sci.* **20**, (1963) 130.
- [3] O. E. Rössler, *Phys. Lett. A* **71**, (1979) 397.
- [4] H. Fujisaka and T. Yamada, *Prog. Theor. Phys.* **69**, (1983) 32.
- [5] L. M. Pecora and T.L. Carroll, *Phys. Rev. Lett.* **64**, (1990) 821.
- [6] K. Pyragas, *Phys. Lett. A* **170**, (1992) 421.
- [7] E. Ott, C. Grebogi, and J. A. Yorke, *Phys. Rev. Lett.* **64**, (1990) 1196.
- [8] S. Boccaletti, J. Kurth, G. Osipov, D. L. Valladares, and C. S. Zhou, *Phys. Rep.* **366**, (2002) 1.
- [9] L. Kocarev and U. Parlitz, *Phys. Rev. Lett.* **76**, (1996) 1816.
- [10] R. Brown and L. Kocarev, *Chaos* **10**, (2000) 344.
- [11] S. Boccaletti, L. M. Pecora, and A. Pelaez, *Phys. Rev. E* **63**, (2001) 066219.
- [12] T. Kapitaniak, L. O. Chua, and G. Q. Zhong, *Int. J. Bifur. Chaos* **6**, (1996) 211.
- [13] G. Chen and X. Dong, *From Chaos to Order* (World Scientific, Singapore, 1998)
- [14] Z. Y. Yan, *Chaos* **15**, (2005) 023902;
- [15] Y. Chen, X. Li, *International Journal of Modern Physics C*, **18**, (2007) 883; Y. Chen, X. Li, *Zeitschrift fuer Naturforschung*, **62a**, (2007) 29.
- [16] M. Itoh, and H. Murakami, *IEICE Trans. Fundamentals E* **77**, (1994) 2092.
- [17] S. S. Yang, *Chaos, Solitons Fractals* **9**, (1998) 1703.
- [18] X. S. Yang, *Phys Lett A* **260**, (1999) 340.
- [19] M. Y. Chen, Z. Z. Han, and Y. Shang, *Int J Bifurcation Chaos* **14**(3), (2004) 1085.
- [20] J. Lu, and Y. Xi, *Solitons Fractals* **17**, (2003) 825.
- [21] M. G. Zheng, and H. Y. Cheng, *chaos, solitons and fractals*, **35**, (2008) 980.
- [22] M. Henon, *Commun. Math. Phys.* **50**, (1976) 69.
- [23] K. Konishi and H. Kokame, *Phys. Lett. A* **248**, (1998) 359.
- [24] G. M. Zaslavsky, M. Edelman, and B.A. Niyazov, *Chaos* **7**, (1997) 159.
- [25] Z. Y. Yan, *Phys. Lett. A* **342**, (2005) 309.
- [26] I. Kanellakopoulos, P. V. Kokotovic, and A. S. Morse, *IEEE Trans. Autom. Control* **36**, (1991) 1241.
- [27] M. Krstic, I. Kaneakopoulos, and P. V. Kokotovic, *Nonlinear Adaptive Control Design* (Wiley, New York, 1995).
- [28] C. Wang and S. S. Ge, *Chaos, Solitons Fractals* **12**, (2001) 1199.
- [29] X. Li, Y. Chen, and Z. B. Li, *Function projective synchronization in discrete-time chaotic systems*, *Zeitschrift fuer Naturforschung*, Accepted.
- [30] X. Y. Wang, *Chaos in Complex Nonlinear Systems* (Publishing House of Electronics Industry, Beijing, 2003).

Non-Standard Discretization of Fractional Differential Equations

G. Hussian Erjaee, Modi Alnasr and Shaher Momani

Abstract— One would expect a discretized difference equation to exhibit dynamical behavior similar to its differential counterpart. The Mickens non-standard discretization method effectively preserves the dynamical behavior of nonlinear ordinary and partial differential equations. In this article, we apply this method to fractional differential equations and thereby increase the accuracy of the solutions.

Keywords— Fractional differential equations, dynamical systems, limit cycle, nonstandard schemes.

I. INTRODUCTION

The goal of any discretization process is to produce a difference equation whose dynamics are as close as possible to those of the original differential equation, but this is often difficult, particularly when we are dealing with nonlinear components in a differential equation. Thus the choice of discretization schemes that produce difference equations whose dynamics resemble those of their continuous counterparts poses a major challenge in numerical analysis. Mickens has introduced two such schemes for asymptotically stable systems [1] and Kahan has introduced another for periodic systems [2]. These and some other similar numerical methods for both differential equations and fractional differential equations are widely used, and the resulting difference equations are dynamically consistent with their continuous counterparts [3-8].

We shall apply Mickens' non-standard method to fractional differential equations, which are increasingly used to model problems in a number of research areas including dynamical systems, mechanical systems, signal processing, control, chaos, chaos synchronization, and others. Some of these applications may be found in [9, 10] and the references therein.

The most important advantage of using fractional differential equations in these and other applications is their *non-local property*.

Manuscript received March, 29, 2001. This work was supported in part by the Qatar University interior research under Grant QUUG-CAS-MATH-08-03.

G. Hussian Erjaee was with the Mathematics Department, Shiraz University, Shiraz, Iran. Now he is with Mathematics Department, Qatar University, Doha, Qatar, phone: +974-525-1739; fax: +974-485-2971; (e-mail: erjaee@qu.edu.qa).

Modi Alnasr is with Mathematics Department, Qatar University, Doha, Qatar (e-mail: modialnasr@qu.edu.qa).
Shaher Momani is with the Mathematics Department, Mutah University, Al-Karak, Jordan (e-mail: shaherm@yahoo.com).

It is well known that the integer order differential operator is a local operator but the fractional order differential operator is non-local. This means that the next state of a system depends not only upon its current state but also upon all of its historical states. This is more realistic and it is one reason why fractional calculus has become more and more popular. On the other hand, the integer order differential operator is indifferent to its history. Because of this, and because of some other significant properties of fractional differential equations, a great deal of effort has been expended over the past decade in attempting to find robust and stable numerical and analytical methods for solving these equations. Numerical and analytical methods have included the finite difference method [11], the Adomian decomposition method [12-14], the variational iteration method [15-17], and the homotopy perturbation method [17, 18]. Among these, variational iteration and Adomian decomposition method are the most transparent because they provide immediate and visible symbolic terms of analytic solutions, as well as numerical approximate solutions to both linear and nonlinear differential equations without linearization or discretization. He [19, 20] has also proposed a new perturbation technique, the *homotopy perturbation method*, which has been applied to various nonlinear problems and particularly to oscillator equations. For more details, see [21-28] and the references therein.

We shall see that the non-standard discretization is another numerical way to solve the fractional differential equations while preserving their crucial non-local property. Specifically, we apply the Mickens non-standard discretization scheme [1] to the Grunwald-Letnikov discretization process for fractional differential equations. We will see in section 3 that for some non-linear fractional differential equations, this leads to faster convergence and more accurate results when compared by standard alternative methods.

II. GENERALIZING THE MICKENS METHOD OF NON-STANDARD DISCRETIZATION

The forward Euler method is one of the simplest discretization schemes. In this method the derivative term $\frac{dy}{dt}$ is replaced by $\frac{y(t+h)-y(t)}{h}$, where h is the step size. However, in the Mickens schemes this term is replaced by $\frac{y(t+h)-y(t)}{\varphi(h)}$, where $\varphi(h)$ is a continuous function of step

size h . In addition to this replacement, if there are nonlinear terms such as $y^2(t)$ in the differential equation, these are replaced by $y(t)y(t+h)$ or $y(t-h)y(t)$. In dimensions two and above, nonlinear terms such as $y(t)x(t)$ are either replaced by, $y(t)x(t+h)$, $y(t+h)x(t)$ or left untouched depending upon the context of the differential equation. While we know of no appropriate general method for choosing the function $\varphi(h)$ or for choosing which nonlinear terms are to be replaced, some special techniques may be found in [1] and [29]. Applying this scheme to the first order ODE

$$\frac{dy}{dt} = f(t, y(t)), \quad y(0) = y_0$$

yields the difference equation

$$y(t_{n+1}) = y(t_n) + \varphi(h)F(t_n, y(t_n), y(t_{n+1})), \quad y(t_0) = y_0$$

where $F(t, y(t), y(t)) = f(t, y(t))$, $t_n = nh$, $t_{n+1} = (n+1)h$ and h is the step size.

Now we generalize this scheme to fractional differential and integral equations. We begin with the single fractional differential equation

$$D^\alpha y(t) = f(t, y(t)), \quad T \geq t \geq 0 \text{ and } y(t_0) = y_0, \quad (1)$$

where $\alpha > 0$ and D^α denotes the fractional derivative in the Caputo sense [30], defined by

$$D^\alpha y(t) = J^{n-\alpha} D^n y(t).$$

Here $n-1 < \alpha \leq n$, $n \in N$ and J^n is the n^{th} -order Riemann-Liouville integral operator defined as

$$J^n y(t) = \frac{1}{\Gamma(n)} \int_0^t (t-\tau)^{n-1} y(\tau) d\tau,$$

with $t > 0$. A limited number of methods have been used to solve the initial value problem (1). We have chosen to use the Grunwald-Letnikov method to enable us to apply Mickens' scheme. This method approximates the one-dimensional fractional derivative as follows [31]:

$$D^\alpha y(t) = \lim_{h \rightarrow 0} h^{-\alpha} \sum_{j=0}^{\lfloor t/h \rfloor} (-1)^j \binom{\alpha}{j} y(t-jh), \quad (2)$$

where $\lfloor t \rfloor$ denotes the integer part of t and h is the step size. Thus equation (1) is discretized as

$$\sum_{j=0}^{\lfloor t_n/h \rfloor} c_j^\alpha y(t_{n-j}) = f(t_n, y(t_n)), \quad n = 1, 2, 3, \dots, \quad (3)$$

where $t_n = nh$ and c_j^α are the Grunwald-Letnikov coefficients defined as

$$c_j^\alpha = h^{-\alpha} (-1)^j \binom{\alpha}{j}, \quad j = 0, 1, 2, \dots,$$

or recursively by

$$c_0^\alpha = h^{-\alpha} \text{ and } c_j^\alpha = \left(1 - \frac{1+\alpha}{j}\right) c_{j-1}^\alpha, \quad j = 1, 2, 3, \dots \quad (4)$$

We will now apply the Mickens discretization scheme to the above results by replacing the step size h by a function of h , $\varphi(h)$, and by changing any nonlinear term to the corresponding one defined above. In the next section we will

present some examples to demonstrate the accuracy and efficiency of our approach.

III. NUMERICAL RESULTS

Our first example is the fractional order logistic equation

$$D^\alpha y(t) = y(1-y), \quad 0 < \alpha \leq 1, \quad y(t_0) = y_0. \quad (5)$$

Using the Grunwald-Letnikov discretization method, equation (5) is discretized as

$$\sum_{j=0}^{\lfloor t_n/h \rfloor} c_j^\alpha y(t_{n-j}) = y(t_n)(1-y(t_n)), \quad y(t_0) = y_0. \quad (6)$$

Replacing h by $\varphi(h)$ and $y^2(t_n)$ by $y(t_n)y(t_{n-1})$ and doing some algebraic manipulation (6) yield

$$y(t_{n+1}) = (\alpha + (\varphi(h))^\alpha) y(t_n) - (\varphi(h))^\alpha y(t_n) y(t_{n-1}) - (\varphi(h))^\alpha \sum_{j=2}^n c_j^\alpha y(t_{n-j}), \quad y(t_0) = y_0, \quad (7)$$

where c_j^α is defined in equation (4). Now we must choose a reasonable $\varphi(h)$. First, it is easy to see that the exact solution of

(5) for $\alpha = 1$ is $y(t) = \frac{e^t y_0}{1 + (e^t - 1)y_0}$. Also, using the Euler

discretization method, we have $\frac{y(t) - y(t_0)}{h} = y(t)(1 - y(t_0))$,

so if we replace h by $\varphi(h) = e^h - 1$, we get the same solution as in the exact case. Therefore it is reasonable to choose $\varphi(h) = e^h - 1$ in (7). We have used the recursive formula (7) to solve the logistic fractional differential equation for various values of α and the results are illustrated in Figure 1-b. As expected, all of the solutions for different initial values and $\alpha = 1$ converge to the stable fixed point of the equation, $\bar{y} = 1$. This convergence is faster than the convergence under the standard Grunwald-Letnikov method (see Figure 1-a). We also compare the convergence of solutions for different values of $\alpha \in (0, 1]$ using the standard Grunwald-Letnikov method (Figure 1-a) and using the Mickens non-standard method (Figure 1-b).

For our second example we have chosen the Brusselator system of fractional differential equations. As discussed elsewhere [32, 33] the integer differential order of this system has extensive dynamics. Indeed, the periodic solutions of this system are very sensitive to the parameters in the system. The fractional order of this system is stated as

$$\begin{cases} D^{\alpha_1} y_1(t) = a - (\mu + 1)y_1 + y_1^2 y_2 \\ D^{\alpha_2} y_2(t) = \mu y_1 - y_1^2 y_2, \end{cases} \quad (8)$$

where $\alpha_i \in (0, 1]$ for $i = 1, 2$. It is well known that this system has a unique limit cycle for $\alpha_1 = \alpha_2 = 1$ and $\mu > 2a + 1$. Furthermore, this limit cycle is stable whenever $2(a-1) < \mu \leq 2a + 1$. Here, to compare our results with those found in [33], we fix $a = 1$ and $\mu = 4$. It was shown in [33] that 0.97 is an upper bound on the so-called *efficient dimension* $\alpha = \alpha_1 + \alpha_2$ on which the limit cycle exists.

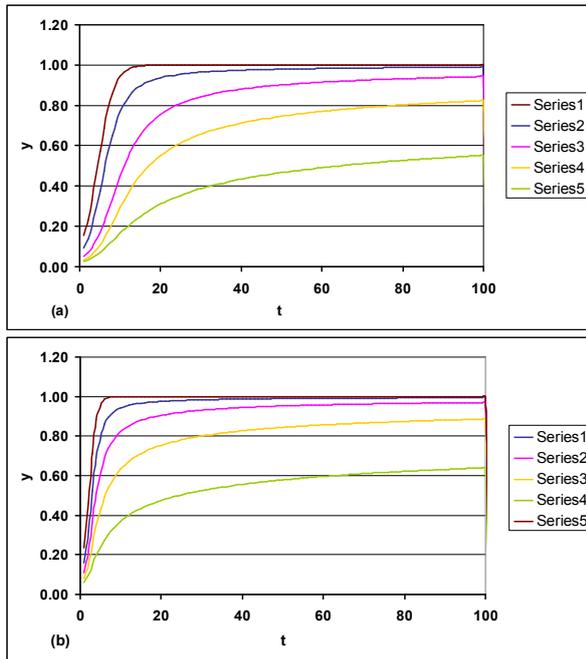


Fig. 1. Different solutions of the Logistic fractional differential equations. In both figures Series 1 to 5 (from top to bottom) shows the solutions for α equal to 1.0, 0.8, 0.6, 0.4 and 0.2, respectively. (a) shows the results found by standard Grunwald-Letnikov method and (b) shows the results, with faster convergence, found by nonstandard Grunwald-Letnikov.

Now we will show that this upper bound is reduced to 0.85 by utilizing the non-standard discretization method. First, by applying the Grunwald-Letnikov method and replacing $y_1^2(t_{n-1})$ by $y_1(t_{n-2})y_1(t_{n-1})$, system (8) is discretized as

$$\begin{cases} y_1(t_n) = \varphi_1(h)^{\alpha_1} \left[a - (\mu + 1)y_1(t_{n-1}) + y_1(t_{n-2})y_1(t_{n-1})y_2(t_{n-1}) \right. \\ \quad \left. - \sum_{j=1}^N \left(1 - \frac{1 + \alpha_1}{j}\right) y_1(t_{n-j}) \right] \\ y_2(t_n) = \varphi_2(h)^{\alpha_2} \left[\mu y_1(t_{n-1}) - y_1(t_{n-2})y_1(t_{n-1})y_2(t_{n-1}) \right. \\ \quad \left. - \sum_{j=1}^N \left(1 - \frac{1 + \alpha_2}{j}\right) y_2(t_{n-j}) \right]. \end{cases} \quad (9)$$

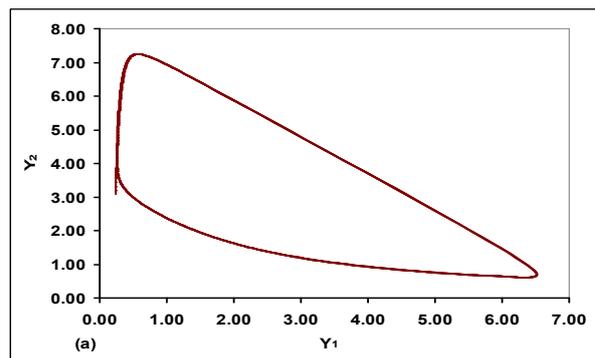
Note that there are some other similar terms such as $y_1(t_n)y_1(t_{n-1})y_2(t_{n-1})$ for replacing the non-linear term $y_1^2 y_2$ in (8), but this makes no different in finding $y(t_n)$ or $y(t_{n+1})$ recursively.

Following our previous example, we choose $\varphi_1(h) = e^h - 1$ and $\varphi_2(h) = 1 - e^{-h}$. Also, to be consistent with the conditions in [33], we choose $y_1(t_0) = 0.2$ and $y_2(t_0) = 0.03$ as our initial conditions with $h = 0.025$. Moreover, in searching for a lower bound on the efficient dimension $\alpha = \alpha_1 + \alpha_2$ for which the limit cycle exists, we use the same algorithm as Wang and Li. That is, we first fix α_1 and change α_2 by multiples of 0.1 or 0.01 up to the point that the

limit cycle of the system (8) exists. Then we fix α_2 and change α_1 in the same manner. Our numerical results show the rapid convergence for various values of α_1 and α_2 in $(0, 1]$. As we see in Figure 2 for $\alpha_1 = \alpha_2 = 1$, the results found by our method and the ones found by the standard Grunwald-Letnikov method, using the dynamical system software Phaser [34], are in complete agreement. To find the limit cycle for smaller values of α_i , such as $\alpha_1 = 0.56$ and $\alpha_2 = 0.48$, we need more iterations. Indeed, for such smaller values we choose $N = 5000$ to obtain reasonable convergence. Finally, as mentioned above, we decrease the possible smallest efficient dimension on which the limit cycle exists from 0.97, found by Wang and Li [33], to $\alpha = 0.45 + 0.40 = 0.85$. The limit cycle for this case is illustrated in Figure 2.

IV. CONSEQUENCES

Obviously, a central issue in the numerical integration of any differential equation is determining a reasonable method of discretizing the equation. This issue is complicated in the case of fractional differential equations due to the existence of long series of computations. As we have seen in this article, using the Mickens non-standard discretization method can improve the accuracy of computation, particularly when the fractional differential equation is nonlinear. This method can be used for either individual or systems of fractional differential equations. In particular, by applying this method to the Brusselator system of fractional differential equations, we reduced the efficient dimension for the existence of the limit cycle to 0.85. Finally, we observe that we can use the Mickens non-standard discretization method in conjunction with not just the Grunwald-Letnikov method, but also with any other discretizing process for fractional differential equations.



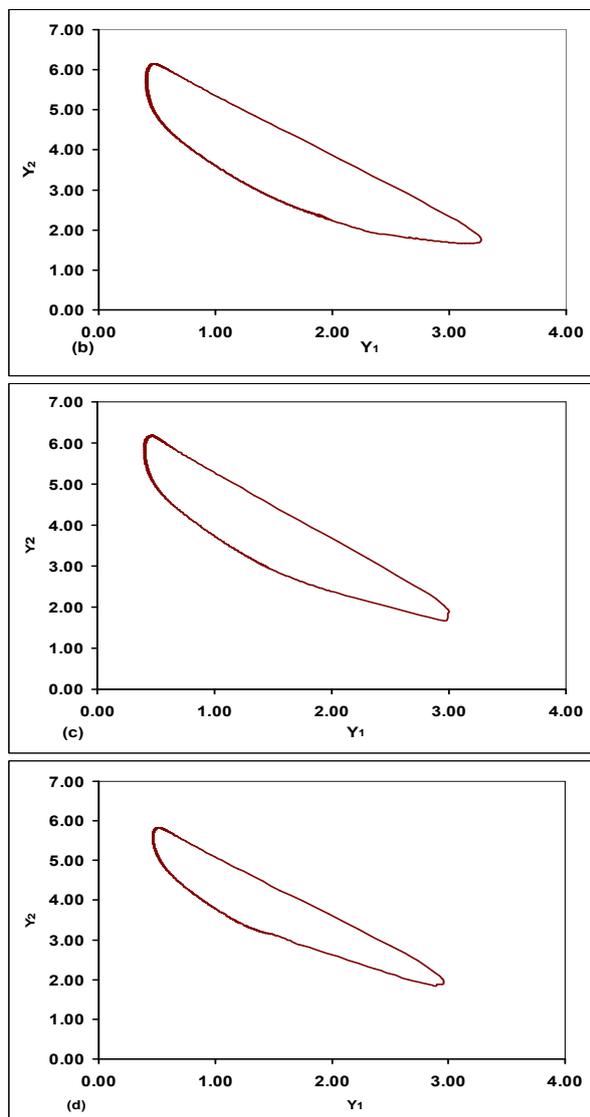


Fig. 2. Limit cycle for the Brusselator fractional differential equations using the nonstandard Grunwald-Letnikov method. The values of (α_1, α_2) are (1.0, 1.0), (0.56, 0.48), (0.50, 0.40) and (0.45, 0.40), respectively in figures (a)-(d). In all these figures the initial value is $(y_1, y_2) = (0.02, 0.03)$ with $h = 0.025$, and just 1000 of the last iterations are illustrated.

REFERENCES

- [1] R. E. Mickens, *Nonstandard Finite Difference Models of Differential Equations*. Singapore: World Scientific, 1994.
- [2] J. M. Sanz-Serna, "An unconventional simplistic integrator of W. Kahan," *Applied Numerical Math.*, vol. 16, pp. 245-250, 1994.
- [3] G.H. Erjaee and F. Dannan, "Stability analysis of periodic solutions to the nonstandard discretized model of the Lotka Volterra predator prey system," *International Journal of Bifurcation and Chaos*, vol. 14, pp. 4301-4308, 2004.
- [4] R. Buckmire, "Application of a Mickens finite-difference scheme to the cylindrical Bratu-Gelfand problem," *Numerical Methods for Partial Differential Equations*, vol. 19, pp. 380-398, 2003.
- [5] K. Gopalsamy, and P. Liu, *New Developments in Difference Equations and Applications*, Gordon and Breach, pp.207-216, 1999.
- [6] GG. Li, "Dynamical behaviors of Timoshenko beam with fractional derivative constitutive relations," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 3(1), pp. 67-73, 2002.
- [7] TL. Bo, L. Xie, and XI. Zheng, "Numerical approach to wind ripple in desert," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8(2), pp. 223-228, 2007.
- [8] H. Guan, CJ. Wu, H. Loewe, and et al., "Three-dimensional numerical study of flow structures of impinging jets in Y-typed micro-mixers," *Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8(3), pp. 425-434, 2007.
- [9] D. Matignon and B. D'andrea-Novel, "Observer-based controllers for fractional differential equations systems," in *Conference on Decision and Control San Diego, CA, December 1997, SIAM, IEEE-CSS*, pp. 4967-4972.
- [10] W.H. Deng, and W.H. Li, "Chaos synchronization of the fractional Lü System," *Physica A*, vol. 353, pp. 61-72, 2005.
- [11] M.M. Meerschaert and C. Tadjeran, "Finite difference approximations for two-sided space-fractional partial differential equations," *Appl. Numer. Math.*, vol. 56, pp. 80-90, 2006.
- [12] S. Momani, "Analytical approximate solution for fractional heat-like and wave-like equations with variable coefficients using the decomposition method," *Appl. Math. Comput.*, vol. 165(2), pp. 459-472, 2005.
- [13] S. Momani, "An explicit and numerical solutions of the fractional KdV equation," *Math. Comput. Simul.*, vol. 70(2), pp. 110-118, 2005.
- [14] S. Momani, "General solutions for the space and time fractional diffusion wave equation," *Journal of Physical Sciences*, accepted.
- [15] S. Momani and Z. Odibat, "Analytical approach to linear fractional partial differential equations arising in fluid mechanics," *Physics Letters A*, vol. 355, pp. 271-279, 2006.
- [16] Z. Odibat and S. Momani, "Numerical methods for solving nonlinear partial differential equations of fractional order," *Applied Mathematical modeling*, accepted.
- [17] S. Momani and Z. Odibat, "Comparison between homotopy perturbation method and the variational iteration method for linear fractional partial differential equations," *Computers and Math. Appl.*, accepted.
- [18] S. Momani and Z. Odibat, "Homotopy perturbation method for nonlinear partial differential equations of fractional order," *Physics Letters A*, accepted.
- [19] J.H. He, "Homotopy perturbation technique," *Comput. Meth. Appl. Mech. Eng.*, vol. 178, pp. 257-262, 1999.
- [20] J.H. He, "A coupling method of homotopy technique and perturbation technique for nonlinear problems," *Int. J. Non-Linear Mech.*, vol. 35(1), pp. 37-43, 2000.
- [21] L. Cveticanin, "Homotopy-perturbation method for pure nonlinear differential equation," *Chaos Solitons & Fractals*, vol. 30(5), pp. 1221-1230, 2006.
- [22] J.H. He, "New interpretation of homotopy perturbation method," *Int. Journal of Modern Physics B*, vol. 20(18), pp. 2561-2568, 2006.
- [23] A. Beléndez and T. Hernández "Application of He's homotopy perturbation method to the Duffing-harmonic oscillator," *Int. Journal of Non-linear Sciences and Numerical Simulation*, vol. 8(1), pp. 79-88, 2007.
- [24] D.H. Shou, "Application of parameter-expanding method to strongly nonlinear oscillators," *Int. Journal of Non-linear Sciences and Numerical Simulation*, vol. 8(1), pp. 121-124, 2007.
- [25] W.M. Ahmad and R. El-Khazali, "Fractional-order dynamical models of love," *Chaos, Solitons & Fractals*, in press.
- [26] X. Gao and J. Yu, "Synchronization of two coupled fractional-order chaotic oscillators," *Chaos, Solitons & Fractals*, vol. 26(1), pp. 141-145, 2005.
- [27] JG. Lu, "Chaotic dynamics and synchronization of fractional-order Arneodo's systems," *Chaos, Solitons & Fractals*, vol. 26(4), pp. 1125-1133, 2005.
- [28] JG. Lu and G. Chen, "A note on the fractional-order Chen system," *Chaos, Solitons & Fractals*, vol. 27(3), pp. 685-688, 2006.
- [29] P. Liu and S. Elaydi, "Discrete competitive and cooperative methods of Lotka-Volterra type," *Computational and Appl. Analysis*, vol. 3, pp. 53-73, 2001.
- [30] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," *Geophys. J. Roy. Astron. Soc.*, vol. 13, pp. 529-539, 1967.

- [31] M.M. Meerschaert and C. Tadjeran, "Finite difference approximations for fractional advection–dispersion flow equations," *J. Comput. Appl. Math.*, vol. 172, pp. 65–77, 2003.
- [32] T.S. Zhou and C.P. Li, "Synchronization in fractional-order differential systems," *Physica D*, vol. 212, pp. 111-115.
- [33] Y. Wang and C. Li, "Does the fractional Brusselator with efficient dimension less than 1 has a limit cycle? *Physics Letters A*, vol. 363, pp. 414–419, 2007.
- [34] PHASER, A universal simulator for dynamical systems, 2007, Available: <http://www.phaser.com>.

Complexified Dynamical Systems from Real Fractional Action Like With Time-Dependent Fractional Exponent on Multi-Fractals Sets

EL-NABULSI Ahmad Rami*

Abstract—Fractional actionlike variational approach with time-dependent dynamical fractional dimensions on multifractal sets of time and space is constructed. Fractional Euler-Lagrange equations are derived and discussed in some details. The case of harmonic oscillators with time-dependent mass and time-dependent frequency is explored and many interesting consequences are revealed, in particular the emergence of complexified dynamical systems from the real fractional action.

Index Terms—fractional action-like variational approach, multifractal sets, fractional Euler-Lagrange equations, complexified harmonic oscillators

I. INTRODUCTION

Fractional Calculus (FC) is a particular field of applied mathematics which dates back to the late part of seventeenth century and grows out of the conventional definitions of the calculus integral and derivative operators in much the similar way fractional exponents is a consequence of exponents with integer value. It is defined as a generalization of differentiation and integration to an arbitrary order with a broad range of applications in different research areas. It plays a vital role in the understanding of complex classical and quantum (conservative and dissipative) dynamical systems with holonomic as well as with nonholonomic constraints. Physicists and mathematicians have begun to investigate the territory of applications of fractional calculus with ever new developments speedily taking place in the field of statistical and nonlinear physics including hydrology (identification of the instantaneous unit hydrograph), seismic exploration (viscoelasticity), heat conduction and diffusion (super and subdiffusion with growth rate or shape of the particle distribution different than Gaussian), polymer physics, biophysics and thermodynamics, Brownian random walks with memory, modeling dispersion and turbulence, oscillating vortex chain, automatic control theory, transfer equation in a medium with fractal geometry, stochasting modeling for ultraslow diffusion, kinetic theories, far-from equilibrium statistical models manifesting scale invariance and scaling processes, dispersive transport in amorphous semiconductors, glasses, liquid crystals, biopolymers, proteins, biosystems, ecosystems, finance and economics, non-local correlations and extensive symmetry breaking, plasma physics, modeling mechanical and electrical properties of real materials, description of rheological properties of rocks, dynamics in complex media, wave propagation in complex and porous media, astrophysics, cosmology, quantum field theory, potential theory, financial time series, signal processing, fluctuations in solids, telecommunications, generation of artificial landscapes, reservoir engineering (poroelasticity), environmental geophysics (modeling of ground-penetrating radar), geophysical fluid dynamics (parameterization of turbulence in

meteorology and oceanography), to ecology and climatology (relationship between forests and greenhouse gases) and so on [1,2,3,4,5]. We refer the reader interested on fractional theory to the comprehensive book [2]. The FC is very rich and abundant applications and physical materializations have been found. However, these applications and the mathematical settings surrounding fractional calculus are far from ironic. While the physical significance is tricky to grasp, the fractional definitions themselves are no more rigorous than those of their integer order counterparts. The fractional derivatives and integrals portray more precisely the non-trivial behavior of complex physical systems whose dynamics are distant from equilibrium, i.e. dealing with fractional derivatives is not more complex than with usual differential operators. Dynamical equations in fractional derivatives describe normally the evolution of physical systems with loss, the fractional exponent of the derivative being a measure of the fraction of the states of the dynamical system that are preserved during the evolution time. Moreover, a growing body of empirical evidence supports the importance of fractional integral in quantum dynamics where important data series might be fractionally integrated. The definition of the fractional order derivative and integral are not unique where several definitions exist ranging from Grunwald-Letnikov fractional derivative, Caputo, Weyl, Feller, Erdelyi-Kober, Riesz fractional derivatives, etc., but the Riemann-Liouville (RL) and Caputo operators are still the most frequently used and have been popularized when fractional integration is performed.

The Fractional Calculus of Variations (FCV) based on fractional calculus was proved lately to be a practical device for description of physics beyond the standard model with holonomic as well as with nonholonomic constraints. Diverse forms of fractional Euler-Lagrange equations were obtained in literature depending on the action and type of fractional derivative used [6]. The main trouble with nearly all of these approaches is the presence of non-local fractional differential operators and the adjoint of a fractional differential operator used to describe the dynamics is not the negative of itself. Further, the derived fractional Euler-Lagrange equations depend on left and right fractional derivatives, even when the dynamics depend only on one of them. Other complicated problems occur during the mathematical manipulations as the manifestation of a very complicated Leibniz rule (the derivative of product of functions) and the absence of any fractional analogue of the chain rule. One successful method to model non-conservative field theories was proposed by the author in 2005 and is entitled the Fractional Action-Like Variational Approach (FALVA) and is based on the notion of left Riemann-Liouville fractional integral functionals with one parameter α but not on fractional-order derivatives of the same order [7,8]. The derived fractional Euler-Lagrange equations are analogous to the standard one but with the company of fractional external force acting on the physical system. Many encouraging results were obtained and discussed ranging from classical dynamics to quantum field theories and high energy

*Department of Nuclear and Energy Engineering, Cheju National University, Ara-dong 1, Jeju 690-756, South Korea
Personal Email: nabulsiyahmadrami@yahoo.fr

physics. It is note-worthy that the formulation of the fractional problems of the COV still needs more amplification as the problem is strongly linked to the fractional quantization process and to the presence of non-local fractional differential operators. Our main aim in the present work is to extend the FALVA in order to describe the dynamic characteristics of Lagrangian systems defined on multifractal time and space sets, in particular when the fractional dimensions of time and space are dynamical, e.g. $d_i = 1 + \varepsilon(\bar{x}_i)$. This notion is based on the ideas of the fractal geometry of nature and is expected to work on a small multifractal intervals set S_i which is build from multifractal subsets $S(\bar{x}_i)$ [9,10]. It is moreover connected to the generalization of the Riemann-Liouville fractional derivatives. Each of them consists in fact of a continuous, but not differentiable bounded set of small elements. Furthermore, both time and space are considered as the only material fields existing in the Universe and hence generating all other physical fields. In this work, we suppose that the fractal dimensions are slightly differs from unity, i.e. $|\varepsilon| \ll 1$ and hence valid for small densities of Lagrangians in points \bar{x}_i . The function $d_i = d(\bar{x}_i)$ is expected to be continuous and describes the fractional dimensions. The method and the theory are developed to describe dynamics of functions defined on multifractal sets of time and space with fractional dimensions and many interesting and appealing consequences are revealed [11-22]. It is remarkable that fractional calculus is based on rational numbers while fractals are irrational in general. As naturally rational numbers are just a special case, fractals and fractal spacetime came first. Historically it is also that way. It is striking that in fact, fractals in quantum mechanics are equally legitimate as in the classical theory.

The paper is organized as follows: In Sec. II, we construct the fractional Euler-Lagrange equations for the fractional action-like variational approach with time-dependent fractional dimension. In Sec. III, we discuss the real and complexified cases of harmonic oscillators. The paper concludes in Sec. IV with a brief summary of main results and future challenge and perspectives. We follow the rationale of [13] where it is assumed that at least one stationary point for the fractional functional exists. We introduce the main notations, conventions and assumptions that underlie the remainder of the present work

1. In the notation $t \rightarrow f(t)$, t is a dummy variable.
2. Exactly, the same function can be written, for example $(\dot{q}, q, \tau) \rightarrow f(\dot{q}, q, \tau)$; \dot{q}, q, τ are here dummy variables.
3. For $(\dot{q}, q, \tau) \rightarrow f(\dot{q}, q, \tau)$, the partial derivative of f with respect to the first argument is denoted by $\partial L / \partial \dot{q}$.
4. Following our previous work, we use in this paper the left fractional Riemann-Liouville integral which is the most widely used definition of an integral of fractional order is via an integral transform defined as

$${}_a I_t^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t f(\tau) (t-\tau)^{\alpha-1} d\tau, \quad 0 < \alpha < 1.$$

5. No fractional-order derivatives will be introduced.

II-FRACTIONAL EULER-LAGRANGE EQUATIONS WITH TIME-DEPENDENT FRACTIONAL DIMENSIONS

In order to build the theory, it is needed to characterize the functionals determined on the functions, given on a multifractal sets. For this, we suggest the following problem [7,8]:

Problem II-1: Find the stationary points of the integral functional on multifractal time and space sets

$$S_L[q] = \int_0^t L(\dot{q}(\tau), q(\tau), \tau) \frac{(t-\tau)^{\alpha(\tau)-1}}{\Gamma(\alpha(\tau))} d\tau = \int_0^t L(\dot{q}(\tau), q(\tau), \tau) \frac{\alpha(\tau)(t-\tau)^{\alpha(\tau)-1}}{\Gamma(1+\alpha(\tau))} d\tau \approx \int_0^t L(\dot{q}(\tau), q(\tau), \tau) \alpha(\tau) (t-\tau)^{\alpha(\tau)-1} d\tau, \quad (1)$$

under the initial condition $q(a) = q_a$.

Here $\dot{q} = dq/d\tau$, $\Gamma(\alpha(\tau))$ is the Euler gamma function defined by

$$\Gamma(\alpha(\tau)) = \int_0^\infty \tau^{\alpha(\tau)-1} \exp(-\tau) d\tau,$$

τ is the intrinsic time and t is the observer time, $t \neq \tau$. $n = \{\alpha_i\} + 1$ and $\{\alpha_i\}$ is the integer part of α_i with $\alpha_i \geq 0$ for $n-1 \leq \alpha_i < n$ and $n=0$ for $\alpha_i < 0$. The smooth Lagrangian function

$$L: [a, b] \times \square^n \times \square^n \rightarrow \square,$$

is a C^2 -function with respect to all its arguments. It is noteworthy that time-dependent fractional dimension plays a leading role in different branches of dynamical systems including self-affine time-sequential data [23], blast furnace iron making process and so on [24].

Remark II-1: In equation (1), we recognized the Riemann-Liouville operator of fractional integration in Lebesgue integrable space:

$${}_a I_t^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_a^t f(\tau) (t-\tau)^{\alpha-1} d\tau, \quad \Re(\alpha) > 0, \quad (2)$$

$$= \frac{d^n}{dt^n} {}_a I_t^{\alpha+n} f(t), \quad -n < \Re(\alpha) \leq 0, n \in \square. \quad (3)$$

Since we have defined fractional differentiation through integration, fractional derivatives are no longer local operations. They are defined over an interval. This may be why Leibnitz believed it was a paradox since his aspiration is to obtain a unique and local derivative. Moreover, because the Euler-Gamma function is defined for all $C \setminus \{0, -1, \dots\}$, we can still define fractional calculus of complex order, however, in this particular case the power function requires a branch cut typically chosen to be a ray originating from a and passing through the origin. Furthermore, due to the fact that the absolute continuity is known to be a sufficient condition to fractional integration, the space of functions chosen above corresponds to the absolutely continuous functions.

Remark II-2: The fractional operators are in reality global (non-locals) operators and limits in the sense of ultra-long time. For that main reason, dynamical systems with fractional order are non-conservatives and hence fractional calculus of variations is extensively used for describing intermediate physical processes and critical phenomena in non-equilibrium complex non-linear systems.

Theorem II-1: If $q(\square)$ are solutions to the previous problem, i.e. $q(\square)$ are critical points of the function (1), then $q(\square)$ satisfy the following fractional Euler-Lagrange equation:

$$\frac{d}{d\tau} \left(\frac{\partial L(\dot{q}(\tau), q(\tau), \tau)}{\partial \dot{q}_k} \right) + \left[\frac{d\alpha(\tau)}{d\tau} \ln(t-\tau) + \frac{\alpha(\tau)-n}{\tau-t} \right] \frac{1}{\Gamma(\alpha(\tau))} \frac{d}{d\tau} \Gamma(\alpha(\tau)) \left[\frac{\partial L(\dot{q}(\tau), q(\tau), \tau)}{\partial \dot{q}_k} - \frac{\partial L(\dot{q}(\tau), q(\tau), \tau)}{\partial q_k} \right] = 0. \quad (4)$$

Proof: We let $q_k^0(\tau)$ be the minimum solution and write $q_k = q_k^0 + \sigma_k$ where $\sigma_k(\tau)$ describes the deviation of $q_k(\tau)$ from the minimum path $q_k^0(\tau)$. Now insert into equation (2) gives:

$$S_L[q] = \int_{t_0}^t L(\dot{q}_k^0(\tau) + \dot{\sigma}_k(\tau), q_k^0(\tau) + \sigma_k(\tau), \tau) \frac{(t-\tau)^{\alpha(\tau)-n}}{\Gamma(\alpha(\tau))} d\tau$$

Performing Taylor expansions to first order in $\dot{\sigma}_k(\tau)$ and $\sigma_k(\tau)$, yields:

$$S_L[q] = \int_{t_0}^t \left\{ L(\dot{q}_k^0(\tau), q_k^0(\tau), \tau) + \frac{\partial L}{\partial \dot{q}_k} \dot{\sigma}_k(\tau) + \frac{\partial L}{\partial q_k} \sigma_k(\tau) \right\} \frac{(t-\tau)^{\alpha(\tau)-n}}{\Gamma(\alpha(\tau))} d\tau$$

After integrating easily the term in $\sigma(\tau)$ by parts, we obtain easily:

$$S_L[q] = \int_{t_0}^t (\dot{q}_k^0(\tau), q_k^0(\tau), \tau) \frac{(t-\tau)^{\alpha(\tau)-n}}{\Gamma(\alpha(\tau))} d\tau - \int_{t_0}^t \sigma(\tau) \left[\frac{(t-\tau)^{\alpha(\tau)-n}}{\Gamma(\alpha(\tau))} \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{q}_k} \right) + (t-\tau)^{\alpha(\tau)-n} \frac{d}{d\tau} \left(\frac{1}{\Gamma(\alpha(\tau))} \right) \frac{\partial L}{\partial q_k} \right] d\tau + \frac{1}{\Gamma(\alpha(\tau))} \left[\frac{d\alpha}{d\tau} \ln(t-\tau) + \frac{\alpha(\tau)-n}{\tau-t} \right] (t-\tau)^{\alpha(\tau)-n} \left[\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} \frac{(t-\tau)^{\alpha(\tau)-n}}{\Gamma(\alpha(\tau))} \right] d\tau$$

from which we get the required results.

Remark II-3: We may substitute into equation (4) the third term inside the bracket by the digamma function which is defined as the logarithmic derivative of the gamma function as [25]:

$$\frac{1}{\Gamma(\alpha(\tau))} \frac{d}{d\tau} \Gamma(\alpha(\tau)) = \frac{d}{d\tau} \ln \Gamma(\alpha(\tau)) = \psi(\tau), \quad (5)$$

which it is the first of the polygamma functions and has the integral representation:

$$\psi(\tau) = -\gamma - \frac{1}{\tau} + \tau \sum_{k=1}^{\infty} \frac{1}{k(\tau+k)}. \quad (6)$$

where γ is the Euler-Mascheroni constant defined by:

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \log n \right) \approx 0.57721566490153286... \quad (7)$$

Corollary II-1: The fractional Euler-Lagrange equation in terms of the polygamma function is written as follows:

$$\frac{\partial L(\dot{q}(\tau), q(\tau), \tau)}{\partial q_k} - \frac{d}{d\tau} \left(\frac{\partial L(\dot{q}(\tau), q(\tau), \tau)}{\partial \dot{q}_k} \right) = \left[\frac{d\alpha(\tau)}{d\tau} \ln(t-\tau) + \frac{\alpha(\tau)-n}{\tau-t} - \psi(\tau) \right] \frac{\partial L(\dot{q}(\tau), q(\tau), \tau)}{\partial \dot{q}_k} \square F_i^{Gen}. \quad (8)$$

Here F_i^{Gen} is the extended modified frictional force which is a common type of non conservative force. It is note-worthy that for $\alpha(\tau) = \alpha = \text{constant}$ and $\alpha = n = 1$, equation (8) reduces to the standard Euler-Lagrange equation.

Remark II-4: For a one-degree of freedom non-conservative dynamical system described by the Lagrangian

$$L(\dot{q}(\tau), q(\tau), \tau) = \frac{1}{2} \dot{q}^2(\tau) - V(q(\tau)),$$

the equation of motion is clearly given by:

$$\ddot{q}(\tau) + \left[\frac{d\alpha(\tau)}{d\tau} \ln(t-\tau) + \frac{\alpha(\tau)-n}{\tau-t} - \psi(\tau) \right] \dot{q}(\tau) = -\frac{\partial V}{\partial q}. \quad (9)$$

III-FROM REAL TO COMPLEXIFIED HARMONIC OSCILLATORS

In order to exemplify, we discuss the harmonic oscillator case. In fact, for the case of a harmonic oscillator with quadratic Lagrangian with time-dependent mass and time-dependent frequency, i.e.

$$L(\dot{q}(\tau), q(\tau), \tau) = \frac{1}{2} m(\tau) \dot{q}^2 - \frac{1}{2} m(\tau) \omega^2(\tau) q^2,$$

equation (8) is reduced to:

$$\ddot{q}(\tau) + \left[\frac{d\alpha(\tau)}{d\tau} \ln(t-\tau) + \frac{\alpha(\tau)-n}{\tau-t} - \psi(\tau) + \frac{1}{m} \frac{dm}{d\tau} \right] \dot{q}(\tau) + \omega^2 q(\tau) = 0, \quad (10)$$

which may be rewritten like:

$$\ddot{q}(\tau) + \left[\frac{d}{d\tau} (\ln(t-\tau)^{\alpha(\tau)-n} m(\tau)) - \psi(\tau) \right] \dot{q}(\tau) + \omega^2 q(\tau) = 0. \quad (11)$$

If for instance $m(\tau) \propto [\ln(t-\tau)^{\alpha(\tau)-n}]^{-1} \psi(\tau) d\tau$, equation (11) is reduced to:

$$\ddot{q}(\tau) + \omega^2(\tau) q(\tau) = 0. \quad (12)$$

It is interesting to have a harmonic oscillator with time-dependent mass $m(\tau)$ function of the dynamical fractional dimension. It is noteworthy that if for instance $t < \tau$, we may substitute the logarithmic term inside equation (11) by

$$\ln(t-\tau) = \ln(-T) = \ln i^2 T = i\pi + \ln T$$

($T = \tau - t > 0$ and $i = \sqrt{-1}$) and therefore equation (11) takes the particular form:

$$\ddot{q}(\tau) + \left[\frac{d}{d\tau} \left((i\pi + \ln T)^{\alpha(\tau)-n} m(\tau) \right) - \psi(\tau) \right] \dot{q}(\tau) + \omega^2 q(\tau) = 0, \quad (13)$$

with $\dot{q}(T) = dq/dT$ and consequently for

$$m(\tau) = (i\pi + \ln T)^{n-\alpha(\tau)} \ln \Gamma(\alpha(\tau)),$$

equation (12) holds. This mass is complexified in the sense that one can rewrite it in the form:

$$m(\tau) = \ln T^{n-\alpha(\tau)} \ln \Gamma(\alpha(\tau)) + i\pi \ln \Gamma(\alpha(\tau)). \quad (14)$$

It is as well appealing to have a complexified harmonic oscillator mass emerging from a non-complexified action. This fact could have motivating consequences in high energy physics and gauge field theories [26]. If in contrast $m(\tau) \propto \ln(t-\tau)^{n-\alpha(\tau)}$, then equation (11) is reduced to:

$$\ddot{q}(\tau) + 2 \frac{\dot{\eta}}{\eta} \dot{q}(\tau) + \omega^2 q(\tau) = 0, \quad (15)$$

where $\eta^{-2}(\tau) \propto \Gamma(\alpha(\tau))$. The term $\eta(\tau)$ may be identified to the square-root of an effective mass $M(\tau)$, i.e. $\eta(\tau) = \sqrt{M(\tau)}$. Hence $M(\tau) = \eta^2(\tau) = \Gamma^{-1}(\alpha(\tau))$. It is an easy exercise to prove that the solution is given by [19,27]:

$$x(\tau) = \rho(\tau) [A \cos(\beta(\tau)) + B \sin(\beta(\tau))], \quad (16)$$

where $\rho(\tau) = \chi(\tau)/\eta(\tau)$ and $\beta(\tau)$ refers to the phase and amplitude of the oscillators and (A, B) are constants which can be determined by imposing the conditions $x(\tau') = x'$ and $x(\tau'') = x''$, then:

$$\eta(\ddot{\chi} - \chi \dot{\beta}^2) + \chi(\omega^2(\tau) \eta - \ddot{\eta}) = 0, \quad (17)$$

$$\ddot{\beta} + 2 \frac{\dot{\chi} \dot{\beta}}{\chi} = 0. \quad (18)$$

It is noteworthy that the wave function for a harmonic oscillator with time-dependent mass $m = m(\tau)$ and frequency $\omega = \omega(\tau)$ described by the fractional Euler-Lagrange equations is given by [28]:

$$\psi_n(x, \tau) = \sqrt{\frac{1}{2^n n!} \sqrt{\frac{m(\tau) \dot{\beta}(\tau)}{\pi \hbar}}} \exp\left(-i \left(n + \frac{1}{2}\right) \beta(\tau)\right) \times \exp\left(\frac{im(\tau)}{2\hbar} \left\{ \frac{\dot{\chi}(\tau)}{\chi(\tau)} - \frac{\dot{\eta}(\tau)}{\eta(\tau)} + i\dot{\beta}(\tau) \right\} x^2\right) \times H_n \left(\sqrt{\frac{m(\tau) \dot{\beta}(\tau)}{\hbar}} x \right). \quad (19)$$

H_n being the Hankel function. The fractional evolution of strongly pulsating mass and quantum damped harmonic oscillator as well as their stationary states could then be deduced from our fractional action-like variational approach.

Remark III-1: In fact, using the substitution $\bar{q}(\tau) = q(\tau)\eta(\tau)$ with $\eta(\tau) = \sqrt{M(\tau)} = \Gamma^{-1/2}(\alpha(\tau))$, equation (15) takes the form:

$$\ddot{\bar{q}}(\tau) + \left[\omega^2 + \frac{1}{4} \left(\frac{\dot{M}(\tau)}{M(\tau)} \right)^2 - \frac{1}{2} \frac{\dot{M}(\tau)}{M(\tau)} \right] \bar{q}(\tau) = 0, \quad (20)$$

which can be rewritten in the form:

$$\ddot{q}(\tau) + \left[\omega^2 + \frac{1}{4} \Gamma^2(\alpha(\tau)) \left(\frac{d}{d\tau} \frac{1}{\Gamma(\alpha(\tau))} \right)^2 - \frac{1}{2} \Gamma(\alpha(\tau)) \frac{d^2}{d\tau^2} \left(\frac{1}{\Gamma(\alpha(\tau))} \right) \right] \ddot{q}(\tau) = 0. \quad (21)$$

Hence we have a particular harmonic oscillator with a time-dependent frequency $\Omega(\tau)$ identified to:

$$\begin{aligned} \Omega^2(\tau) &\equiv \omega^2 + \frac{1}{4} \left(\frac{\dot{M}(\tau)}{M(\tau)} \right)^2 - \frac{1}{2} \frac{\ddot{M}(\tau)}{M(\tau)} \\ &= \omega^2 + \frac{1}{4} \Gamma^2(\alpha(\tau)) \left(\frac{d}{d\tau} \frac{1}{\Gamma(\alpha(\tau))} \right)^2 - \frac{1}{2} \Gamma(\alpha(\tau)) \frac{d^2}{d\tau^2} \left(\frac{1}{\Gamma(\alpha(\tau))} \right). \end{aligned} \quad (22)$$

The general solution is given by [19,27]

$$\ddot{q}(\tau) = A(\tau) [C \cos(\sigma(\tau) + D \sin(\sigma(\tau))), \quad (23)$$

C and D are constants which may be determined by using two conditions on the classical trajectory and where the amplitude $A(\tau)$ and the phase $\sigma(\tau)$ satisfy equations

$$A^3(\tau) \ddot{A}(\tau) + \omega^2(\tau) A^4(\tau) = E^2, \quad (24)$$

$$\dot{\sigma}(\tau) A^2(\tau) = E, \quad (25)$$

$0 < E \in \square$ and can be taken equal to one.

Accordingly, we have a fractional description of a harmonic oscillator with time-dependent mass and time-dependent frequency exhibiting diverse types of behavior and which may be practical in quantum dynamical systems. Generally the quantum-mechanical behavior of various damped harmonic oscillators is done making use of the path-integral method and second quantization methods. The fractional approach elaborated here offers a novel approach.

More exemplification could be done if we consider merely a simple pendulum of length l attached to the circumference of a body of small radius and mass M assumed to be time-dependent. The linear kinetic energy is $K(\theta) = \frac{1}{2} M(\tau) l^2 \dot{\theta}^2$ and the potential energy for small oscillations is $V(\theta) = \frac{1}{2} M(\tau) g l \theta^2$. Here θ is the angular coordinate. The Lagrangian of the dynamical system is then given by $L = K - V$. As a result, equation (8) yields the following non-linear differential equation:

$$\ddot{\theta}(\tau) + \left(\frac{d\alpha(\tau)}{d\tau} \ln(t-\tau) + \frac{\alpha(\tau)-n}{\tau-t} - \psi(\tau) \right) \dot{\theta}(\tau) + \omega^2 \theta(\tau) = 0, \quad (25)$$

which for $\alpha(\tau) = n - \ln(t-\tau)$, takes the special form:

$$\ddot{\theta}(\tau) - \psi(\tau) \dot{\theta}(\tau) + \omega^2 \theta(\tau) = 0. \quad (26)$$

where time-derivative is done with respect to τ , $\omega^2 = g/l$ and g being the gravity constant. Equation (25) may be identified to equation (15) if for instance $\eta^2(\tau) = \Gamma^{-1}(\alpha(\tau))$.

The term $\eta(\tau)$ may be identified here to the square-root of an effective complexified mass $\tilde{M}(\tau)$, i.e. $\eta(\tau) = \sqrt{-\tilde{M}(\tau)}$. Hence $M(\tau) = -\eta^2(\tau) = -\Gamma^{-1}(\alpha(\tau))$.

Remark III-2: Making use of the substitution $\ln(t-\tau) = i\pi + \ln T$, we may rewrite equation (8) like:

$$\begin{aligned} &\frac{\partial L(\dot{q}(T), q(T), T)}{\partial q_k} - \frac{d}{dT} \left(\frac{\partial L(\dot{q}(T), q(T), T)}{\partial \dot{q}_k} \right) \\ &= \left[(i\pi + \ln T) \frac{d\alpha(T)}{dT} + \frac{\alpha(T)-n+1}{T} + \gamma - T \sum_{k=1}^{\infty} \frac{1}{k(T+k)} \right] \frac{\partial L(\dot{q}(T), q(T), T)}{\partial \dot{q}_k}. \end{aligned} \quad (27)$$

The presence of the complex number inside equation (26) motivates us to propose the following conjecture: "In fact, we may start with a real Lagrangian system with N -degrees of freedom and a Lagrangian depending analytically on the dynamical variables, then complexified the system and

subsequently reconsidered as a Lagrangian system with $2N$ -real degrees of freedom".

For this, we make the dependent variable complex and write $q(\tau) = q_1(\tau) + i q_2(\tau)$. Naturally, under complexification, the potential becomes complex, i.e. $V(q_1, q_2) = V_1(q_1, q_2) + i V_2(q_1, q_2)$ where $V_1(q_1, q_2)$ and $V_2(q_1, q_2)$ satisfy the Cauchy-Riemann condition:

$$\frac{\partial V_1}{\partial q_1} = \frac{\partial V_2}{\partial q_2}, \quad (28)$$

$$\frac{\partial V_1}{\partial q_2} = -\frac{\partial V_2}{\partial q_1}. \quad (29)$$

Similarly the Lagrangian takes the special form:

$$L(q) \rightarrow L_1 + i L_2 = \left[\frac{1}{2} (\dot{q}_1^2 + \dot{q}_2^2) - V_1(q_1, q_2) \right] + i \left[\dot{q}_1 \dot{q}_2 - V_2(q_1, q_2) \right], \quad (30)$$

with

$$\frac{\partial}{\partial q} = \frac{1}{2} \left(\frac{\partial}{\partial q_1} - i \frac{\partial}{\partial q_2} \right). \quad (31)$$

Therefore equation (27) is splitted into two equations (real and complex parts respectively):

$$\frac{\partial L}{\partial q_1} - \frac{d}{dT} \left(\frac{\partial L}{\partial \dot{q}_1} \right) = \left(\ln T \frac{d\alpha(T)}{dT} + \frac{\alpha(T)-n}{T} - \psi(T) \right) \frac{\partial L}{\partial q_1} + \pi \frac{d\alpha(T)}{dT} \frac{\partial L}{\partial q_2}, \quad (32)$$

$$\frac{\partial L}{\partial q_2} - \frac{d}{dT} \left(\frac{\partial L}{\partial \dot{q}_2} \right) = \left(\ln T \frac{d\alpha(T)}{dT} + \frac{\alpha(T)-n}{T} - \psi(T) \right) \frac{\partial L}{\partial q_2} - \pi \frac{d\alpha(T)}{dT} \frac{\partial L}{\partial q_1}. \quad (33)$$

To illustrate, we reconsider the simple pendulum of length l attached to the boundary of a body of small radius and mass M assumed to be time-independent. The linear kinetic energy is $K(\theta) = \frac{1}{2} M l^2 \dot{\theta}^2$ and the potential energy for small oscillations is $V(\theta) = \frac{1}{2} M g l \theta^2$. Here $\theta = \theta_1 + i \theta_2$ is the angular coordinate. Hence we are dealing with a complexified harmonic oscillator model. It is noteworthy that complexified harmonic oscillators play a crucial role in Crypto-gauge invariant models related to \mathcal{PT} -symmetric models [29], e.g. Lotka-Volterra predator-prey model and the Euler equations for the free rotation of a rigid body [30]. In fact, there exist quantum mechanical models with specific complex terms in the Hamiltonian that admit real spectra and unitary evolution. In a general context, it was observed that the real part of the Hamiltonian can generate the dynamics in a real phase space and that the imaginary part, treated as a constraint, can generate symmetry transformation. The Lagrangian of the dynamical system is then given by:

$$L = K - V = \frac{1}{2} M \left[l(\dot{\theta}_1^2 - \dot{\theta}_2^2) - g(\theta_1^2 - \theta_2^2) \right] + i M \left[l^2 \dot{\theta}_1 \dot{\theta}_2 - g \theta_1 \theta_2 \right]. \quad (34)$$

As a result, equations (32) and (33) yield the following non-linear differential equations:

$$\ddot{\theta}_1 + \left(\ln T \frac{d\alpha(T)}{dT} + \frac{\alpha(T)-n}{T} - \psi(T) \right) \dot{\theta}_1 - \pi \frac{d\alpha(T)}{dT} \dot{\theta}_2 + \omega^2 \theta_1 = 0, \quad (35)$$

$$\ddot{\theta}_2 + \left(\ln T \frac{d\alpha(T)}{dT} + \frac{\alpha(T)-n}{T} - \psi(T) \right) \dot{\theta}_2 + \pi \frac{d\alpha(T)}{dT} \dot{\theta}_1 + \omega^2 \theta_2 = 0. \quad (36)$$

Here $\dot{\theta} = d\theta/dT$. Numerical analyses of these coupled non-linear differential equations are under progress. However, it is remarkable that, if for instance, $\alpha(T) = QT^{1-m}$, $(m, Q) \in \square$, then equations (35) and (36) may be rewritten as:

$$\ddot{\theta}_1 + \left((1-m)Q \frac{\ln T}{T^m} + \frac{Q}{T^{m+2}} - \frac{n}{T} - \psi(T) \right) \dot{\theta}_1 - \frac{\pi Q(1-m)}{T^m} \dot{\theta}_2 + \omega^2 \theta_1 = 0, \quad (37)$$

$$\ddot{\theta}_2 + \left((1-m)Q \frac{\ln T}{T^m} + \frac{Q}{T^{m+2}} - \frac{n}{T} - \psi(T) \right) \dot{\theta}_2 + \frac{\pi Q(1-m)}{T^m} \dot{\theta}_1 + \omega^2 \theta_2 = 0. \quad (38)$$

Therefore, for $0 < m < 1$ which corresponds for an increasing fractional dimension with time, equations (37) and (38) are simplified for very large time to:

$$\ddot{\theta}_1 + \left((1-m)Q \frac{\ln T}{T^m} - \frac{n}{T} - \psi(T) \right) \dot{\theta}_1 + \omega^2 \theta_1 = 0, \quad (39)$$

$$\ddot{\theta}_2 + \left((1-m)Q \frac{\ln T}{T^m} - \frac{n}{T} - \psi(T) \right) \dot{\theta}_2 + \omega^2 \theta_2 = 0. \quad (40)$$

If, in contrast, $m > 1$ which corresponds for a decreasing fractional dimension with time, then equations (37) and (38) are simplified for very large time to:

$$\ddot{\theta}_1 - \left(\frac{n}{T} + \psi(T) \right) \dot{\theta}_1 + \omega^2 \theta_1 = 0, \quad (41)$$

$$\ddot{\theta}_2 - \left(\frac{n}{T} + \psi(T) \right) \dot{\theta}_2 + \omega^2 \theta_2 = 0. \quad (42)$$

The previous equations require numerical solutions and it is the author's speculation that they may play important role on complexified oscillatory dynamical systems with time-dependent mass and time-dependent frequency. Work in this direction is also under progress.

IV-CONCLUSIONS AND PERSPECTIVES

To the best of our knowledge, this work represents the first attempt to explore the complexified Euler-Lagrange equations for a dynamical systems starting from a real fractional action-like integral approach. That will be the beginning, we believe, of new exciting investigations. Our contribution is, however, only theoretical and, in that sense, more modest. The emerging complexified fractional theory is still an open problem under development, and up to this point, many of the theories and formulations presented in this work are only mathematical exercises. We anticipate they will open up in the future a new stimulating research area in different branch of mathematical physics and provide us by a powerful tool to understand many fundamental problems in the area of complexified dynamical systems. Future research efforts may be directed towards formulating predictions that can be tracked tested numerically. Since in our framework we argued that time-dependent fractional dimensions could have appealing and interesting consequences in classical dynamical systems, it follows naturally that it should also attractive features in quantum field theory with complexified gauge as by itself, Nature is fractal. Therefore, it is not a big surprise that quantum spacetime and consequently quantum field theory (in particular the particle physics classification) could be as well and supplementary accompanied with a time-dependent fractional dimension. For all these stated reasons, it seems for us indispensable to incorporate in a future work the fractional formalism based on fractional operators and in particular the fractional problems of the Calculus of Variations with time-dependent fractional dimensions within the context of chaos, fractals, scale relativity [31], a Cantorian-fractal spacetime [32] and complexification of gauge theories [33]. Concurrent research efforts are needed to confirm or falsify, develop or disprove the fractional dynamics discussed here including our preliminary findings.

Acknowledgments: I would like to thanks Professor Z. M. Odibat from the Faculty of Science and Technology, University of Le Havre, France for inviting me to participate on the 3rd International Conference on Complex Systems and Applications, in particular on the special session "Chaos and synchronization of fractional order systems".

REFERENCES

[1] K. B. Oldham, J. Spanier, *The fractional calculus* (New York., London, Acad. Press., 1974).

- [2] S. Samko, A. Kilbas and O. Marichev, *Fractional integrals and derivatives: Theory and applications*, (Gordon and Breach, New York, 1993).
- [3] K. S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations*, (John Wiley & Sons Inc., New York, 1993).
- [4] I. Podlubny, *An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of their Solution and some of their Applications*, (Academic Press, New York-London, 1999).
- [5] R. Hilfer, Editor, *Applications of Fractional Calculus in Physics*, (World Scientific Publishing Co., New Jersey, London, elmh6b Hong Kong, 2000).
- [6] M. Klimek, *Czechoslovak J. Phys.* 55(11), (2005) 1447 and references therein.
- [7] R. A. El-Nabulsi, *Fizika* A14, 4, (2005) 289.
- [8] R. A. El-Nabulsi, *Int. J. Appl. Math.* 17, No 3, (2005) 299.
- [9] L. Ya. Kobelev, arXiv.org: physics/0001035.
- [10] L. Ya. Kobelev, arXiv.org: gr-qc/00010543.
- [11] R. A. El-Nabulsi, *Fractional dynamics, fractional weak bosons masses and physics beyond the standard model*, *Chaos, Solitons and Fractals* (in press). doi:10.1016/j.chaos.2008.08.033.
- [12] R. A. El-Nabulsi and D. F. M. Torres, *J. Math. Phys.* 49, (2008) 053521.
- [13] R. A. El-Nabulsi, *Int. J. Geom. Meth. Mod. Phys.* 5, 6 (2008), 863 and references therein.
- [14] R. A. El-Nabulsi, *Int. J. Geom. Meth. Mod. Phys.* 6, 1, (2009), 27.
- [15] R. A. El-Nabulsi and D. F. M. Torres, *Math. Meth. Appl. Sci.* 30, 15, (2007) 1931.
- [16] R. A. El-Nabulsi, *Fizika* B17, 3 (2008), 369.
- [17] R. A. El-Nabulsi, *Fractional action-like variational problems in holonomic, non-holonomic and semi-holonomic constrained and dissipative dynamical systems*, *Chaos, Solitons and Fractals* (in press). doi:10.1016/j.chaos.2008.10.022.
- [18] R. A. El-Nabulsi, *On the fractional minimal length Heisenberg-Weyl uncertainty relation from fractional Riccati generalized momentum operator*, *Chaos, Solitons and Fractals* (in press). doi:10.1016/j.chaos.2008.10.031.
- [19] R. A. El-Nabulsi, *The fractional calculus of variations from extended Erdelyi-Kober operator*, *Int. J. Mod. Phys. B* (in press).
- [20] R. A. El-Nabulsi, *Fizika* A17, 2 (2008), 71.
- [21] R. A. El-Nabulsi, *Fizika* A16, 3 (2007), 137.
- [22] R. A. El-Nabulsi, *Fractional quantum Euler-Cauchy equation in the Schrödinger picture, complexified harmonic oscillators and emergence of complexified Lagrangian and Hamiltonian dynamics*, *Mod. Phys. Lett. B* (accepted for publication, in press).
- [23] S. Salvador and N. Masahiro, *J. Phys. Soc. Japan*, 64, 9 (1995) 3226.
- [24] Z.-M. Zhou, *Physica* A376, (2006) 133.
- [25] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs and Mathematical Tables*, Dover Publications (1965).
- [26] M. Omar and M. S. Habib, *Int. J. Theor. Phys.* 47, 4 (2008) 1112
- [27] G. S. Djordjevic and B. Dragovich, quant-ph/0005027.
- [28] A. Valtakoski, *Mathematical aspects of functional integration*, HELSINGIN YLIOPISTO FYSIIKAN LAITOS TEOREETTISEN FYSIIKAN OSASTO PL 9 (Siltavuorenpenger 20 D) 00014 Helsingin Yliopisto (2000).
- [29] S. Ghosh and B. R. Majdi, *J. Phys.* A41, (2008) 065306.
- [30] C. M. Bender, D. D. Holm and D. W. Hook, *J. Phys.* A40, (2008) F793 and references therein
- [31] L. Nottale, *Fractal space-time and micro physics*, Singapore: World Scientific (1993).
- [32] M. S. El Naschie, *Chaos Solitons and Fractals* 1, 5, (1992) 485.
- [33] R. Loll, J. M. Mourao and J. N. Tavares, *J. Geom. Phys.* 18, (1996) 1.

Designing modified projective synchronization for fractional order chaotic systems

Runzi Luo, Shucheng Deng, Zhengmin Wei

Abstract—In this paper, we investigate the modified projective synchronization of fractional order chaotic systems between the drive and response systems. A new method for constructing modified projective synchronization of fractional order systems is presented based on the stability criterion of linear systems. Numerical simulations of modified projective synchronization between the fractional order Lorenz chaotic system and fractional order Lü-chen chaotic system show the effectiveness of the proposed method.

Index Terms—Keywords: Modified projective synchronization; Fractional order; Lorenz chaotic system

I. INTRODUCTION

FRACTIONAL calculus is a generalization of ordinary (integer order) integration and differentiation to its non-integer (fractional) order counterpart. It has many applications to physics, engineering and control processing [1-3]. Many systems in interdisciplinary fields, such as dielectric polarization, electrode-electrolyte polarization, electromagnetic waves and viscoelastic systems are known to display fractional order dynamics[3]. It is known that some fractional-order differential systems behave chaotically, for example, the fractional-order Duffing system [4], the fractional-order Chua system [5], the fractional-order Chen system [6], the fractional-order Lü system [7], the fractional-order unified system [8]. Recently, due to its potential applications in secure communication and control processing, synchronization of chaotic fractional systems starts to attract increasing attention [9-16]. However, most of research efforts mentioned above have concentrated on studying complete synchronization (CS), generalized projective synchronization where the drive and response systems could be synchronized up to a scaling factor α .

In the present paper, we consider modified projective synchronization (MPS), where the responses of the synchronized dynamical states synchronize up to a constant scaling matrix. Based on the stability criterion of linear systems, a new approach for constructing modified projective synchronization of fractional order systems is attained. Numerical simulations show the effectiveness of the proposed method.

The rest of this paper is organized as follows. In Section 2 the definition of modified projective synchronization is introduced and the synchronization criterion is given. In Section 3, numerical simulation is given to show the effectiveness of the proposed method. Finally, some conclusions are given in Section 4.

R. Z. Luo, S. C. Deng and Z. M. Wei are with the Department of Mathematics, Nanchang University, 330031, P. R. China, E-mail: luo_rz@163.com

II. MODIFIED PROJECTIVE SYNCHRONIZATION SYSTEMS DESIGN

First, we review the fractional derivatives. Many authors formally use the Riemann-Liouville fractional derivatives, defined by

$$D^\alpha x(t) = \frac{d^m}{dt^m} J^{m-\alpha} x(t), \alpha > 0,$$

where $m = \lceil \alpha \rceil$, i.e., m is the first integer which is not less than α . J^β is the β -order Riemann-Liouville integral operator with expression:

$$J^\beta y(t) = \frac{1}{\Gamma(\beta)} \int_0^t (t-\tau)^{\beta-1} y(\tau) d\tau, \beta > 0,$$

here Γ stands for Gamma function. In this paper, the following definition is used

$$D^\alpha x(t) = J^{m-\alpha} x^{(m)}(t), \alpha > 0,$$

where $m = \lceil \alpha \rceil$, and the operator D^α is generally called " α -order Caputo differential operator".

Now, we introduce the modified projective synchronization(MPS) of fractional order chaotic systems. For simplicity, we take the notation $D^q(\cdot)$, which is described as $D^q x(t) = \begin{pmatrix} D^{q_1} x_1(t) \\ D^{q_2} x_2(t) \\ \dots \\ D^{q_n} x_n(t) \end{pmatrix}$, where $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$, and $q = \text{diag}(q_1, q_2, \dots, q_n)^T$, $0 < q_i \leq 1, i = 1, 2, \dots, n$.

Consider the following chaotic systems:

$$\begin{cases} D^q x = f(x, t) \leftarrow \text{drive system,} \\ D^q y = g(x, y, t) \leftarrow \text{response system,} \end{cases} \quad (1)$$

where $y = (y_1, y_2, \dots, y_n)$. If there exists a constant matrix $\alpha = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)$, such that $\lim_{t \rightarrow +\infty} \|x - \alpha y\| = 0$, then we regard that y synchronize to x up to scaling matrix α , and call such synchronization "modified projective synchronization". Obviously, CS and generalized projective synchronization are the special cases of MPS where $\alpha_1 = \alpha_2 = \dots = \alpha_n = 1$ and $\alpha_1 = \alpha_2 = \dots = \alpha_n$, respectively.

Assume the fractional order chaotic drive system under study can be written as:

$$D^q x = \frac{d^q x}{dt^q} = f(x, t) = Ax + [f(x, t) - Ax] = Ax + h(x, t) \quad (2)$$

where $x \in R^n$, $f : R^n \rightarrow R^n$ and $h : R^n \rightarrow R^n$ are nonlinear functions, A is an $n \times n$ constant diagonal matrix, i.e., $A = \text{diag}(a_1, a_2, \dots, a_n)$, where $a_i < 0, i = 1, 2, \dots, n$.

Suppose the controlled response system is

$$D^q y = \frac{d^q y}{dt^q} = g(y, t) + u = Ay + [g(y, t) - Ay] + u = Ay + S(y, t) + u, \quad (3)$$

where $y \in R^n$ and $S : R^n \rightarrow R^n$ is nonlinear functions, u is a controller to be designed below. Then we have the following main result.

Theorem 1. For an invertible diagonal matrix α , if $u = \alpha^{-1}h(x, t) - S(y, t)$, then modified projective synchronization between systems (2) and (3) will occur.

Proof. Define the error $e = x - \alpha y$ between systems (2) and (3). Then one can obtain $D^q e = \frac{d^q e}{dt^q} = \frac{d^q(x - \alpha y)}{dt^q} = A(x - \alpha y) + \alpha S(y, t) + \alpha u - R(x, t) = Ae$. Thus the above equation can be rewritten as

$$\begin{cases} D^{q_1} e_1 = a_1 e_1, \\ D^{q_2} e_2 = a_2 e_2, \\ \dots \\ D^{q_n} e_n = a_n e_n, \end{cases}$$

For equation $D^{q_i} e_i = a_i e_i, i = 1, 2, \dots, n$, since a_1 is negative, according to the stability criterion of linear system we have $\lim_{t \rightarrow +\infty} e(t) = 0$. Therefore the state vectors $x(t)$ and $y(t)$ of different systems (2) and (3) are modified projective synchronized.

III. MODIFIED PROJECTIVE SYNCHRONIZATION BETWEEN FRACTIONAL ORDER LORENZ CHAOTIC SYSTEM AND FRACTIONAL ORDER Lü-CHEN SYSTEM

The Lü-chen system was first introduced by Lü and Chen[17], which can be described by three-dimensional quadratic autonomous ordinary differential equations and can simultaneously display two 1-scroll chaotic attractor with only three equilibria, and two 2-scroll chaotic attractors with five equilibria. This system is given by

$$\begin{cases} \dot{x} = -\frac{ab}{a+b}x - yz + c, \\ \dot{y} = ay + xz, \\ \dot{z} = bz + xy, \end{cases} \quad (4)$$

where a, b, c are real constants and x, y, z are state variables.

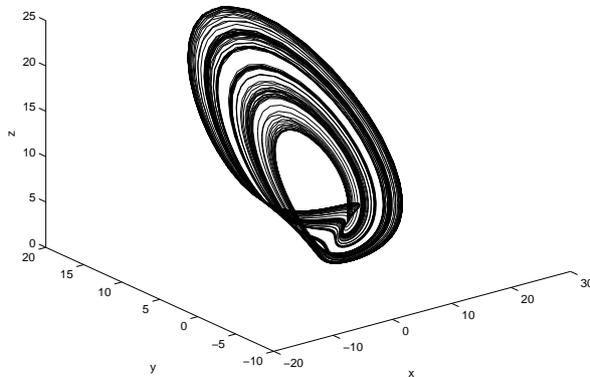


Figure 1: The 1-scroll chaotic attractor with $a=-10, b=-4$ and $c=18.1$.

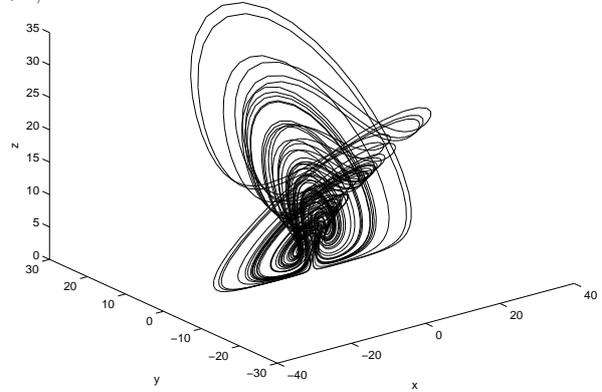


Figure 2: The 2-scroll chaotic attractor with $a=-10, b=-4$ and $c=0$.

The Lü-chen system is found to be chaotic in a wide range of parameters and has many interesting complex dynamical behaviors. For example, it is chaotic for the parameters $a = -10, b = -4$ and $|c| < 19.2$. In particular, it displays 1-scroll chaotic attractor when $a = -10, b = -4, c = 18.1$ as shown in Fig. 1. and 2-scroll chaotic attractors when $a = -10, b = -4, c = 0$ as shown in Fig. 2.

The fractional order version of Lü-chen system is given as follows:

$$\begin{cases} \frac{d^{q_1} x}{dt^{q_1}} = -\frac{ab}{a+b}x - yz + c, \\ \frac{d^{q_2} y}{dt^{q_2}} = ay + xz, \\ \frac{d^{q_3} z}{dt^{q_3}} = bz + xy, \end{cases} \quad (5)$$

its order q_i is subject to $0 < q_i \leq 1, i = 1, 2, 3$. Figure 3 and 4 display the chaotic attractors of system (5) with different $q_i, i = 1, 2, 3$.

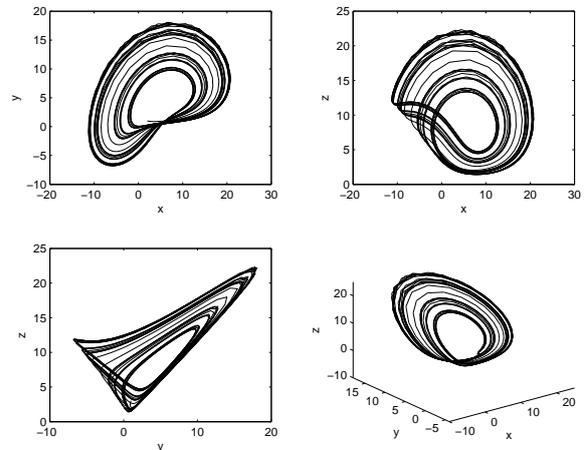


Figure 3: The 1-scroll chaotic attractor of fractional order Lü-chen system with $a=-10, b=-4, c=18.1$ and $q_1 = 0.985, q_2 = 0.99, q_3 = 0.995$.

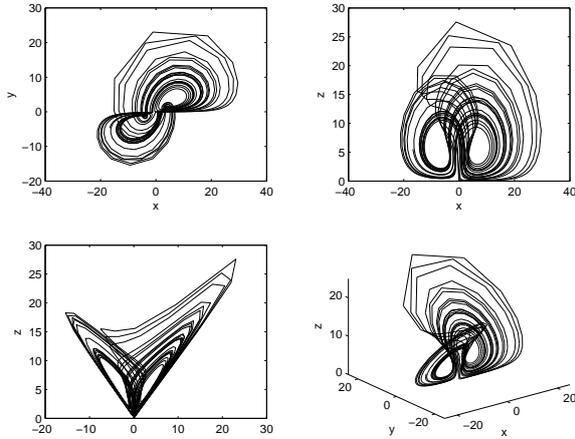


Figure 4: The 2-scroll attractor of fractional order Lü-chen system with $a = -10, b = -4, c = 0$ and $q_1 = 0.85, q_2 = 0.9, q_3 = 0.95$.

The fractional order Lorenz chaotic system is

$$\begin{cases} \frac{d^{q_1} x_1}{dt^{q_1}} = \delta(x_2 - x_1), \\ \frac{d^{q_2} x_2}{dt^{q_2}} = rx_1 - x_1x_3 - x_2, \\ \frac{d^{q_3} x_3}{dt^{q_3}} = x_1x_2 - dx_3, \end{cases} \quad (6)$$

where δ, r, b are three positive real constants. System (6) is chaotic when $q_1 = 0.985, q_2 = 0.99, q_3 = 0.995$ and $\delta = 10, r = 28, d = \frac{8}{3}$. The chaotic attractor of fractional order Lorenz chaotic system is depicted in Figure 5.

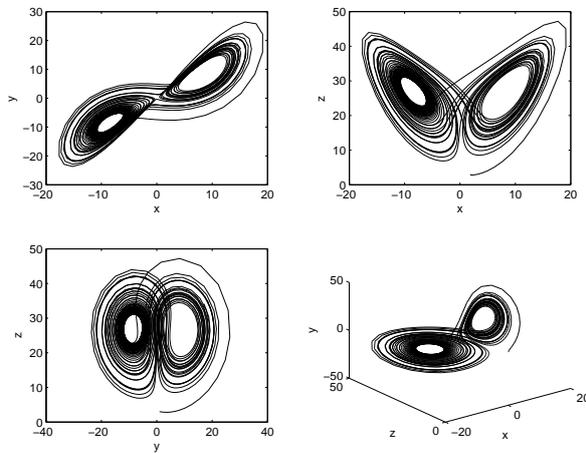


Figure 5: The chaotic attractor of fractional order Lorenz chaotic system with $q_1 = 0.985, q_2 = 0.99, q_3 = 0.995$ and $\delta = 10, r = 28, d = \frac{8}{3}$.

Suppose system (6) is the drive system, which can be rewritten as: $\frac{d^q x}{dt^q} = Ax + h(x, t)$, where $A = \begin{pmatrix} -\delta & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -d \end{pmatrix}$, $h(x, t) = [\delta x_2, rx_1 - x_1x_3, x_1x_2]^T$.

Let system (5) be the response system, then the controlled response system can be read as:

$$\begin{cases} \frac{d^{q_1} y_1}{dt^{q_1}} = -\frac{ab}{a+b}y_1 - y_2y_3 + c + u_1, \\ \frac{d^{q_2} y_2}{dt^{q_2}} = ay_2 + y_1y_3 + u_2, \\ \frac{d^{q_3} y_3}{dt^{q_3}} = by_3 + y_1y_2 + u_3. \end{cases} \quad (7)$$

System (7) can be described as: $\frac{d^q y}{dt^q} = g(y, t) + u = Ay + S(y, t) + u$, where $u = [u_1, u_2, u_3]^T$ is a controller. Assume $\alpha = \text{diag}(2, 1, -1)$, then $u = \begin{pmatrix} -\delta y_1 + \frac{ab}{a+b}y_1 + y_2y_3 - c + \frac{\delta}{2}x_2 \\ -y_2 - (ay_2 + y_1y_3) + rx_1 - x_1x_3 \\ -dy_3 - (by_3 + y_1y_2) - x_1x_3 \end{pmatrix}$. By Theorem 1, systems (6) and (7) will achieve modified projective synchronization. In our simulation, the initial states of the drive system (6) are $x_1 = 1, x_2 = 1, x_3 = 3$ and initial states of the response system (7) are $y_1 = 1, y_2 = 2, y_3 = 2$. The numerical simulation results with $q_1 = 0.985, q_2 = 0.99, q_3 = 0.995$ and $\delta = 10, r = 28, d = \frac{8}{3}$ are shown in Figures 6 and 7, respectively. Figure 6 displays the evolution of state variables of systems (6) and (7), Figure 7 shows the error signals between systems (6) and (7). It is easy to see that systems (6) and (7) can achieve modified projective synchronization quickly.

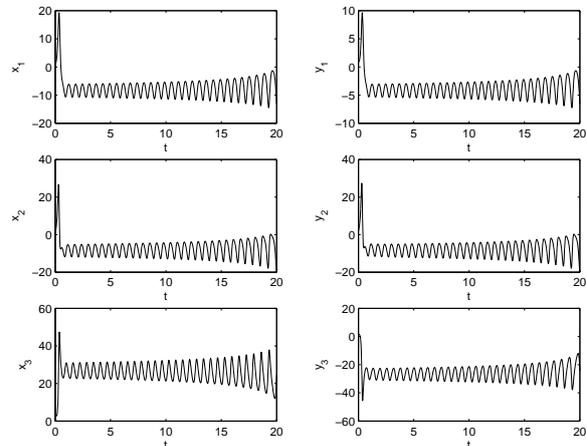


Figure 6: The evolution of state variables of systems (6) and (7) with $q_1 = 0.985, q_2 = 0.99, q_3 = 0.995$ and $\delta = 10, r = 28, d = \frac{8}{3}$.

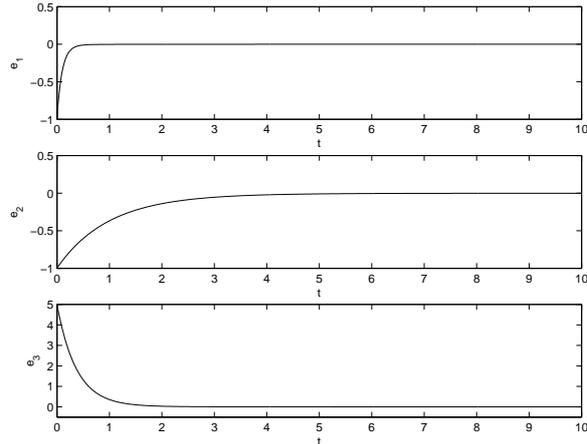


Figure 7: The error signals between systems (6) and (7) with $q_1 = 0.985, q_2 = 0.99, q_3 = 0.995$ and $\delta = 10, r = 28, d = \frac{8}{3}$.

IV. CONCLUSION

In this paper, we have investigated the modified projective synchronization of fractional order chaotic systems with a

general kind of proportional relationships between the drive and response systems. Based on the stability criterion of linear systems, a new approach for constructing modified projective synchronization of fractional order systems is attained. From theory and numerical simulation we have shown that the different fractional order chaotic systems can be globally asymptotically synchronized by utilizing the proposed scheme.

Shucheng Deng received his B.Sc. degree from Department of Mathematics, Nanchang University, in 2005, where he is currently working toward the the M.Sc. degree in operational research and cybernetics. His research interests are stability theory, dynamical systems, impulsive control, chaos synchronization and communication security.

ACKNOWLEDGMENT

This paper is supported by the Natural Science Foundation of Jiangxi Province, P.R. China under Grant No. 2007GZS2126.

REFERENCES

- [1] I. Podlubny, *Fractional Differential Equations*, New York: Academic Press, 1999.
- [2] S. G. Samko, A. A. Kilbas, O. I. Marichev, *Fractional Integrals and Derivatives : Theory and Applications*, Yverdon: Gordon and Breach, 1993.
- [3] R. Hilfer, *Applications of Fractional Calculus in Physics*, New Jersey: World Scientific, 2001.
- [4] X. Gao, J. Yu, "Chaos in the fractional order periodically forced complex Duffing's oscillators," *Chaos, Solitons & Fractals*, vol. 26, pp. 1125-1133, 2005.
- [5] T. T. Hartley, C. F. Lorenzo, H. K. Qammer, "Chaos on a fractional Chua's system," *IEEE Trans Circ Syst Theory Appl*, vol. 42, pp. 485-490, 1995.
- [6] C. P. Li, G. J. Peng, "Chaos in Chen's system with a fractional order," *Chaos, Solitons & Fractals*, vol. 20, pp. 443-450, 2004.
- [7] W. H. Deng, C. P. Li, "Chaos synchronization of the fractional Lü system," *Physica A*, vol. 353, pp. 61-72, 2005.
- [8] J. W. Wang, Y. B. Zhang, "Designing synchronization schemes for chaotic fractional-order unified systems," *Chaos, Solitons and Fractals*, vol. 30, pp. 1265-1272, 2006.
- [9] P. Zhou, "Chaotic synchronization for a class of fractional-order chaotic systems," *Chinese Physics*, vol. 16, pp. 1263-1266, 2007.
- [10] X. Gao, J. B. Yu, "Synchronization of two coupled fractional- chaotic oscillators," *Chaos, Solitons and Fractals*, vol. 26, p. 141-145, 2005.
- [11] J. W. Wang, X. H. Xiong, Y. B. Zhang, "Extending synchronization scheme to chaotic fractional-order Chen systems," *Physica A*, vol. 370, pp. 279-285, 2006.
- [12] C. P. Li, W. H. Deng, D. Xu, "Chaos synchronization of the Chua system with a fractional order," *Physica A*, vol. 360, pp. 171-185, 2006.
- [13] J. G. Lu, "Synchronization of a class of fractional-order chaotic systems via a scalar transmitted signal," *Chaos, Solitons and Fractals*, vol. 27, pp. 519-525, 2006.
- [14] X. Y. Wang, Y. J. He, "Projective synchronization of fractional order chaotic system based on linear separation," *Physics Letters A*, vol. 372, pp. 435-441, 2008.
- [15] G. J. Peng, Y. L. Jiang, "Generalized projective synchronization of a class of fractional-order chaotic systems via a scalar transmitted signal," *Physics Letters A*, vol. 372, pp. 3963-3970, 2008.
- [16] G. J. Peng, Y. L. Jiang, F. Chen, "Generalized projective synchronization of fractional order chaotic systems," *Physica A*, vol. 387, pp. 3738-3746, 2008.
- [17] J. H. Lü, G. Chen, "A new chaotic system and beyond," *Int J Bifurcat Chaos*, vol. 14, pp. 1507-1537, 2004.

Zhengmin Wei received his B.Sc. degree from Department of Mathematics, Fuyang Teachers College, in 2007. He is currently working toward the the M.Sc. degree in operational research and cybernetics in Nanchang University. His research interests are systems analysis, stability theory, impulsive control, chaos synchronization, nonlinear oscillations.

Runzi Luo received the M.Sc. and Ph.D. degrees from Shanghai University, Shanghai, China, in 2001 and 2005, respectively, all in mathematics. He joined the Department of Mathematics, Nanchang University, P. R. China, in 2005, where he became an Associate Professor in 2008. His research areas include combinatorial optimization, systems analysis, stability theory, dynamical systems, impulsive control, chaos synchronization, nonlinear oscillations, and communication security.

On some stability conditions and hyperchaos synchronization in the new fractional order hyperchaotic Chen system

Ahmed E. Matouk ^{a,b}

Abstract— This work investigates some stability conditions and hyperchaos in the new fractional order hyperchaotic Chen system which shows hyperchaos with order less than 4. The lowest order found to yield chaos for the new fractional order hyperchaotic Chen system is 3.76. The analytical conditions for achieving synchronization in this system via linear control are investigated theoretically by using the Laplace transform theory. Numerical simulations are used to verify the existence of hyperchaos and to show the effectiveness of the proposed synchronization technique.

Index Terms—fractional order, Routh-Hurwitz, hyperchaotic Chen system, synchronization

I. INTRODUCTION

The idea of fractional calculus has been known since the work of Leibniz and L'Hopital in 1695 [1]. It has useful applications in physics, engineering [2], mathematical biology [3-4] and finance [5].

The fractional order derivatives have many definitions. The Caputo definition of fractional derivative [6] is used throughout this paper and is given as follows:

$$D^\alpha f(x) = I^{m-\alpha} f^{(m)}(x), \quad \alpha > 0, \quad (1)$$

where $f^{(m)}$ represents the m -order derivative of $f(x)$, $m = [\alpha]$ is the first integer which is not less than α , and

$$I^q g(x) = \frac{1}{\Gamma(q)} \int_0^x (x-t)^{q-1} g(t) dt, \quad q > 0, \quad (2)$$

is the q -order Riemann-Liouville integral operator, where $\Gamma(q)$ is the gamma function. The operator D^α is called the " α -order Caputo differential operator". The geometric and physical interpretation of the fractional derivatives was given in

^a Mathematics Department, Faculty of Science, Mansoura University, Mansoura, 35516, Egypt.

^b Mathematics Department, Faculty of science, Hail University, Hail, 2440, Saudi Arabia.
E.mail: amatouk@hotmail.com

Ref. [7]. On the other hand, studying chaos and hyperchaos in fractional order systems has recently a particular interest by scientists. A regular chaotic system has one positive Lyapunov exponent. However, a hyperchaotic system has more than one positive Lyapunov exponent which shows more complex behaviors and abundant dynamics than chaotic system. Therefore, hyperchaotic systems have better applications in secure communications [8] than chaotic ones. Recently, some fractional order hyperchaotic systems have been investigated, such as the fractional order hyperchaotic Rössler system [9], the fractional order hyperchaotic Lü system [10], the fractional order hyperchaotic Lorenz system [11] and the new fractional order hyperchaotic system [12].

In this study, some Routh-Hurwitz conditions are derived to discuss local stability in some fractional order hyperchaotic systems. The proposed conditions are applied successfully to the new fractional order hyperchaotic Chen system. Numerical results show that hyperchaos does exist in the proposed system with order less than 4 and Lyapunov exponents are also calculated for this system. Moreover, the Laplace transform theory is used to achieve synchronization between two identical new fractional order hyperchaotic Chen systems via linear control technique.

II. SOME ROUTH-HURWITZ CONDITIONS FOR THE FRACTIONAL ORDER HYPERCHAOTIC SYSTEMS

Consider the 4-dimensional fractional order hyperchaotic system

$$\begin{aligned} D^\alpha x_1(t) &= f_1(x_1, x_2, x_3, x_4), & D^\alpha x_2(t) &= f_2(x_1, x_2, x_3, x_4), \\ D^\alpha x_3(t) &= f_3(x_1, x_2, x_3, x_4), & D^\alpha x_4(t) &= f_4(x_1, x_2, x_3, x_4), \end{aligned} \quad (3)$$

where the fractional derivative in Eq. (3) is in the sense of Caputo and $0 \leq \alpha < 1$. If $\bar{E} = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4)$ is an equilibrium solution of (3) then \bar{E} is locally asymptotically stable if all the eigenvalues λ of the Jacobian matrix

$$J = \begin{bmatrix} \partial f_1 / \partial x_1 & \partial f_1 / \partial x_2 & \partial f_1 / \partial x_3 & \partial f_1 / \partial x_4 \\ \partial f_2 / \partial x_1 & \partial f_2 / \partial x_2 & \partial f_2 / \partial x_3 & \partial f_2 / \partial x_4 \\ \partial f_3 / \partial x_1 & \partial f_3 / \partial x_2 & \partial f_3 / \partial x_3 & \partial f_3 / \partial x_4 \\ \partial f_4 / \partial x_1 & \partial f_4 / \partial x_2 & \partial f_4 / \partial x_3 & \partial f_4 / \partial x_4 \end{bmatrix},$$

evaluated at \bar{E} satisfies the Matignon's condition [13]:

$$|\arg(\lambda_i)| > \alpha\pi/2, \quad (i=1,2,3,4). \quad (4)$$

According to the Matignon's conditions (4), the stability region of the fractional order system with order α is illustrated in figure 1. In this figure one sets σ, ω to denote the real and imaginary parts of the eigenvalues respectively and $j = \sqrt{-1}$. The eigenvalues equation of the equilibrium point \bar{E} is given as

$$P(\lambda) = \lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4 = 0, \quad (5)$$

whose discriminant $D(P)$ is given by:

$$\begin{aligned} D(P) = & -4a_3^3a_1^3 + a_2^2a_1^2a_3^2 + 18a_3^3a_2a_1 - 6a_1^2a_2^2a_4 \\ & - 4a_2^3a_1^2a_4 - 80a_3a_2^2a_1a_4 + 144a_3^2a_2a_4 \\ & - 192a_3a_2^2a_1 + 144a_2a_1^2a_4^2 + 18a_3a_2a_1^3a_4 \\ & - 27a_3^4 - 4a_2^3a_3^2 - 128a_4^2a_2^2 - 27a_1^4a_4^2 \\ & + 256a_4^3 + 16a_1^4a_4. \end{aligned} \quad (6)$$

Theorem 1.

- (i) If c_1, c_2, c_3 are Routh-Hurwitz determinants which are defined as follows:

$$c_1 = a_1, \quad c_2 = a_1a_2 - a_3, \quad c_3 = a_1a_2a_3 - a_1^2a_4 - a_3^2, \quad (7)$$

then for $\alpha=1$, the equilibrium point \bar{E} to is locally asymptotically stable if and only if

$$c_1 > 0, \quad c_2 > 0, \quad c_3 > 0, \quad a_4 > 0. \quad (8)$$

Also, the conditions (8) are sufficient conditions for the equilibrium point \bar{E} to be locally asymptotically stable for all $\alpha \in [0,1)$.

- (ii) If $D(P) > 0$, $a_1 > 0$, $a_2 < 0$ and $\alpha > 2/3$ then the equilibrium point \bar{E} is unstable.
 (iii) If $D(P) < 0$, $a_1 > 0$, $a_2 > 0$, $a_3 > 0$, $a_4 > 0$, and $\alpha < 1/3$, then the equilibrium point \bar{E} is locally asymptotically stable. Also, if $D(P) < 0$, $a_1 < 0$, $a_2 > 0$, $a_3 < 0$, $a_4 > 0$, then the equilibrium point \bar{E} is unstable.
 (iv) If $D(P) < 0$, $a_1 > 0$, $a_2 > 0$, $a_3 > 0$, $a_4 > 0$ and $a_2 = \frac{a_1a_4 + a_3}{a_3 + a_1}$, then the equilibrium point \bar{E} is locally asymptotically stable, for all $\alpha \in (0,1)$.

- (v) $a_4 > 0$, is the necessary condition for the equilibrium point \bar{E} to be locally asymptotically stable.

Proof. To prove the case (i), assume that the conditions (8) are satisfied, then all real eigenvalues and all real parts of complex conjugate eigenvalues of equation (5) are negative (this is shown clearly from the case when $\alpha = 1$ which is proved in the classical Routh-Hurwitz theory [14]), hence, these conditions (8) implies that all the eigenvalues of (5) lie in the left-half plane of figure 1 (the stable region). Therefore \bar{E} is locally asymptotically stable.

To prove (ii), notice that if $D(P) > 0$ then there exists 4 distinct real roots or two pairs of complex eigenvalues $\lambda_{1,2} = s \pm jt$, and $\lambda_{3,4} = p \pm jq$ where $j = \sqrt{-1}$. In the case of real roots, $a_2 < 0$ implies that the equilibrium point \bar{E} is unstable. In the other case:

$a_2 = p^2 + q^2 + s^2 + t^2 + 4sp < 0$ then $p^2 \sec^2 \theta + s^2 + t^2 < -4sp$, where $\theta = |\arg \lambda_{3,4}|$. Using the condition $a_1 > 0$ then $p^2 \sec^2 \theta < -4sp < 4p^2$. This implies that $\theta < \pi/3$. Hence, (ii) is proved.

To prove (iii), notice that if $D(P) < 0$ then there exist two real roots $\lambda_1 = r_1$, $\lambda_2 = r_2$, and one pair of complex eigenvalues $\lambda_{3,4} = p_1 \pm jq_1$. Then the conditions $a_1 > 0, a_2 > 0, a_3 > 0, a_4 > 0$ implies that r_1, r_2 are negative and the condition $a_2 > 0$ implies that $p^2 \sec^2 \theta > -2p(r_1 + r_2) - r_1r_2$, where $\theta = |\arg \lambda_{3,4}|$. Using the condition $a_1 > 0$ implies that $-(r_1 + r_2) > 2p$, therefore $p^2 \sec^2 \theta > -(r_1r_2) + 4p^2$. Since $a_3 > 0$ then $\frac{(r_1 + r_2)}{2} p \sec^2 \theta < -r_1r_2$, thus, $p^2 \sec^2 \theta > \frac{r_1 + r_2}{2} p \sec^2 \theta + 4p^2$, again using the condition $a_1 > 0$, the last inequality is reduced to $3p^2 \sec^2 \theta > 4p^2$ which implies that $\sec^2 \theta > 4/3 \Rightarrow \theta > \pi/6$. Then the first part of (iii) is proved. On the other hand, if the conditions $a_1 < 0, a_2 > 0, a_3 < 0, a_4 > 0$ are satisfied then using Descartes' rule of signs, it follows that there is no negative real roots for the characteristic polynomial, this implies that $r_1, r_2 > 0$ and therefore the equilibrium point \bar{E} is unstable.

To prove (iv), notice that the two real roots are negative and the condition $a_2 = \frac{a_1a_4 + a_3}{a_3 + a_1}$ implies that the other two eigenvalues lie on the imaginary axis. Consequently, all the roots of (5) lie in the stable region, and (iv) is now proved.

The part (v) is proved in [15] for general n , which includes

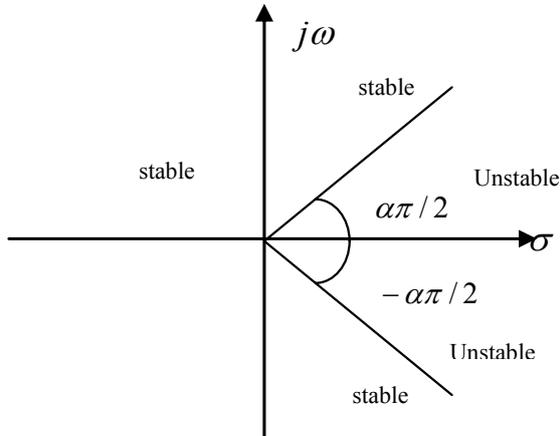


Fig. 1. Stability region of the fractional order system.

our current case. □

III. THE NEW FRACTIONAL ORDER HYPERCHAOTIC CHEN SYSTEM

In the following, we investigate the stability conditions and hyperchaos in the new fractional order hyperchaotic Chen system. This system will be integrated numerically to show hyperchaos using an efficient method for solving fractional order differential equations that is the predictor-correctors scheme or more precisely, **PECE** (Predict, Evaluate, Correct, Evaluate) technique which has been investigated in [16-17], and represents a generalization of the Adams-Bashforth-Moulton algorithm. It is used throughout this paper.

Now, we consider the new fractional order hyperchaotic Chen system as follows:

$$\begin{aligned} D^\alpha x &= a(y-x) + u, & D^\alpha y &= \gamma x - xz + cy, \\ D^\alpha z &= xy - bz, & D^\alpha u &= yz + du. \end{aligned} \quad (9)$$

The integer order form of system (9) is studied in [18]. System (9) has only one equilibrium point $E_0 = (0, 0, 0)$. At the parameter values $(a, b, c, d, \gamma) = (35, 3, 12, 0.3, 7)$ and $\alpha = 0.97$, system (9) has two positive Lyapunov exponents $\lambda_1 \approx 1.214$ and $\lambda_2 \approx 0.138$ which are calculated using the algorithm given in [19]. Figure 2 shows the hyperchaotic attractor of system (9) for the previously mentioned parameter values and fractional order α . However the lowest fractional order at which system (9) exhibits chaotic attractor is $\alpha = 0.94$, i.e., the lowest order found to yield chaos for system (9) is 3.76.

The characteristic polynomial for the equilibrium point E_0 is given from

$$\begin{aligned} \lambda^4 + (-c+a+b-d)\lambda^3 + (cd-\gamma a-ad-bd+ab-cb-ac)\lambda^2 \\ + (-acb+acd+cbd+\gamma ad-abd-\gamma ab)\lambda + abd(c+\gamma) = 0. \end{aligned} \quad (10)$$

Equation (10) has the roots $\lambda_1 = d, \lambda_2 = -b$ and $\lambda_{3,4} = \frac{c-a \pm \sqrt{(a+c)^2 + 4a\gamma}}{2}$. If all these eigenvalues satisfy the conditions $|\arg(\lambda_i)| > \alpha\pi/2$, ($i=1,2,3,4$), then system (9) is locally asymptotically stable at the equilibrium point E_0 . Moreover, using the above-mentioned parameter values, it is easy to verify that $D(P) > 0$, $a_1 > 0$, $a_2 < 0$, $a_4 > 0$. Thus, Theorem 1 part (ii) implies that the equilibrium point E_0 is unstable for $\alpha > 2/3$.

IV. SYNCHRONIZATION OF THE NEW FRACTIONAL ORDER HYPERCHAOTIC CHEN SYSTEM

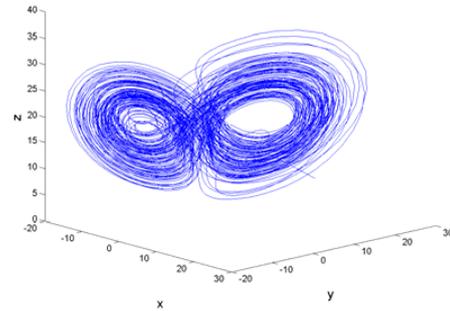


Fig. 2. 3-D plot of the new fractional order hyperchaotic Chen attractor in x-y-z space.

In the following, our aim is to achieve chaos synchronization of the new fractional order hyperchaotic Chen system. The drive and response systems are given as follows:

$$\begin{aligned} D^\alpha x_m &= a(y_m - x_m) + u_m, & D^\alpha y_m &= \gamma x_m - x_m z_m + cy_m, \\ D^\alpha z_m &= x_m y_m - bz_m, & D^\alpha u_m &= y_m z_m + du_m, \end{aligned} \quad (11)$$

and

$$\begin{aligned} D^\alpha x_s &= a(y_s - x_s) + u_s + v_1, & D^\alpha y_s &= \gamma x_s - x_s z_s + cy_s + v_2, \\ D^\alpha z_s &= x_s y_s - bz_s + v_3, & D^\alpha u_s &= y_s z_s + du_s + v_4, \end{aligned} \quad (12)$$

where v_1, v_2, v_3 and v_4 are the controllers. Define the error variables as follows:

$$e_1 = x_s - x_m, \quad e_2 = y_s - y_m, \quad e_3 = z_s - z_m, \quad e_4 = u_s - u_m. \quad (13)$$

By subtracting (11) from (12) and using (13), we obtain:

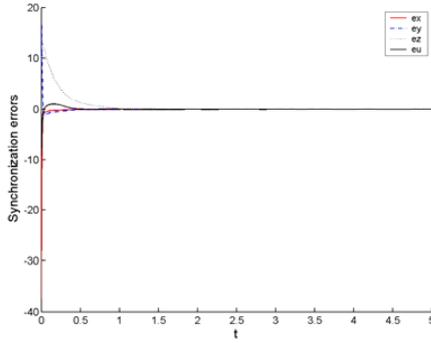


Fig. 3. Synchronization errors between the drive and response systems tend to zero when using fractional order $\alpha=0.97$ and feedback control gains $k_1=100, k_2=100, k_3=1, k_4=30$.

$$\begin{aligned} D^\alpha e_1 &= a(e_2 - e_1) + e_4 + v_1, \\ D^\alpha e_2 &= \gamma e_1 - z_m e_1 - x_m e_3 - e_1 e_3 + c e_2 + v_2, \\ D^\alpha e_3 &= -b e_3 + y_m e_1 + x_m e_2 + e_1 e_2 + v_3, \\ D^\alpha e_4 &= d e_4 + z_m e_2 + y_m e_3 + e_2 e_3 + v_4. \end{aligned} \quad (14)$$

Now, by letting

$$v_1 = -k_1 e_1 - e_4, \quad v_2 = -k_2 e_2, \quad v_3 = -k_3 e_3, \quad v_4 = -k_4 e_4, \quad (15)$$

where $k_1, k_2, k_3, k_4 \geq 0$, the error system (14) is reduced to

$$\begin{aligned} D^\alpha e_1 &= -(a+k_1)e_1 + a e_2, \\ D^\alpha e_2 &= \gamma e_1 + (c-k_2)e_2 - z_m e_1 - x_m e_3 - e_1 e_3, \\ D^\alpha e_3 &= -(b+k_3)e_3 + y_m e_1 + x_m e_2 + e_1 e_2, \\ D^\alpha e_4 &= (d-k_4)e_4 + z_m e_2 + y_m e_3 + e_2 e_3. \end{aligned} \quad (16)$$

By taking the Laplace transform in both sides of (16), letting $E_i(s) = L\{e_i(t)\}$ where $(i=1,2,3,4)$, and applying $L\{d^\alpha e_i / dt^\alpha\} = s^\alpha E_i(s) - s^{\alpha-1} e_i(0)$, we obtain:

$$\begin{aligned} s^\alpha E_1(s) - s^{\alpha-1} e_1(0) &= -(a+k_1)E_1(s) + aE_2(s), \\ s^\alpha E_2(s) - s^{\alpha-1} e_2(0) &= \gamma E_1(s) + (c-k_2)E_2(s) - L\{x_m e_3\} \\ &\quad - L\{z_m e_1\} - E_1(s)E_3(s), \\ s^\alpha E_3(s) - s^{\alpha-1} e_3(0) &= -(b+k_3)E_3(s) + L\{y_m e_1\} \\ &\quad + L\{x_m e_2\} + E_1(s)E_2(s), \\ s^\alpha E_4(s) - s^{\alpha-1} e_4(0) &= (d-k_4)E_4(s) + L\{y_m e_3\} \\ &\quad + L\{z_m e_2\} + E_2(s)E_3(s). \end{aligned} \quad (17)$$

Theorem 2. If $E_1(s), E_2(s)$ are bounded, $k_2 - c \neq 0$ and $k_4 - d \neq 0$, then the drive and response systems (11) and (12) will be synchronized under a suitable choice of k_1, k_2, k_3 and k_4 .

Proof. By rewriting equation (17) as follows:

$$\begin{aligned} E_1(s) &= \frac{aE_2(s)}{s^\alpha + a + k_1} + \frac{s^{\alpha-1} e_1(0)}{s^\alpha + a + k_1}, \\ E_2(s) &= \frac{\gamma E_1(s)}{s^\alpha - c + k_2} - \frac{L\{z_m e_1\}}{s^\alpha - c + k_2} - \frac{L\{x_m e_3\}}{s^\alpha - c + k_2} - \frac{E_1(s)E_3(s)}{s^\alpha - c + k_2} \\ &\quad + \frac{s^{\alpha-1} e_2(0)}{s^\alpha - c + k_2}, \\ E_3(s) &= \frac{L\{y_m e_1\}}{s^\alpha + b + k_3} + \frac{L\{x_m e_2\}}{s^\alpha + b + k_3} + \frac{E_1(s)E_2(s)}{s^\alpha + b + k_3} + \frac{s^{\alpha-1} e_3(0)}{s^\alpha + b + k_3}, \\ E_4(s) &= \frac{L\{y_m e_3\}}{s^\alpha - d + k_4} + \frac{L\{z_m e_2\}}{s^\alpha - d + k_4} + \frac{E_2(s)E_3(s)}{s^\alpha - d + k_4} + \frac{s^{\alpha-1} e_4(0)}{s^\alpha - d + k_4}. \end{aligned} \quad (18)$$

According to the Final-value theorem of the Laplace transform, it follows that

$$\begin{aligned} \lim_{t \rightarrow \infty} e_1(t) &= \lim_{s \rightarrow 0^+} s E_1(s) = \frac{a}{a+k_1} \lim_{s \rightarrow 0^+} s E_2(s) = \frac{a}{a+k_1} \lim_{t \rightarrow \infty} e_2(t), \\ \lim_{t \rightarrow \infty} e_2(t) &= \lim_{s \rightarrow 0^+} s E_2(s) = \frac{\gamma \lim_{t \rightarrow \infty} e_1(t)}{k_2 - c} - \frac{\lim_{t \rightarrow \infty} sL\{x_m e_3\}}{k_2 - c} \\ &\quad - \frac{\lim_{s \rightarrow 0^+} sL\{z_m e_1\}}{k_2 - c} - \frac{\lim_{t \rightarrow \infty} e_1(t) \cdot \lim_{t \rightarrow \infty} e_3(t)}{k_2 - c}, \\ \lim_{t \rightarrow \infty} e_3(t) &= \lim_{s \rightarrow 0^+} s E_3(s) = \frac{\lim_{s \rightarrow 0^+} sL\{y_m e_1\}}{b+k_3} + \frac{\lim_{s \rightarrow 0^+} sL\{x_m e_2\}}{b+k_3} \\ &\quad + \frac{\lim_{t \rightarrow \infty} e_1(t) \cdot \lim_{t \rightarrow \infty} e_2(t)}{b+k_3}, \\ \lim_{t \rightarrow \infty} e_4(t) &= \lim_{s \rightarrow 0^+} s E_4(s) = \frac{\lim_{s \rightarrow 0^+} sL\{y_m e_3\}}{k_4 - d} + \frac{\lim_{s \rightarrow 0^+} sL\{z_m e_2\}}{k_4 - d} \\ &\quad + \frac{\lim_{t \rightarrow \infty} e_2(t) \cdot \lim_{t \rightarrow \infty} e_3(t)}{k_4 - d}. \end{aligned} \quad (19)$$

Since $E_1(s), E_2(s)$ are bounded, $k_2 - c \neq 0$ then $\lim_{t \rightarrow \infty} e_1(t) = \lim_{t \rightarrow \infty} e_2(t) = 0$. Now, owing to the attractiveness of the attractor, there exists $\varepsilon > 0$ such that $|x_i(t)| \leq \varepsilon < \infty, |y_i(t)| \leq \varepsilon < \infty, |z_i(t)| \leq \varepsilon < \infty$ and $|u_i(t)| \leq \varepsilon < \infty$ where i refers to the subscript of the drive or response variables. Consequently, $\lim_{t \rightarrow \infty} e_3(t) = 0$ and $\lim_{t \rightarrow \infty} e_4(t) = 0$ provided that $k_4 - d \neq 0$. Hence, we have proved that

$$\lim_{t \rightarrow \infty} e_i(t) = 0, \quad i = 1, 2, 3, 4. \quad (20)$$

Thus, the synchronization between the drive and response systems (11) and (12) is achieved. \square

A. Numerical results

Based on the PECE scheme, the drive and response systems (11) and (12) are integrated numerically using the above-mentioned parameter values and fractional order

$\alpha = 0.97$, with the initial values $x_m(0) = 30$, $y_m(0) = 7$, $z_m(0) = 10$, $u_m(0) = 40$ and $x_s(0) = -15$, $y_s(0) = 25$, $z_s(0) = 25$, $u_s(0) = 30$. From figure 3, it is clear that synchronization is achieved when $k_1 = 100, k_2 = 100, k_3 = 1$ and $k_4 = 30$.

V. CONCLUSION

Some stability conditions in fractional order hyperchaotic systems have been derived and applied to the new fractional order hyperchaotic Chen system. Numerical simulations and Lyapunov exponents have been used to show that hyperchaos exists in this system with order less than 4. It has been shown that according to the Laplace transformation theory, one achieve synchronization of the new fractional order hyperchaotic Chen system when choosing suitable linear controllers.

ACKNOWLEDGMENTS

I would like to thank Prof. E. Ahmed, Prof. A.S. Hegazi and Dr. Matjaž Perc for their support. Thanks to Hail University for the financial support of this project.

REFERENCES

- [1] P.L. Butzer and U. Westphal, "An introduction to fractional calculus," World Scientific, Singapore, 2000.
- [2] H.H. Sun, A.A. Abdelwahab and B. Onaral, "Linear approximation of transfer function with a pole of fractional order," IEEE Trans. Auto. Contr. 29 (1984) 441-444.
- [3] E. Ahmed and A.S. Elgazzar, "On fractional order differential equations model for nonlocal epidemics," Physica A 379 (2007) 607-614.
- [4] A.M.A. El-Sayed, A.E.M. El-Mesiry and H.A.A. El-Saka, "On the fractional-order logistic equation," Applied Mathematics Letters 20 (2007) 817-823.
- [5] N. Laskin, "Fractional market dynamics," Physica A 287 (2000) 482-492.
- [6] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," Geophys. J. R. Astron. Soc. 13 (1967) 529-539.
- [7] F. Ben Adda, "Geometric interpretation of the fractional derivative," J. Fract. Calc. 11 (1997) 21-52.
- [8] C. Li, X. Liao and K.W. Wong, "Lag synchronization of hyperchaos with application to secure communications," Chaos Solitons & Fractals 23 (2005) 183-193.
- [9] C. Li and G. Chen, "Chaos and hyperchaos in the fractional-order Rössler equations," Physica A 341 (2004) 55-61.
- [10] J. Liu and X. Li, "Synchronization of fractional hyperchaotic Lü system via unidirectional coupling method," in: Proceedings of the 7th World Congress on Intelligent Control and Automation, June 25-27, 2008, Chongqing, China, pp. 4653-4658.
- [11] X.Y. Wang and J.M. Song, "Synchronization of the fractional order hyperchaos Lorenz systems with activation feedback control," Commun Nonlinear Sci Numer Simulat 14 (2009) 3351-3357.
- [12] H. Deng, T. Li, Q. Wang and H. Li, "A fractional-order hyperchaotic system and its synchronization," Chaos Solitons & Fractals (2008), doi:10.1016/j.chaos.2008.04.034.
- [13] D. Matignon, "Stability results for fractional differential equations with applications to control processing," in: Computational Engineering in System Application, vol. 2, Lille, France, 1996, pp. 963-968.
- [14] A.P. Mishina and I.V. Proskuryakov, "Higher Algebra," Nauka, Moscow, 1965.
- [15] E. Ahmed, A.M.A. El-Sayed and H.A.A. El-Saka, "On some Routh-Hurwitz conditions for fractional order differential equations and their applications in Lorenz, Rössler, Chua and Chen systems," Physics Lett. A 358 (2006) 1-4.
- [16] K. Diethelm and N.J. Ford, "Analysis of fractional differential equations," J. Math. Analysis and Applications 265 (2002) 229-248.
- [17] K. Diethelm, N.J. Ford and A.D. Freed, "A predictor-corrector approach for the numerical solution of fractional differential equations," Nonlinear Dynam. 29 (2002) 3-22.
- [18] Z. Yan, "Controlling hyperchaos in the new hyperchaotic Chen system," Applied Mathematics and Comp. 168 (2005) 1239-1250.
- [19] D.I. Abarbanel, "Analysis of observed chaotic data," Springer-Verlag, New York, 1996.

Backstepping Control of Fractional-Order Chaotic Systems

E. Naseri*, A. Ranjbar*, S.H. Hosseinnia*, S. Momani**

*Intelligent system research group, Faculty of Electrical and Computer Engineering,
Noushirvani University of Technology, Babol, Iran, P.O. Box 47135-484,
(a.ranjbar@nit.ac.ir), (h.hosseinnia@stu.nit.ac.ir)

** Department of Mathematics, Mutah University, P.O. Box: 7, Al-Karak, Jordan
(shaherm@yaho.com)

Abstract- In this paper, chaos and its control is studied in fractional-order chaotic systems. Backstepping method is proposed to synchronize two identical fractional-order Chen systems. The simulation results show that this method can effectively synchronize two identical chaotic systems.

I. INTRODUCTION

Fractional calculus is an old mathematical topic, which has been originated from 17th century. Nowadays, some fractional-order differential systems such as Chua circuit [1], Duffing system [2], jerk model [3], Chen system [4], the fractional-order Lü system [5], Rossler system [6], Arneodo system [7] and Newton–Leipnik system [8] have been found to demonstrate chaotic behaviors. Sensitive dependence on initial conditions is an important characteristic of chaotic systems. For this reason, chaotic systems are difficult to be controlled or synchronized. Control of these systems has been considered as an important and challenging problem [9]. Control of chaotic systems would have been supposed impossible with uncontrollable and unpredictable dynamic.

The problem of designing a system, whose behavior mimics that of another chaotic system, is called synchronization. Two chaotic systems are usually called drive (master) and response (slave) systems respectively. Different control technique e.g. a chattering-free fuzzy sliding-mode control (FSMC) strategy for synchronization of chaotic systems even in presence of uncertainty has been proposed in [10]. In [11] authors have proposed an active sliding mode control to synchronize two chaotic systems with parametric uncertainty. An algorithm to determine parameters of active sliding mode controller in synchronizing different chaotic systems has been studied in [12]. In [13] an adaptive sliding mode controller has also been presented for a class of master–slave chaotic synchronization systems with uncertainties.

Over the past decade, Backstepping has become the most popular design method for adaptive nonlinear control because it can guarantee global stabilities, tracking, and transient performance for a board class of strict feedback systems. In [14-16], it has been shown that many well-known chaotic

systems as paradigms in research of chaos, including Duffing oscillator, Van der Pol oscillator, Rossler system, Lorenz system, Lü system, Chen system and several type of Chua's circuit, can be transformed into a class of nonlinear system in the so-called nonautonomous form, and the backstepping and tuning functions control schemes have been employed an extended to control these chaotic systems with key parameters unknown. Global stability and tracking have been achieved. In particular, the output of the controlled chaotic system has been designed to asymptotically track any smooth and bounded reference signals generated from a known reference model which may be a chaotic system.

In the following section, the basic definition and numerical methods of solving fractional-order systems is studied. Backstepping method is applied to synchronize the fractional-order chaotic system (Chen) in Section 3. Finally, our findings are summarized in the Conclusion.

The paper is organized as follows; the basic definition and numerical methods of solving fractional-order systems is studied in the next section. Backstepping method is applied to synchronize tow identical fractional-order chaotic system (Chen system as a case study) in Section III. Finally, Our findings are summarized in the Conclusion.

II. FRACTIONAL DERIVATIVE AND ITS APPROXIMATION METHODS

The differ-integral operator ${}_a D_t^q$, a combination of differentiation and integration operator, is commonly used in fractional calculus. This operator represents both the fractional differential equation and the fractional integral in a single expression; [17] defines it:

$${}_a D_t^q = \begin{cases} \frac{d^q}{dt^q} & q > 0 \\ 1 & q = 0 \\ \int_a^t (d\tau)^{-q} & q < 0 \end{cases} \quad (2.1.1)$$

There are several definitions for fractional differential equations [17]. The three most commonly used definitions are

the Grunwald-Letnikov, Riemann-Liouville, and Caputo definitions.

Definition 1. A real function $f(x)$, $x > 0$, is said to be in the space C_μ , $\mu \in \mathbb{R}$ if there exists a real number $p > \mu$ such that $f(x) = x^p f_1(x)$, where $f_1(x) \in [0, \infty)$. Clearly $C_\mu \subset C_\beta$ if $\beta \leq \mu$.

Definition 2. A function $f(x)$, $x > 0$, is said to be in the space C_μ^m , $m \in \mathbb{N} \cup \{0\}$, if $f^{(m)} \in C_\mu$.

Definition 3. The Grunwald-Letnikov fractional differential equation operator of order q [17]:

$${}_a D_t^q f(t) = \frac{d^q f(t)}{d(t-a)^q} = \lim_{N \rightarrow \infty} \left[\frac{t-a}{N} \right]^{-q} \sum_{j=0}^{N-1} (-1)^j f\left(t - j \left[\frac{t-a}{N} \right]\right) \quad (2.2)$$

Definition 4. The left sided Riemann-Liouville fractional differential equation of order $q \geq 0$, of a function $f \in C_q$, $q \geq -1$, is defined as [17]:

$${}_a D_t^q f(t) = \begin{cases} \frac{1}{\Gamma(-q)} \int_a^t (t-\tau)^{-q-1} f(\tau) d\tau & q < 0 \\ f(t) & q = 0 \\ D^n [{}_a D_t^{q-n} f(t)] & q > 0 \end{cases} \quad (2.3)$$

where, n is the smallest integer larger than q , i.e., $n-1 \leq q < n$ and Γ is the Gamma function:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \quad (2.4)$$

For a wide class of functions, the Grunwald-Letnikov and the Riemann-Liouville definitions are equivalent [17].

Definition 5. Let $f \in C_{-1}^m$, $m \in \mathbb{N}$. Then the (left sided) Caputo fractional differential equation of $f(x)$ is defined as [17]:

$${}_0 D_t^q f(t) = \begin{cases} \frac{1}{\Gamma(m-q)} \int_0^t (t-\tau)^{m-q-1} \frac{d^m f(\tau)}{d\tau^m} d\tau, & m-1 < q < m \\ \frac{d^m f(x,t)}{d\tau^m}, & q = m \in \mathbb{N} \end{cases} \quad (2.5)$$

where, m is the smallest integer larger than q . Primarily, the Caputo fractional differential equation computes an ordinary differential equation followed by a fractional integral to achieve the desired order of fractional derivative and then the Riemann-Liouville fractional differential equation is computed in the reverse order. The Caputo fractional differential equation allows traditional initial and boundary conditions to be included in the formulation of the problem, but for homogeneous initial condition assumption, these two operators coincide. For more details on the geometric and physical interpretation for fractional differential equations of both the Riemann-Liouville and Caputo types, see [17].

Unlike the numerical procedure in ordinary differential equations, the numerical evaluation of fractional differential equations is quite complex. One of the approximation methods

of solving FDEs is based on frequency domain which completely discussed in many papers such as [3, 18], so that we withhold to explain it again. The other algorithm to find an approximation for fractional-order systems is based on the predictor-correctors scheme [18]. This method is in essence an improved version of Adams-Bashforth-Moulton algorithm [19-21]. It is based on the predictor-correctors scheme [21, 22]. Although the following proposed numerical procedure has been used to solve some specific problems, it will certainly be used for similar equations. As a practical experience, this method is found as a fundamental algorithm for these types of problems. The method will be explained systematically through some examples. Consider the following fractional differential equation:

$$D^\alpha y(t) = r(t, y(t)), \quad 0 \leq t \leq T, \quad (2.6)$$

$$m-1 < \alpha \leq m,$$

$$y^{(k)}(0) = y_0^{(k)}, \quad k = 0, 1, \dots, m-1.$$

The solution of the above differential equation is equivalent to Volterra integral series [23]:

$$y(t) = \sum_{k=0}^{[\alpha]-1} y_0^{(k)} \frac{t^k}{k!} + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} r(s, y(s)) ds. \quad (2.7)$$

The step size is equally spaced by $h = T/N$ where, $t_n = nh$ ($n = 0, 1, \dots, N$). Then equation (7) can be rewritten as follows:

$$y_h(t_{n+1}) = \sum_{k=0}^{[\alpha]-1} y_0^{(k)} \frac{t^k}{k!} + \frac{h^\alpha}{\Gamma(\alpha+2)} \left\{ r(t_{n+1}, y_h^p(t_{n+1})) + \sum_{j=0}^n a_{j,n+1} r(t_j, y_h(t_j)) \right\}, \quad (2.2.3)$$

where,

$$a_{j,n+1} = \begin{cases} n^{\alpha+1} - (n-\alpha)(n+1)^\alpha, & j=0 \\ (n-j+2)^{\alpha+1} + (n-j)^{\alpha+1} - 2(n-j+1)^{\alpha+1}, & 1 \leq j \leq n \\ 1, & j=n+1 \end{cases} \quad (2.9)$$

$$b_{j,n+1} = \frac{h^\alpha}{\alpha} ((n+1-j)^\alpha - (n-j)^\alpha),$$

and,

$$y_h^p(t_{n+1}) = \sum_{k=0}^{[\alpha]-1} y_0^{(k)} \frac{t^k}{k!} + \frac{1}{\Gamma(\alpha)} \sum_{j=0}^n b_{j,n+1} r[t_j, y_h(t_j)] \quad (2.10)$$

The error of this approximation is of order p , which can be described [18] by following relation

$$O(h^p) = \max_{j=0,1,\dots,N} |y(t_j) - y_h(t_j)| \quad (2.11)$$

where, $p = \min(2, 1 + \alpha)$

III. BACKSTEPPING CONTROL OF FRACTIONAL ORDER CHAOTIC SYSTEMS

In this section we apply a novel method based on the backstepping method to synchronize two identical fractional-

order Chen systems which has been already synchronized via Active Sliding Mode Control in [24].

The Chen system was introduced by Chen and Ueta in 1999 [25]. Some researchers [26-28] have investigated chaotic behavior of the fractional-order Chen system that is described by:

$$\begin{cases} D_t^q x_1 = a(x_2 - x_1) \\ D_t^q x_2 = (c-a)x_1 - x_1x_3 + cx_2 \\ D_t^q x_3 = x_1x_2 - bx_3 \end{cases} \quad (3.1)$$

Considering Eq. (3.1) as master system, slave systems are defined as follows:

$$\begin{cases} D_t^q y_1 = a(y_2 - y_1) + u_1 \\ D_t^q y_2 = (c-a)y_1 - y_1y_3 + cy_2 + u_2 \\ D_t^q y_3 = y_1y_2 - by_3 + u_3 \end{cases} \quad (3.2)$$

Define the track error as:

$$e_i = y_i - x_i \quad (3.3)$$

The error dynamic is described by

$$\begin{cases} D_t^q e_1 = a(e_2 - e_1) + u_1 \\ D_t^q e_2 = (c-a-x_3)e_1 + ce_2 - e_3(e_1 + x_1) + u_2 \\ D_t^q e_3 = e_1x_2 + e_2(e_1 + x_1) - be_3 + u_3 \end{cases} \quad (3.4)$$

In this section, the backstepping design technique is applied to obtain control law of error system (3.4). The design procedure is divided into these steps shown as follows.

Step 1: In this step we consider the stability of the first equation of Eq. (3.4)

$$D_t^q w_1 = a(e_2 - w_1) + u_1 \quad (3.5)$$

where

$$w_1 = e_1.$$

and e_2 and u_1 are controllers.

Choose the first Lyapunov functional candidate as follow

$$V_1 = \frac{1}{2} w_1^2 > 0 \quad (3.6)$$

The time derivative of V_1 along trajectories of error dynamic (3.6) is

$$\begin{aligned} \dot{V}_1 &= w_1 \dot{w}_1 = w_1 D_t^{1-q} (D_t^q w_1) \\ &= w_1 D_t^{1-q} (a(e_2 - w_1) + u_1) \end{aligned} \quad (3.7)$$

Assuming controllers, $e_2 = \alpha_1(w_1)$, $u_1 = aw_1 - k_1 D_t^{1-q} w_1$, Eq. (3.7) can be written as:

$$\dot{V}_1(e_1) = -k_1 w_1^2 + w_1 D_t^{1-q} (a\alpha_1) \quad (3.8)$$

where k_1 is a positive constant. If $\alpha_1(w_1) = 0$, Eq. (3.8) can be rewritten as

$$\dot{V}_1(e_1) = -k_1 w_1^2 \quad (3.9)$$

This means that the zero solution of Eq. (3.6) is asymptotically stable.

Step 2: When e_2 is considered as controllers, $\alpha_1(e_1)$ is estimative function. Defining

$$w_2 = e_2 - \alpha_1(w_1) \quad (3.10)$$

In this step we consider the stability of study (w_1, w_2) system

$$\begin{cases} D_t^q w_1 = a(w_2 - w_1) + u_1 \\ D_t^q w_2 = (c-a-x_3)w_1 + cw_2 - e_3(w_1 + x_1) + u_2 \end{cases} \quad (3.11)$$

Substituting $u_1 = aw_1 - k_1 D_t^{1-q} w_1$ in Eq. (3.11) we obtain

$$\begin{cases} D_t^q w_1 = aw_2 - k_1 D_t^{1-q} w_1 \\ D_t^q w_2 = (c-a-x_3)w_1 + cw_2 - e_3(w_1 + x_1) + u_2 \end{cases} \quad (3.12)$$

where e_3 and u_2 are controllers. Now, we candidate the second Lyapunov function as

$$V_2(w_1, w_2) = V_1(w_1) + \frac{1}{2} w_2^2 > 0 \quad (3.13)$$

The time derivative of V_2 along trajectories of error dynamic (3.9) is

$$\begin{aligned} \dot{V}_2(w_1, w_2) &= -k_1 w_1^2 + w_2 D_t^{1-q} (D_t^q w_2) \\ &= -k_1 w_1^2 + w_2 D_t^{1-q} ((c-a-x_3)w_1 \\ &\quad + cw_2 - e_3(w_1 + x_1) + u_2) \end{aligned} \quad (3.14)$$

Assuming controllers

$$u_2 = -w_1(c-a-x_3) - cw_2 - k_2 D_t^{q-1} w_2, e_3 = \alpha_2(w_1, w_2), \text{ Eq. (3.14) can be written as:}$$

$$\dot{V}_2(w_1, w_2) = -k_1 w_1^2 - k_2 w_2^2 + w_2 D_t^{1-q} (-\alpha_2(w_1 + w_1)) \quad (3.15)$$

where k_2 is a positive constant. If $\alpha_2(w_1, w_2) = 0$, Eq. (3.15) can be written as

$$\dot{V}_2(w_1, w_2) = -k_1 w_1^2 - k_2 w_2^2 < 0 \quad (3.16)$$

This will guaranty that the zero solution of Eq. (3.13) is asymptotically stable.

Step 3: When e_3 is considered as controllers in Eq. (3.12), $\alpha_2(w_1, w_2)$ is estimative function. Defining

$$w_3 = e_3 - \alpha_2(w_1, w_2) \quad (3.17)$$

we study (w_1, w_2, w_3) system

$$\begin{cases} D_t^q w_1 = a(w_2 - w_1) + u_1 \\ D_t^q w_2 = (c - a - x_3)w_1 + cw_2 - w_3(w_1 + x_1) + u_2 \\ D_t^q w_3 = w_1x_2 + w_2(w_1 + x_1) - bw_3 + u_3 \end{cases} \quad (3.18)$$

Substituting $u_1 = a w_1 - k_1 D_t^{1-q} w_1$ and

$u_2 = -w_1(c - a - x_3) - cw_2 - k_2 D_t^{q-1} w_2$ in Eq. (3.18) we obtain:

$$\begin{cases} D_t^q w_1 = a w_2 - k_1 D_t^{1-q} w_1 \\ D_t^q w_2 = -k_2 D_t^{q-1} w_2 - w_3(w_1 + x_1) \\ D_t^q w_3 = w_1x_2 + w_2(w_1 + x_1) - bw_3 + u_3 \end{cases} \quad (3.19)$$

where u_3 is controller. Finally, we candidate the third Lyapunov function as

$$V_3(w_1, w_2, w_3) = V_1(w_1) + V_2(w_1, w_2) + \frac{1}{2} w_3^2 > 0 \quad (3.20)$$

Its time derivative is

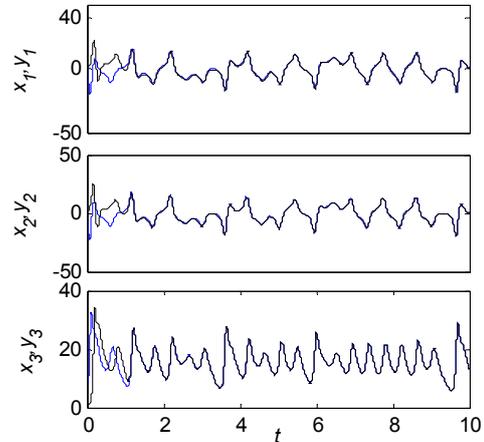
$$\begin{aligned} \dot{V}_3(w_1, w_2, w_3) &= -k_1 w_1^2 - k_2 w_2^2 + w_3 \cdot D_t^{1-q} (D_t^q w_3) \\ &= -k_1 w_1^2 - k_2 w_2^2 + \\ &\quad w_3 \cdot D_t^{1-q} (w_1x_2 + w_2(w_1 + x_1) - bw_3 + u_3) \end{aligned} \quad (3.21)$$

Assuming controller, $u_3 = -w_1x_2 - w_2(w_1 + x_1) + bw_3 - k_3 D_t^{q-1} w_3$, Eq. (3.21) can be rewritten as:

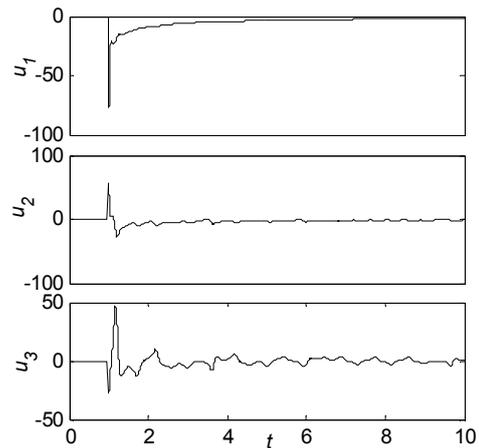
$$\dot{V}_3(w_1, w_2, w_3) = -k_1 w_1^2 - k_2 w_2^2 - k_3 w_3^2 < 0 \quad (3.22)$$

where k_3 is a positive constant. The Eq. (3.22) guaranties that the controllers i.e. u_1, u_2 , and u_3 will stabilize the Eq. (3.4). We applied Backstepping control to synchronize two fractional-order Chen systems in the following two cases: Case 1: order of chaotic systems is 2.85 ($q = 0.95$) and $(a, b, c) = (40, 3, 28)$; Case 2: order of chaotic systems is 2.7 ($q = 0.9$) and $(a, b, c) = (35, 3, 28)$. The controller parameter in Case 1, and Case 2, are chosen as $(k_1, k_2, k_3) = (20, 10, 10)$ and $(k_1, k_2, k_3) = (10, 10, 10)$ respectively. Initial conditions are chosen $(x_{10}, x_{20}, x_{30}) = (3, -6, 9)$ and $(y_{10}, y_{20}, y_{30}) = (1, 1, 1)$. Numerical simulations have carried out using the SIMULINK based on the frequency domain approximation. To solve the sets of fractional-order differential equations related to the master and slave systems, the CRONE Toolbox is used and Runge-Kutta solver with fixed step size 0.0001 is used. Fig. (3) shows simulation results of Case No. 1. Simulation results of Case No. 2 are given in Fig. (4). As it can be seen from Figs. (3.a) and (4.a), the designed controller is successfully able to synchronize two identical fractional-order Chen systems. It has been shown that all of the state variables of the slave converge to that of the

master. The simulation results verify the performance of the Backstepping controller.



(a)

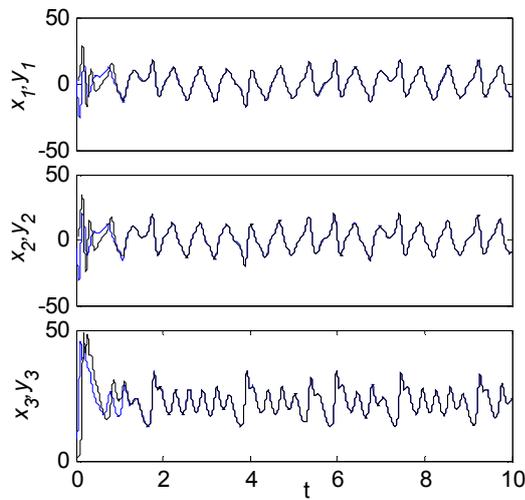


(b)

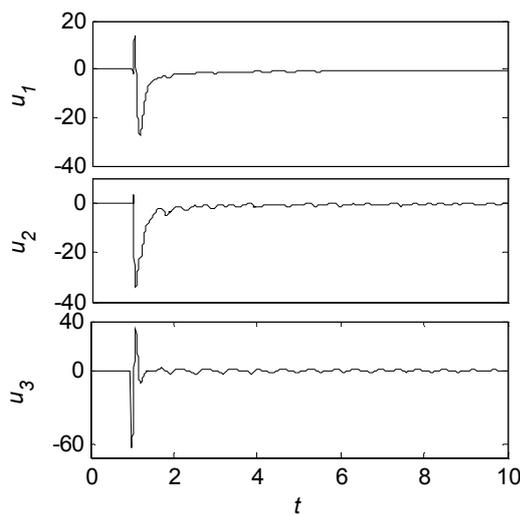
Fig. (3) Simulation Results of synchronization of two identical Chen systems (Case 1)

IV. CONCLUSION

In this paper we applied a proposed controller based on backstepping method to synchronize fractional order Chen systems. The simulation results show that this method can easily control fractional-order systems. As it can be seen from figures all of the control signal are feasible and can be applied to the other fractional-order systems such as fractional-order Hyperchaotic systems.



(a)



(b)

Fig. (3) Simulation Results of synchronization of two identical Chen systems (Case 2)

REFERENCES

- [1] T.T. Hartley, C.F. Lorenzo, H.K. Qammer, Chaos in a fractional order Chua's system, *IEEE Trans. CAS-I* 1995; 42: 485–490.
- [2] P. Arena, R. Caponetto, L. Fortuna, D. Porto, Chaos in a fractional order Duffing system, in: *Proceedings ECCTD, Budapest, 1997*; 1259–1262.
- [3] W.M. Ahmad, J.C. Sprott, Chaos in fractional-order autonomous nonlinear systems, *Chaos, Solitons Fractals* 2003; 16: 339–351.
- [4] J.G. Lu, G. Chen, A note on the fractional-order Chen system, *Chaos, Solitons Fractals* 2006; 27 (3): 685–688.
- [5] J.G. Lu, Chaotic dynamics of the fractional-order Lü system and its synchronization, *Phys. Lett. A* 2006; 354 (4): 305–311.

- [6] C. Li, G. Chen, Chaos and hyperchaos in the fractional-order Rössler equations, *Phys. A: Stat. Mech. Appl.* 2004; 341: 55–61.
- [7] J.G. Lu, Chaotic dynamics and synchronization of fractional-order Arneodo's systems, *Chaos, Solitons Fractals* 2005; 26 (4): 1125–1133.
- [8] L.J. Sheu, H.K. Chen, J.H. Chen, L.M. Tam, W.C. Chen, K.T. Lin, Y. Kang, Chaos in the Newton-Leipnik system with fractional order, *Chaos, Solitons & Fractals* 2008; 36(1): 98–103.
- [9] Shu-Yong Liu, Xiang Yu and Shi-Jian Zhu, Study on the chaos anti-control technology in nonlinear vibration isolation system, *Journal of Sound and Vibration*, 4 March 2008; 310 (4-5): 855–864.
- [10] W. Deng, C.P. Li, Chaos synchronization of the fractional Lu system, *Phys. A* 2005; 353: 61–72.
- [11] W. Deng, C. Li, Synchronization of chaotic fractional Chen system, *J. Phys. Soc. Jpn* 2005; 74 (6): 1645–1648.
- [12] C.P. Li, W.H. Deng, D. Xu, Chaos synchronization of the Chua system with a fractional order, *Phys. A* 2006; 360: 171–185.
- [13] V.I. Utkin, Variable structure systems with sliding mode, *IEEE Trans. Automat. Contr.*, 1977; 22: 212–222.
- [14] Jian Zhang, Chunguang Li, Hongbin Zhang, Juebang Yu, Chaos synchronization using single variable feedback based on backstepping method *Chaos, Solitons & Fractals*, 21(5) (2004) 1183–1193.
- [15] Yongguang Yu, Suochun Zhang, Adaptive backstepping synchronization of uncertain chaotic system, *Chaos, Solitons & Fractals*, 21(3) (2004) 643–649.
- [16] Xiaohui Tan, Jiye Zhang, Yiren Yang, Synchronizing chaotic systems using backstepping design, *Chaos, Solitons & Fractals*, 16(1) (2003) 37–45.
- [17] I. Podlubny, *Fractional Differential Equations*, Academic Press, New York, 1999.
- [18] Mohammad Saleh Tavazoei, Mohammad Haeri, A necessary condition for double scroll attractor existence in fractional-order systems, *Phys. Lett. A*, 367 (2007) 102–113.
- [19] K. Diethelm, N.J. Ford, A.D. Freed, Detailed error analysis for a fractional Adams method, *Numer. Algorithms*, 36 (2004) 31–52.
- [20] C. Li, G. Peng, Chaos in Chen's system with a fractional order, *Chaos Solitons Fractals* 22 (2004) 443–450.
- [21] K. Diethelm, N.J. Ford, A.D. Freed, A predictor–corrector approach for the numerical solution of fractional differential equations, *Nonlinear Dyn.* 29 (2002) 3–22.
- [22] K. Diethelm, An algorithm for the numerical solution of differential equations of fractional order, *Electron. Trans. Numer. Anal.* 5 (1997) 1–6.
- [23] K. Diethelm, N.J. Ford, Analysis of fractional differential equations, *J. Math. Anal. Appl.* 265 (2002) 229–248.
- [24] M. S. Tavazoei, M. Haeri, Synchronization of chaotic fractional-order systems via active sliding mode controller, *Physica A* 2008; 387: 57–70.
- [25] G. Chen, T. Ueta, Yet another attractor, *Int. J. Bifurcation Chaos* 9 (1999) 1465–1466.
- [26] J.G. Lu, G. Chen, A note on the fractional-order Chen system, *Chaos, Solitons Fractals* 27 (3) (2006) 685–688.
- [27] C. Li, G. Peng, Chaos in Chen's system with a fractional order, *Chaos, Solitons Fractals* 22 (2004) 443–450.
- [28] C. Li, G. Peng, Chaos in the fractional order Chen system with and its control, *Chaos, Solitons Fractals* 22 (2004) 549–554.

Control of Genesio-Tesi and Chen Chaotic Systems Using a Fractional-Order Controller

M. Mahmoudian*, A. Ranjbar*, E. Naseri*, S.H. Hosseinnia*, S. Momani**
**Intelligent system research group, Faculty of Electrical and Computer Engineering,
 Noushirvani University of Technology, Babol, Iran, P.O. Box 47135-484,
 (a.ranjbar@nit.ac.ir), (h.hosseinnia@stu.nit.ac.ir)*
*** Department of Mathematics, Mutah University, P.O. Box: 7, Al-Karak, Jordan
 (shaherm@yaho.com)*

Abstract- In order to control *Genesio-Tesi* and *Chen* chaotic systems, a fractional controller has been presented. This controller has been composed of a combination of fractional derivative and an integer derivative in form of $s - s^\alpha$. This kind of controller, the system with integer derivatives is turned into a system with fractional ones. The idea is to expand the stability region of fractional system because of using the fractional derivative. A proper range of the fraction parameter will also be designed to stabilize the closed loop system. The performance of the proposed idea is shown through the simulation.

I. INTRODUCTION

Fractional calculus is an old mathematical topic from 17th century. Nowadays, It has been found that in interdisciplinary fields, many systems can be described by fractional differential equations. For example: the fractional-order Chua circuit [1] Duffing system [2] jerk model [3] Chen system [4] Lü dynamic [5] Rössler system [6] and Newton–Leipnik system [7] demonstrate a chaotic behaviour. Chaotic systems have recently been much considered due to their potential usage in different fields of science and technology particularly in electronic systems [8] secure communication [9] computer [10]. To develop the chaotic theory, a control approach of chaotic systems has been considered as an important problem [11]. It was initially assumed that, control of chaotic systems is impossible and they have uncontrollable and unpredictable dynamic. The imagination was changed when three researchers [12] have shown other vice. The Endeavour has been proceeded to control the chaos using different approach, e.g. feedback linearization [13-15] Delayed feedback control [16] OPF [17] and TDFC [18].

Recently, some scientist applied the fractional-order controller to control fractional and integer order dynamics of chaos. In [19, 20] an Adaptive fractional control is proposed to control and synchronize chaos [19, 20], and controller parameter is updated based on a proper adaptation mechanism the. A sliding mode control and active control is presented to synchronize the fractional-order chaotic system [21,22]. In [23], a simple fractional controller has been proposed to control such chaotic systems. In this method, a system with integer derivative turns in to a system with fractional derivatives. Thereafter a proper interval for α will be designed to stabilize the closed loop system. This method will be applied to control *Genesio-Tesi* and *Chen* chaotic systems, in this paper.

This paper is organized as follows:

In section II, basic stability requirement for fractional systems will be shortly described. A brief description of fractional controller will be presented in section III. Section IV is devoted to implement the proposed method on *Genesio-Tesi* and *Chen* chaotic systems. Ultimately, the work will be concluded at section V.

II. STABILITY ANALYSIS OF FRACTIONAL-ORDER SYSTEM

A fractional order linear time invariant (FO-LTI) system can be represented by the following state-space format:

$$\begin{cases} D^\alpha x = Ax + Bu \\ y = Cx \end{cases} \quad (1)$$

where, $x \in \mathbb{R}^n$, $u \in \mathbb{R}$ and $y \in \mathbb{R}^r$ denote states, input and output vectors of the system. The appropriate coefficients will be shown by $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times r}$ and $C \in \mathbb{R}^{r \times n}$ respectively whilst α is the fractional commensurate order. Fractional order differential equations are at least as stable as their integer orders counterparts. This is because; systems with memory are typically more stable than their memory-less alternatives [24-26]. It has been shown that an autonomous dynamic $D^\alpha x = Ax$, $x(0) = x_0$ is asymptotically stable if the following condition is met [27]:

$$|\arg(\text{eig}(A))| > \alpha\pi/2, \quad (2)$$

where, $0 < \alpha < 1$ and $\text{eig}(A)$ represents the eigenvalues of matrix A . This means each component of states decays towards 0, like $t^{-\alpha}$. In addition, this system is stable if $|\arg(\text{eig}(A))| \geq \alpha\pi/2$. Those critical eigenvalues which satisfy $|\arg(\text{eig}(A))| > \alpha\pi/2$ have geometric multiplicity of 1. The equality holds when the dynamic has not geometric multiplicity of 1.

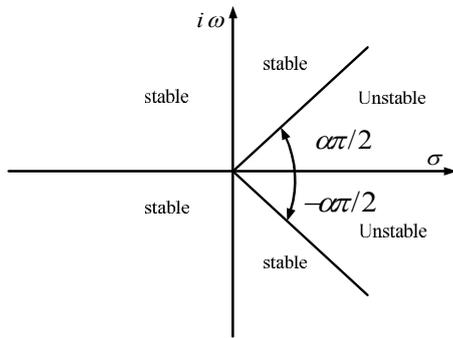


Fig. 1. Stability region of the FO-LTI system with fractional order, $0 < \alpha < 1$

The stability region for $0 < \alpha < 1$ are shown in Figure 1. Now, consider the following autonomous commensurate fractional order system:

$$D^\alpha x = f(x), \quad (3)$$

where, $0 < \alpha < 1$ and $x_2 \in R^n$. The equilibrium points of system (3) are calculated when:

$$f(x) = 0. \quad (4)$$

These points are locally asymptotically stable if all eigenvalues of Jacobian matrix $J = \partial f / \partial x$, which are evaluated at the equilibrium points- satisfy the following condition [27-28]:

$$|\arg(\text{eig}(J))| > \alpha\pi/2. \quad (5)$$

III. A FRACTIONAL CONTROLLER, $S - S^\alpha$

A controller is often used to stabilize the system or improve the performance indices. A classic stabilizing treatment may be found by a pole placement approach. This may be achieved when a state feedback controller assuming availability of the state is applied. An alternative fractional controller [23], expands the effective range of stability according to equation (5). In this case, there is no need to locate the poles in other places explicitly. A normal integer type controller will be substituted with a fractional derivative. In Figure (2) a schematic diagram of the proposed closed loop fractional controller is shown.

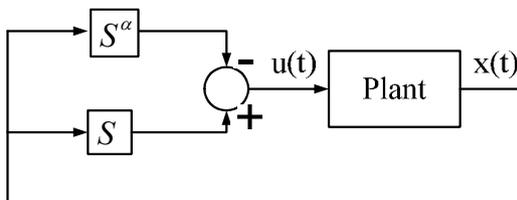


Fig. 2. schematic diagram of closed loop fractional controller

Consider a chaotic system can be represented by the following dynamic:

$$\dot{x} = f(x) + u \quad (6)$$

A fractional controller alters the dynamic in (6) into the following autonomous dynamic:

$$D^\alpha x = f(x) \quad (7)$$

To stabilize the overall system, a proper interval for α can be established. If λ_i shows the i^{th} eigenvalues of system, an upper bound for α can be written as [27-28]:

$$\tan(\alpha\pi/2) < \left| \frac{\text{Im}(\lambda_i)}{\text{Re}(\lambda_i)} \right| \Rightarrow \alpha < \frac{2}{\pi} \tan^{-1} \left| \frac{\text{Im}(\lambda_i)}{\text{Re}(\lambda_i)} \right| \quad (8)$$

In the following section, this controller is implemented on a Genesio-Tesi and Chen chaotic systems.

IV. IMPLEMENTATION OF THE FRACTIONAL CONTROLLER ON CHAOTIC SYSTEMS

IV.I. CONTROL OF CHEN CHAOTIC SYSTEM

Chen system was originally introduced by in [29]. Some researchers [30-31] have used the results and also investigated the chaotic behaviour of the fractional-order. Chen system that is described by:

$$\begin{cases} \frac{dx_1}{dt} = A(x_2 - x_1) \\ \frac{dx_2}{dt} = (g - A)x_1 - x_1x_3 + gx_2 \\ \frac{dx_3}{dt} = x_1x_2 - Bx_3 \\ y = x_2 \end{cases} \quad (9)$$

Parameters A, B and g are chosen respectively as 40, 3 and 31.

In order to investigate the stability of the system, Jacobian matrix and corresponding Eigenvalues have to be found. These in return also require the equilibrium points, which are as follows:

$$\begin{cases} (x_{1e}, x_{2e}, x_{3e}) = (0, 0, 0) \\ (x_{1e}, x_{2e}, x_{3e}) = (8.12, 8.12, 22) \\ (x_{1e}, x_{2e}, x_{3e}) = (-8.12, -8.12, 22) \end{cases}$$

The appropriate Jacobian matrix is as follows:

$$J = \begin{bmatrix} -40 & 40 & 0 \\ -X_{3e} - 9 & 31 & -X_{1e} \\ X_{2e} & X_{1e} & -3 \end{bmatrix}$$

The corresponding Eigenvalues at each equilibrium points are evaluated as:

$$\begin{cases} (25.5, -34.5, -3) \\ (-20.2, 4.1 + 15.5i, 4.1 - 15.5i) \\ (-20.2, 4.1 + 15.5i, 4.1 - 15.5i) \end{cases}$$

It is seen that first set of eigenvalues show instability of the system and therefore instability of the corresponding equilibrium point i.e. the origin. Therefore, according to equation (8), α must satisfy $\alpha < 0.83$ to maintain the stability of the system. A fractional controller for $\alpha = 0.8$ together with initial states equal to $X_1(0) = -1.0032$, $X_2(0) = 2.3545$ and $X_3(0) = -0.87$ is chosen. This controller is applied on *Chen* chaotic system. Corresponding results are shown in Figures 3 to 5. The controlled states and control input are shown in Figure (3) and (4) respectively. The control action is triggered at $t = 5$ s. As it can be seen, the states are got settled and stabilized soon after. To show the effectiveness of the controller, the states are also shown in the phase plane in Figure (5). It can be seen that the states are converged to the stable equilibrium point, i.e. $(X_1(0) = -20, X_2(0) = -10, X_3(0) = 6)$. Similarly, the performance of the proposed controller will also be verified on other chaotic system e.g. *Genesio-Tesi*.

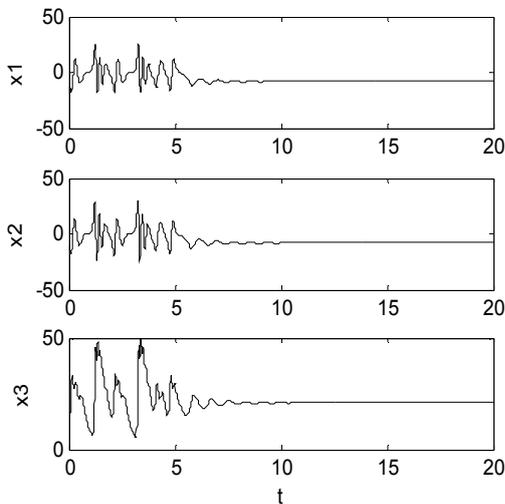


Fig. 3. The states of *Chen* chaotic system under a fractional controller

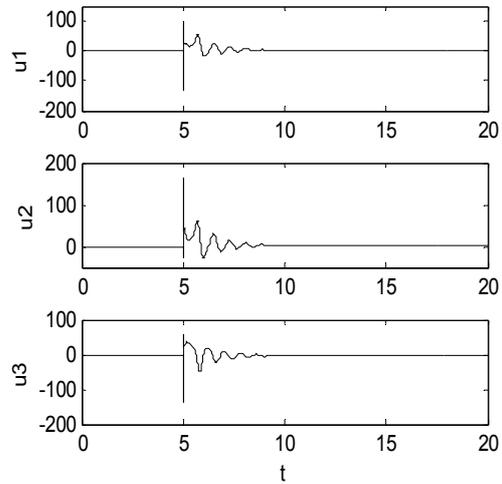


Fig. 4. Control Input in *Chen* Chaotic system

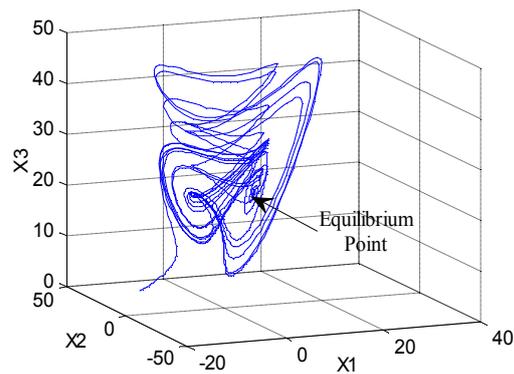


Fig. 5. Phase portrait of *Chen* chaotic system under a fractional controller

IV.I. CONTROL OF GENESIO-TESI CHAOTIC SYSTEM

The effectiveness of the controller will be shown when it is applied on the other system i.e. *Genesio-Tesi*. Consider the following *Genesio-Tesi* chaotic system [32]:

$$\begin{cases} \frac{dx_1}{dt} = x_2 \\ \frac{dx_2}{dt} = x_3 \\ \frac{dx_3}{dt} = -cx_1 - bx_2 - ax_3 + x_1^2 \end{cases} \quad (10)$$

Corresponding parameters are selected as: $a = 1.2, b = 2.92, c = 6$. The same procedure will be done to the system. To investigate the stability, equilibrium point, Jacobian

matrix and corresponding Eigenvalues are computed. Equilibrium point are found as:

$$\begin{cases} (x_{1e}, x_{2e}, x_{3e}) = (0, 0, 0) \\ (x_{1e}, x_{2e}, x_{3e}) = (6, 0, 0) \end{cases}$$

Then Jacobian matrix is found as:

$$J = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 + 2x_{1e} & -2.92 & -1.2 \end{bmatrix}$$

Corresponding eigenvalues are then found at the equilibrium point as:

$$\begin{cases} (-1.64, 0.22 + 1.89i, 0.22 - 1.89i) \\ (1.1, -1.15 + 2.03i, -1.15 - 2.03i) \end{cases}$$

The stability range for α is found using equation (8) as $\alpha < 0.92$. A simple fractional controller assuming $\alpha = 0.8$ and $X_1(0) = -1.0032$, $X_2(0) = 2.3545$, $X_3(0) = -0.087$ is implemented on the chaotic system. Controlled states and control input are shown in Figure (6) and (8) respectively, when the control takes action at $t=10$ s. The phase portrait is also plotted in Figure (7). From this graphs, the stability of the controlled system and therefore the effectiveness of the proposed controller can be seen.

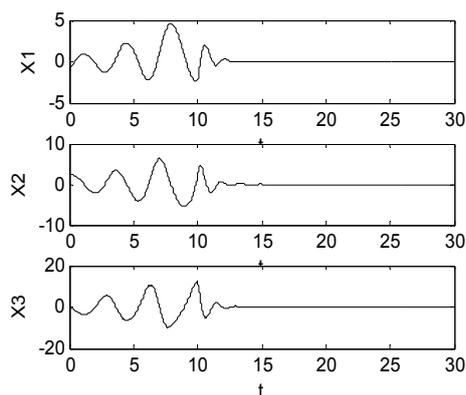


Fig. 6. The state transient behaviour of *Genesis-Tesi* when the control is triggered at $t=10$ s.

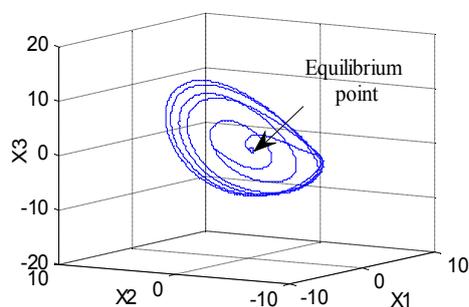


Fig. 7. Phase portrait of *Genesis-Tesi* chaotic system under a fractional controller

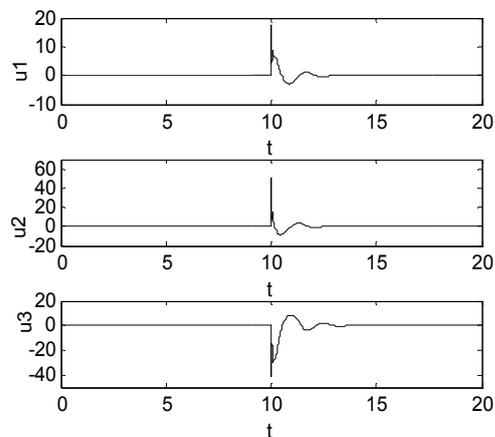


Fig. 8. Control Input in *Genesis-Tesi* Chaotic system

V. CONCLUSION

Significance of a fractional controller is shown on two *Chen* and *Genesis-Tesi* chaotic systems. In addition of simplicity of the controller, maintaining the stability are the advantages of the proposed controller. The range of the stability is increased without need for a pole placement. This is achieved using a simple fractional controller. The convergence and the stability of the closed loop are shown achieved on two chaotic systems, through simulation.

REFERENCES

- [1] Hartley T.T, C.F. Lorenzo, H.K. Qammer (1995), Chaos in a fractional order Chua's system, IEEE Trans. 42,485-490.
- [2] Arena. P, R. Caponetto, L. Fortuna, D. Porto (1997), Chaos in a fractional order Duffing system, in: Proceedings ECCTD Budapest, 1259-1262.
- [3] Ahmad W. M, J.C. Sprott (2003), Chaos in fractional-order autonomous nonlinear systems, Chaos, Solitons Fractals, 16, 339-351.
- [4] Lu. J.G., G. Chen (2006), A note on the fractional-order Chen system, Chaos, Solitons Fractals, 27 (3), 685-688.
- [5] Lu. J.G (2005), Chaotic dynamics and synchronization of fractional-order Arneodo's systems, Chaos, Solitons Fractals, 26 (4), 1125-1133.
- [6] Li. C, G. Chen (2004), Chaos and hyperchaos in the fractional-order Rossler equations, Physic Review Letter. A, 341, 55-61.
- [7] Sheu. L.J., H.K. Chen, J.H. Chen, L.M. Tam, W.C. Chen, K.T. Lin, Y. Kang (2006), Chaos in the Newton-Leipnik system with fractional order, Chaos, Solitons Fractals, 06.013
- [8] Roy R., T. W. Murph, T. D. Maier, Z. Gillis and E. R. Hunt (2000), Control of a chaotic Layer: Experimental stabilization of a Globally Coupled System, Physic Review Letter 68 1259-1262.
- [9] Yang .T (2004) "A survey of chaotic secure communication systems" Int. Journal of Computational Cognition 2(2) 81-130.
- [10] Chen .L, X. F. Wang and Z. Z. Han (2003) "Controlling Bifurcation and Chaos in internet Congestion Control Method, Proceedings of the 2003 International Symposium on Circuits and Systems , 3, 132-135.
- [11] Liu Shu-Yong, Xiang Yu and Shi-Jian Zhu(2008), Study on the chaos anti-control technology in nonlinear vibration isolation system, Journal of Sound and Vibration, 310(4-5), 855-864.

- [12] Ott E., Grebogi, C., York, J. A (1990), Controlling Chaos, *Physic Review Letter*, 64, 1196-99.
- [13] Gallegos JA. (1994), Nonlinear regulation of a Lorenz system by feedback linearization techniques, *Dyn Control*, 4, 277-98.
- [14] Yassen MT (2005). Controlling chaos and synchronization for new chaotic system using linear feedback control, *Chaos Solitons Fract*, 26, 913-20.
- [15] F. Abdous, A. Ranjbar N., S.H. Hosein Nia, A. Sheikhol Eslami, B. Abdous (2008), The Hopf Bifurcation control in Internet Congestion Control System, Second International Conference on Electrical Engineering (ICEE), Lahore (Pakistan).
- [16] F. Abdous, A. Ranjbar N., S.H. Hosein Nia, A. Sheikhol Eslami, (2008), Chaos Control of Voltage Fluctuations in DC Arc Furnaces Using Time Delay Feedback Control, Second International Conference on Electrical Engineering (ICEE), Lahore (Pakistan).
- [17] E. R Hunt (1991), Stabilization High-Period Orbits in a Chaos System, *Physic Review Letter*, 67, 1953-55.
- [18] Pyragas k (1991), Continuous control of chaos by self controlling feedback, *Physic Review Letter*, 67, 1953-55.
- [19] Hosseinnia, S. H., R. Ghaderi, A. Ranjbar N., J. Sadati, S. Momani, Synchronization of Fractional Chaotic Systems Via Fractional-Order Adaptive Controller, in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [20] Hosseinnia, S. H., R. Ghaderi, A. Ranjbar N., J. Sadati, S. Momani, Designation of an adaptive PID controller to synchronize fractional Corder hard spring Phi⁶-Van der Pol oscillator, in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [21] Hosseinnia, S. H., Reza Ghaderi, Abolfazl Ranjbar, N., M. Mahmoudian, Shaher Momani, Synchronization of Uncertain Fractional-Order Duffing-Holmes Chaotic System Via Sliding Mode Control, in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [22] M. Shahiri T., A. Ranjbar N., R. Ghaderi, S. H. Hosseinnia, S. Momani, Control and Synchronization of Chaotic Fractional-Order Coulet System Via Active Controller, in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [23] Tavazoei M. S. and Mohammad Haeri (2008). Chaos control via a simple fractional-order controller *Physics Letters A*, 372(6), 798-80.
- [24] Ahmed.E., A.M.A. El-Sayed, H.A.A. El-Saka (2007), Equilibrium points, stability and numerical solutions of fractional order predator-prey and rabies models, *J. Math. Anal.* 325 (1), 542-553.
- [25] Hosein Nia S.H., A.N. Ranjbar, D.D. Ganji, H. Soltani, J. Ghasemi (2008a), Maintaining the Stability of Nonlinear Differential Equations by the Enhancement of HPM, *Physics Letters A*, 372 2855-2861.
- [26] Hosseinnia S.H., A.N. Ranjbar, D.D. Ganji, H. Soltani, J. Ghasemi (2008b), A Solution of Riccati Nonlinear Differential Equation Using Enhanced Homotopy Perturbation method (EHPM), *International Journal of Engineering*, 21(1), 27-38.
- [27] Matignon.D (1996), Stability results for fractional differential equations with applications to control processing, in: *Computational Engineering in Systems Applications*, IEEE-SMC, 2, 963-968.
- [28] Ahmed.E., A.M.A. El-Sayed, H.A.A. El-Saka (2007), *J. Math. Anal. Appl.*, 325 (1), 542-553.
- [29] Chen .G, T. Ueta (1999), Yet another attractor, *Int. J. Bifurcation Chaos*, 9, 1465-1466.
- [30] Li. C, G. Peng (2004), Chaos in Chen's system with a fractional order, *Chaos, Solitons Fractals*, 22, 443-450.
- [31] Li. C, G. Peng (2004), Chaos in the fractional order Chen system with and its control, *Chaos, Solitons Fractals*, 22, 549-554.
- [32] Genesio R, Tesi A. A harmonic balance methods for the analysis of chaotic dynamics in nonlinear systems. *Automatica* 1992;28:531-48.

Synchronization of Chaotic Fractional-Order Coulet System via ASMC

M. Shahiri T.* , A. Ranjbar*, R. Ghaderi*, S.H. Hosseinnia*, S. Momani**

*Intelligent system research group, Faculty of Electrical and Computer Engineering,
Noushirvani University of Technology, Babol, Iran, P.O. Box 47135-484,
(a.ranjbar@nit.ac.ir), (h.hosseinnia@stu.nit.ac.ir)

**Department of Mathematics, Mutah University, P.O. Box: 7, Al-Karak, Jordan
(shaherm@yaho.com)

Abstract- The main objective of this study is to investigate on Chaotic behavior of fractional-order modeled Coulet system and its controllability. It has been shown that this problem could lead to synchronization of two master and slave systems with the same or different fractional-order. The proposed method which is based on active sliding mode control (ASMC), has been developed to synchronize two chaotic systems with the same or partially different attractor. The numerical simulation results, verify the significance of the proposed controller even for chaotic synchronization task.

I. INTRODUCTION

Although Fractional Order Calculus (FOC) has 300-year of history, its applications in physics and engineering have just begun [1]. In many systems, such as viscoelastic systems [2], dielectric polarization, and electromagnetic waves, FOC models exhibited better utility. Furthermore, emergence of effective analytical and numerical methods in differentiation and integration of non-integer (fractional) order equations, in recent years, makes FOC more attractive for the control systems community.

Recently the interest of chaotic synchronization has been extensively growth [3-7]. The fact that [8] nonlinear chaotic systems may keep their natural chaotic behavior when their models become fractional has a critical effect in this manner.

A pioneering work on the concept of "chaotic synchronization" is presented in [9]. Another's work has been continued through presentation of a method to synchronize two identical chaotic systems with different initial conditions [10]. Different types of chaotic synchronization methods in terms of *complete synchronization*, *generalized synchronization*, *phase synchronization* and *lag synchronization* have been reported. [11-16]

Recently, some researchers applied the fractional-order controller to control fractional and integer order dynamics of chaotic systems. In [17,18] an adaptive fractional controller is proposed to control and synchronize chaos, and controller parameter is updated based on a proper adaptation mechanism. A sliding mode control and active control is presented to synchronize the fractional-order chaotic system in [19,20]. Fractional sliding mode is proposed to control chaos in [21], and a fractional controller in combination with state feedback is proposed in [22]. In [23] Bifurcation in fractional order

Newton-Leipnik system was investigated. The projective method is used to synchronize fractional order rigid body system in [24].

Also synchronization of chaotic fractional order Coulet systems has already studied in [20]. In this paper the synchronization of this system will be developed via active sliding mode control with the novelty of synchronizing of two master and slave systems with different fractional orders. This case attracts the designers, particularly in coding and decoding applications.

Active sliding mode control technique is a discontinuous strategy that relies on two stages of designation. The first step is to select an appropriate active controller to facilitate the design of sequent sliding mode controller. The second stage is to design a sliding mode controller to achieve the synchronization. Although, this method is already applied to synchronize fractional order chaotic Lü and Chen systems [25], there has been a lack of report in fractional order chaotic Coulet system.

This paper is organized as follows:

Coulet system will be described in section 2. Active sliding mode control will be presented in section 3. This controller is applied to synchronize two identical fractional-order Coulet systems in section 4. In section 5, the proposed method will be developed to synchronize two master and slave systems with different fractional-orders. Ultimately, the work will be concluded at section 5.

II. SYSTEM DESCRIPTION

Some of nonlinear systems represent deterministic treatment called *chaotic behavior*. These systems are very sensitive to initial conditions. This means two identical distinct systems but with a minor deviation in their initial condition, may result completely different. In the other words, having bounded initial conditions known, would not guarantee to predict the behavior correctly.

Consider the following Arnéodo-Coulet dynamic equation [26]:

This system has exhibited chaotic dynamics for various values of four parameters [26,27]. To investigate the chaotic behavior

of the system, simulation with the following set $\{a=0.8, b=-1.1, c=-0.45, d=-1\}$ has been considered.

The system in (1) will be written in the state space format as:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = cx_3 + bx_2 + ax_1 + dx_1^3 \end{cases} \quad (2)$$

The phase portrait of the system, represented in Fig.1, shows the chaotic behavior. This would cause the synchronization task to be hard and complex.

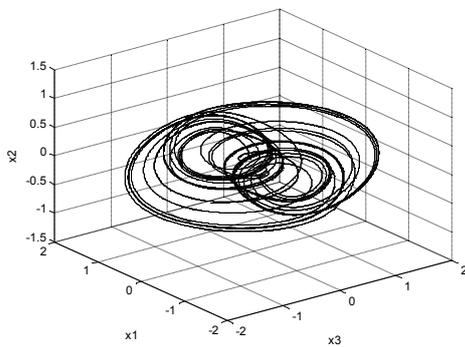


Fig. 1. Phase portrait of Chaotic Coulet system

Although some works have been reported on chaotic Coulet system, there is still less report on fractional order Coulet system[20]. In this research, an active sliding mode controller is proposed on fractional order chaotic system.

To construct the fractional order Coulet chaotic system model, equation (3) will be used, in which q , the fractional commensurate factor will be changed in accordance to system behavior in practice. It can be shown that for some range of q , the fractional order Coulet system is unstable [28]. A resonance property of fractional order Coulet chaotic system is shown in Fig.2 for different values of $q=0.97, 0.95$, and 0.9 . Based on the type of practical chaotic behavior of system, the appropriate value of q could be chosen in modeling stage.

The numerical simulations have carried out based on the frequency domain approximation.

$$\begin{cases} D^q x_1 = x_2 \\ D^q x_2 = x_3 \\ D^q x_3 = cx_3 + bx_2 + ax_1 + dx_1^3 \end{cases} \quad (3)$$

The primary task of this research is to stabilize system, (as it will be shown) using an active sliding mode control design.

III. ACTIVE SLIDING MODE CONTROLLER DESIGN

This method has already been applied to synchronize fractional order chaotic Lü and Chen systems [25].

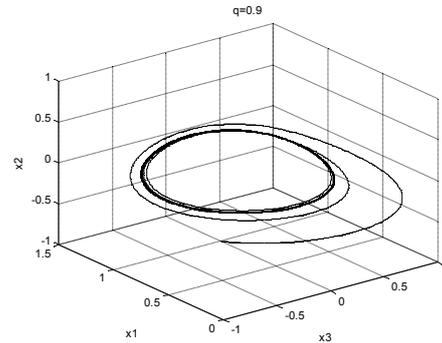
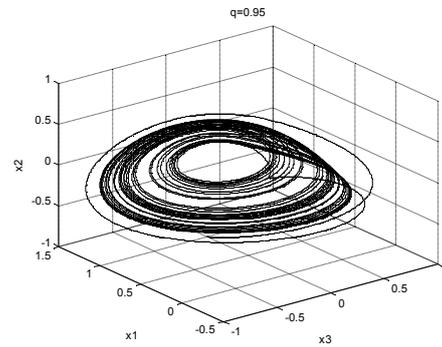
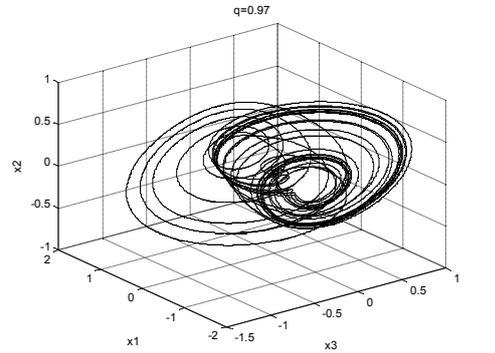


Fig. 2. Phase portrait of chaotic Coulet system vs. different values of the fraction parameter

The subsequent objective is to show the efficiency the fractional modeling of dynamic. The method is basically a combination of an active and sliding mode controller. The designation procedure of active sliding mode controller is primarily given and then the stability issue of the proposed method is proven. Consider the following nonlinear incommensurate chaotic fractional system of order q assuming ($0 < q < 1$):

$$D^q x = Ax + g(x) \quad (4)$$

Where $x \in \mathbb{R}^3$ denotes 3-D state vector, $A \in \mathbb{R}^{3 \times 3}$ represents the linear part of the dynamic and $g: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is the nonlinear part of the system. The procedure description uses a

synchronization architecture where equation (4) represents the master dynamic. Meantime, the slave dynamic is defined inclusion of a control signal $u(t) \in \mathbb{R}^3$ by:

$$D^q y = Ay + h(y) + u(t) \quad (5)$$

where $y \in \mathbb{R}^3$ is the slave 3-D state vector, A is the same parameter matrices as the master has, and $h: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ simply has the same role as g in the master. Synchronization means finding the appropriate control signal $u(t) \in \mathbb{R}^3$ to derive states of the slave system to evolve as the states of the master. The synchronization goal will be achieved through the error definition which is as follows:

$$D^q e = Ae + h(y) - g(x) + u(t) \quad (6)$$

where $e = y - x$. The controller law ($u(t) \in \mathbb{R}^3$) should cause:

$$\lim_{t \rightarrow \infty} \|e(t)\| = 0 \quad (7)$$

In accordance with the active control design procedure [29-31], nonlinear part of the error dynamic is eliminated by the following choice of the input vector:

$$u(t) = H(t) - h(y) + g(x) \quad (8)$$

The error in (6) can be rewritten as:

$$D^q e = Ae + H(t) \quad (9)$$

Equation (9) describes the error with a recently defined control input $H(t)$. In active sliding mode control, $H(t)$ is designed based on a sliding mode control law, as:

$$H(t) = KW(t) \quad (10)$$

where $K = [k_1 \ k_2 \ k_3]^T$ is a constant gain vector and $W(t) \in \mathbb{R}$ is the control input that satisfies in:

$$W(t) = \begin{cases} W^+(t), & S(e) \geq 0 \\ W^-(t), & S(e) < 0 \end{cases} \quad (11)$$

in which $S = S(e)$ is a switching surface that describes the desired dynamic. The resultant error is then written by:

$$D^q e = Ae + KW(t) \quad (12)$$

The sliding surface can be defined as:

$$S(e) = Ce = c_1 e_1 + c_2 e_2 + c_3 e_3 \quad (13)$$

An equivalent control is found when $\dot{S}(e) = 0$ which is a *necessary condition* for the state trajectory to stay on the switching surface $S(e) = 0$. Hence, the controlled system satisfies the following conditions in the steady state:

$$S(e) = 0 \quad \text{and} \quad \dot{S}(e) = 0 \quad (14)$$

Based on equation (12) to (14), It could be deduced:

$$\dot{S}(e) = \frac{\partial S(e)}{\partial e} \dot{e} = \frac{\partial S(e)}{\partial e} D^{1-q} (D^q e) = \quad (15)$$

$$CD^{1-q} [Ae + KW(t)] = 0$$

Hence,

$$D^{1-q} W(t) = -(CK)^{-1} CAD^{1-q} e(t) \quad (16)$$

The equivalent control W_{eq} is a solution of equation (16):

$$W_{eq}(t) = -(CK)^{-1} CAe(t) \quad (17)$$

which is realizable whenever CK takes non-zero value. Replacing $W(t)$ in equation (12) from $W_{eq}(t)$ in equation (17), the error dynamic will be determined by the following relation:

$$D^q e = (I - K(CK)^{-1}C)Ae \quad (18)$$

As a classic work, the constant plus proportional rate reaching law will be considered [32-37]. Accordingly the reaching law is obtained as:

$$D^q S = -p \operatorname{sgn}(S) - rS \quad (19)$$

Where $\operatorname{sgn}(\cdot)$ denotes the sign function. Gains $p > 0$ and $r > 0$ are determined such that the sliding condition is met and the sliding mode motion occurs. From equations (12) and (13), one may find that:

$$D^q S = CD^q e = C[Ae + KW(t)] \quad (20)$$

However, from equations (19) and (20), the control input will be provided by:

$$W(t) = -(CK)^{-1} [C(rI + A)e + p \operatorname{sgn}(S)] \quad (21)$$

According to Theorem 1, as long as all eigenvalues of $(A - K(CK)^{-1}C(rI + A))$ (λ_i 's $i = 1, 2, 3$) satisfy the condition $|\arg(\lambda_i)| > q\pi/2$, the system in (18) is asymptotically stable [25].

IV. ACTIVE SLIDING MODE SYNCHRONIZATION OF COULLET SYSTEM

In this section, the described method will be used to synchronize two identical fractional orders Couplet system with the following initial conditions: $(x_{10}, x_{20}, x_{30}) = (1, -1, 0)$

, $(y_{10}, y_{20}, y_{30}) = (0.2, 0.2, 0.2)$.

Consider two fractional order Couplet systems as *master* and *slave* systems respectively:

$$\text{master} \begin{cases} D^q x_1 = x_2 \\ D^q x_2 = x_3 \\ D^q x_3 = cx_3 + bx_2 + ax_1 + dx_1^3 \end{cases} \quad (22)$$

$$\text{slave} \begin{cases} D^q y_1 = y_2 + u_1 \\ D^q y_2 = y_3 + u_2 \\ D^q y_3 = cy_3 + by_2 + ay_1 + dy_1^3 + u_3 \end{cases} \quad (23)$$

Applying described method, matrix A , nonlinear part $g(x)$ and $h(y)$ are achieved accordingly:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & b & c \end{bmatrix}, \quad g(x) = \begin{bmatrix} 0 \\ 0 \\ dx_1^3 \end{bmatrix}, \quad h(y) = \begin{bmatrix} 0 \\ 0 \\ dy_1^3 \end{bmatrix} \quad (24)$$

The error is defined as the discrepancy of the relevant states i.e. $e_i = y_i - x_i$ for $i = 1, 2, 3$.

Deducing the master dynamic from the slave, leads to:

$$D^q e = Ae + h(y) - g(x) + U \quad (25)$$

Finally, a desired input control is calculated as (27):

$$B_1 = -(CK)_3^{-1} = (c_1k_1 + c_2k_2 + c_3k_3)^{-1} \quad (26)$$

$$B_2 = [C(rI + A)e + p \operatorname{sgn}(S)] =$$

$$(c_1r + c_3a)e_1 + (c_1 + c_2r + c_3b)e_2 +$$

$$(c_2 + c_3(c + r))e_3 + p \operatorname{sgn}(S)$$

$$\begin{cases} u_1(t) = -k_1B_1B_2 \\ u_2(t) = -k_2B_1B_2 \\ u_3(t) = -k_3B_1B_2 - dy_1^3 + dx_1^3 \end{cases} \quad (27)$$

Assume that orders of the master and the slave are $q=0.9$ and parameters of system as $(a, b, c) = (5, -10, -3.8)$. Parameters of the controller are chosen as $K = (0.1, 0.3, 0.4)$, $C = (10, 3, 4)$ and $p = 0.5$. This selection of parameters results in eigenvalues $(\lambda_1, \lambda_2, \lambda_3) = (-10, -0.5793 \pm 1.3602i)$ which located in a stable region ($|\arg(\lambda_i)| > 0.9 * \pi / 2$). Numerical simulations for this case have carried out based on ode45 solver. Fig.3 shows the effectiveness of the proposed controller to synchronize two fractional-order modeled systems. It should be noted that control $u(t)$, has been activated at $t = 0$.

V. SYNCHRONIZATION OF TWO DIFFERENT FRACTIONAL-ORDER COULET SYSTEM VIA ACTIVE SLIDING MODE CONTROL

In this section two fractional order master and slave with different fractional order will be synchronized via proposed method. This method can be useful in synchronizing two systems with partially different attractor, which may usually happen in coding and decoding applications.

Consider master and slave system as :

$$\text{master} \begin{cases} D^{q_1} x_1 = x_2 \\ D^{q_1} x_2 = x_3 \\ D^{q_1} x_3 = cx_3 + bx_2 + ax_1 + dx_1^3 \end{cases} \quad (28)$$

$$\text{slave} \begin{cases} D^{q_2} y_1 = y_2 + u_1 \\ D^{q_2} y_2 = y_3 + u_2 \\ D^{q_2} y_3 = cy_3 + by_2 + ay_1 + dy_1^3 + u_3 \end{cases} \quad (29)$$

the error dynamic can be written as:

$$\begin{cases} D^{q_1} e_1 = e_2 + D^{q_1} y_1 - D^{q_2} y_1 \\ D^{q_1} e_2 = e_3 + D^{q_1} y_2 - D^{q_2} y_2 \\ D^{q_1} e_3 = ce_3 + be_2 + ae_1 + dy_1^3 - dx_1^3 + D^{q_1} y_3 - D^{q_2} y_3 \end{cases} \quad (30)$$

Again matrix A , nonlinear part $g(x)$ and $h(y)$ are described accordingly as:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & b & c \end{bmatrix}, \quad g(x) = \begin{bmatrix} 0 \\ 0 \\ dx_1^3 \end{bmatrix}, \quad (31)$$

$$h(y) = \begin{bmatrix} D^{q_1} y_1 - D^{q_2} y_1 \\ D^{q_1} y_2 - D^{q_2} y_2 \\ dy_1^3 + D^{q_1} y_3 - D^{q_2} y_3 \end{bmatrix}$$

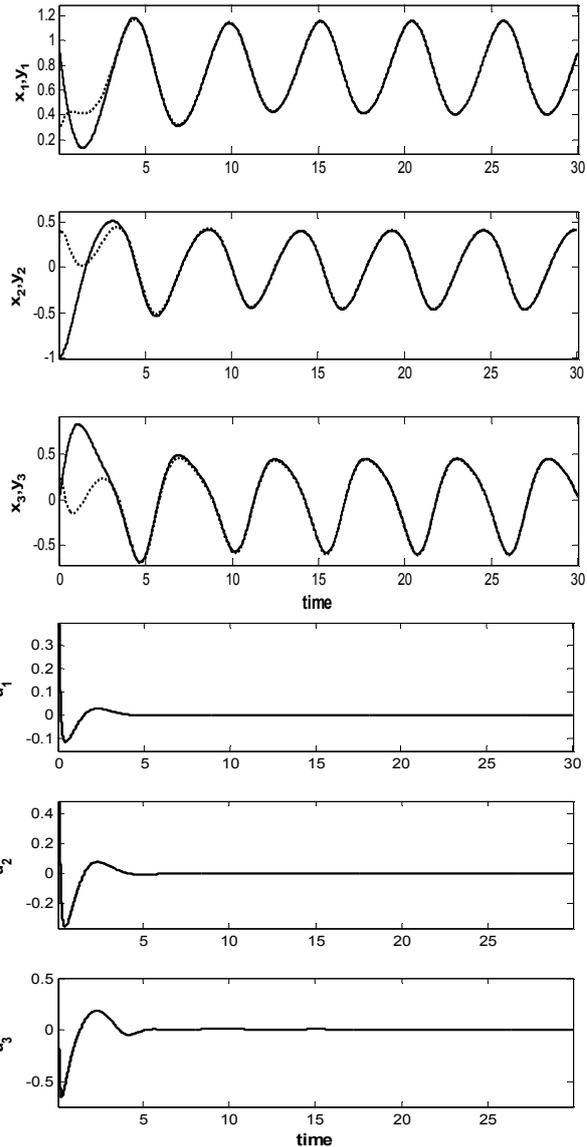


Fig.3 :a) Simulation Results of ASM synchronization of two identical fractional order Coulet systems for $q = 0.9$, $p = 0.5$, $k = [10, 0.3, 0.4]^T$ and $C = [10, 3, 4]$. b) Control signal for states x_1 , x_2 and x_3 .

The objective is to find suitable controllers u_i ($i = 1, 2, 3$) which made the error dynamic asymptotically stable.

Assuming B_1, B_2 the same as (26), the suitable controllers can be defined as:

$$\begin{cases} u_1(t) = -k_1 B_1 B_2 + D^{q_1} y_1 - D^{q_2} y_1 \\ u_2(t) = -k_2 B_1 B_2 + D^{q_1} y_2 - D^{q_2} y_2 \\ u_3(t) = -k_3 B_1 B_2 - d y_1^3 + d x_1^3 + D^{q_1} y_3 - D^{q_2} y_3 \end{cases} \quad (32)$$

Assume two different fractional order Coulet system with the initial conditions as master and slave. The fractional order of master is $q_1=0.9$ with initial condition as $(x_{10}, x_{20}, x_{30}) = (1, -1, 0)$ and for slave system are $q_2=0.88$ and $(y_{10}, y_{20}, y_{30}) = (0.2, 0.2, 0.2)$. Assume the parameters of system as $(a, b, c) = (5, -10, -3.8)$. Parameters of the controller are chosen as $K = (0.1, 0.3, 4)$, $C = (10, 4, 0.5)$ and $p = 0.5$. This selection of parameters results in eigenvalues $(\lambda_1, \lambda_2, \lambda_3) = (-10, -2.3649 \pm 2.1492i)$ which satisfy the stable condition ($|\arg(\lambda_i)| > q_1 * \pi / 2$).

In Fig.4 the results of proposed method and its effectiveness in synchronizing two different fractional-order Coulet systems are shown.

VI. CONCLUSION

In this paper, synchronization of two chaotic Coulet system with fractional orders models (as master and slave) with the same order, is investigated. Active sliding mode method has been developed to imply the task. It has been shown that by proper selection of the control parameters (K_i, c_i, p, r), the master and slave systems are synchronized. Furthermore all eigenvalues (λ_i 's $i = 1, 2, 3$) of synchronized system will satisfy the suffusion condition i.e $|\arg(\lambda_i)| > q\pi/2$. This means the error is stabilized so in the long term analysis, the synchronization would be guaranteed.

The proposed method has also been developed to synchronize two master and slave system with different fractional-order. Numerical simulations have shown the efficiency of the proposed controller in the mentioned task.

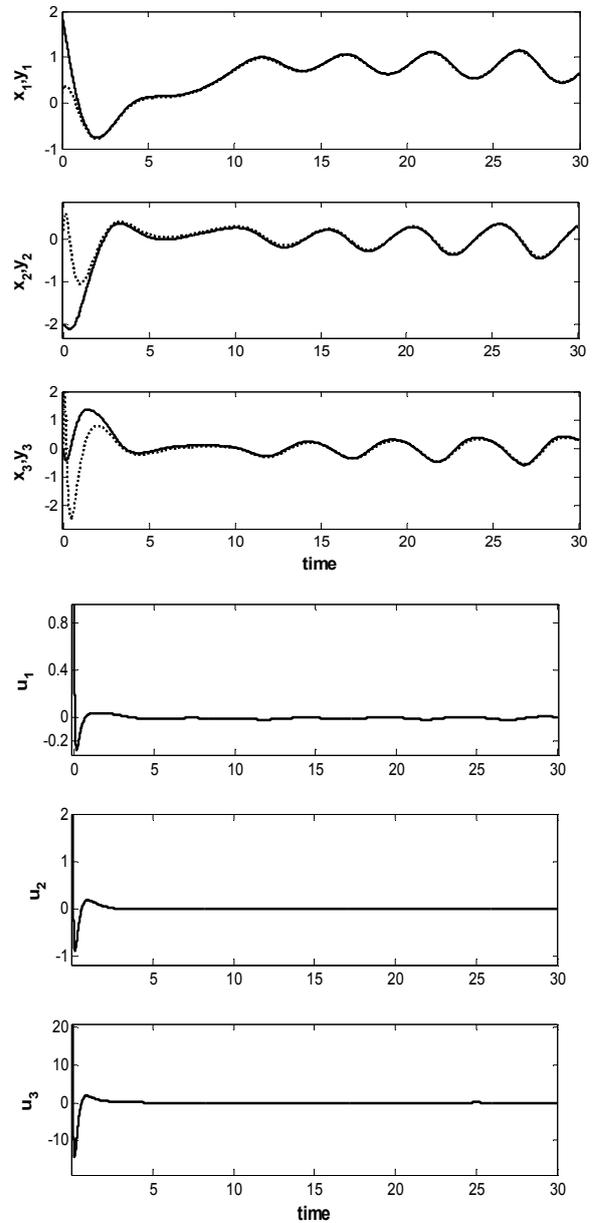


Fig.4 :a) Simulation Results of ASM synchronization of two fractional order Coulet systems with different orders. Parameters of the systems and controller are respectively chosen as $q_1=0.9$, $q_2=0.88$, $(a, b, c) = (5, -10, -3.8)$, $K = (0.1, 0.3, 4)$, $C = (10, 4, 0.5)$ and $p = 0.5$. b) Control signal for states x_1 , x_2 and x_3 .

REFERENCES

- [1] R. Hilfer, "Application of fractional calculus in physics" New Jersey: World Scientific, 2001.
- [2] RL. Bagley and RA. Calico. Fractional order state equations for the control of viscoelastically damped structures. Journal of Guidance Control and Dynamic, 14, 304-311, 1991.

- [3] G. Chen, X. Dong. On feedback control of chaotic continuous-time systems. *IEEE Transactions on Circuits and Systems I*, 40 (9) 591-601, 1993.
- [4] G. Chen, X. Dong. *From Chaos to Order: Methodologies, Perspectives and Applications*. World Scientific, Singapore, pp. 12–27, 1998.
- [5] C.C. Fuh and P.C. Tung. Controlling chaos using differential geometric method. *Physics Review Letter* 75 (16), 2952–2955, 1995.
- [6] X.Y. Wang. *Chaos in the Complex Nonlinearity System*, Electronics Industry Press, Beijing, pp. 91–113, 2003.
- [7] G.R. Wang, X.L. Yu, S.G. Chen. *Chaotic Control, Synchronization and Utilizing*. National Defence Industry Press, Beijing, pp. 281–455, 2001.
- [8] Ahmad W. and J.C. Sprott. Chaos in fractional order system autonomous nonlinear systems. *Chaos, Solitons & Fractals*;16, 339–351, 2003.
- [9] L.M. Pecora, T.L. Carroll. Synchronization in chaotic systems. *Physics Review Letter*, 64(8) (1990), 821-824.
- [10] T.L. Carroll, L.M. Pecora. Synchronization chaotic circuits, *IEEE Trans. Circuits. Syst.* 38 (4), 453-456, 1991.
- [11] M-C Ho, Y-C. Hung and C-H. Chou . Phase and anti-phase synchronization of two chaotic systems by using active control. *Physics Letter A*, 296 , 43–8, 2002.
- [12] G.R. Michael, S.P. Arkady and K. Jürgen. From phase to lag synchronization in coupled chaotic oscillators. *Physics Review Letter* 78, 4193–4196, 1997.
- [13] E.M. Shahverdiev, S. Sivaprakasam, K.A. Shore. Lag synchronization in time-delayed systems. *Physics Letter A*, 292, 320–324, 2002.
- [14] T. Shinbrot, E. Ott, C. Grebogi and JA Yorke. Using Small Perturbations to Control Chaos. *Nature*, 363(6), 411-7, 1993.
- [15] X.S. Yang. On the existence of generalized synchronizer in unidirectionally coupled systems. *Applied Mathematics and Computation*, 122(1), 71-79, 2001.
- [16] X. Yu and Y. Song. Chaos Synchronization via controlling partial state chaotic system. *International Journal of Bifurcation. Chaos* ,11 (6) .pp 1737,2001.
- [17] Hosseinnia, S. H., R. Ghaderi, A. Ranjbar N., J. Sadati, S. Momani, Synchronization of Fractional Chaotic Systems Via Fractional-Order Adaptive Controller, , in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [18] Hosseinnia, S. H., R. Ghaderi, A. Ranjbar N., J. Sadati, S. Momani, Designation of an adaptive PID controller to synchronize fractional Corder hard spring Phi⁶-Van der Pol oscillator, , in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [19] Hosseinnia, S. H., Reza Ghaderi, Abolfazl Ranjbar, N., M. Mahmoudian, Shaher Momani, Synchronization of Uncertain Fractional-Order Duffing-Holmes Chaotic System Via Sliding Mode Control, in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [20] M. Shahiri T., A. Ranjbar N., R. Ghaderi, S. H. Hosseinnia, S. Momani, Control and Synchronization of Chaotic Fractional-Order Couillet System Via Active Controller, in proc. : 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [21] H. Delavari, A. Ranjbar N., R. Ghaderi, S. Momani, Fractional Sliding Mode Control for Synchronization of Chaotic Systems , 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [22] S.H. Hosseinnia*, R. Ghaderi*, A. Ranjbar N.*, F. Abdous*, S. Momani Fractional Order Controller in Combination with State Feedback Controller to Stabilize Chaotic Systems, 3rd IFAC Workshop on Fractional Differentiation and its Applications, Ankara, Turkey, 2008.
- [23] S.J. Sadati*, A. Ranjbar N.*, S.H. Hosseinnia*, R. Ghaderi*, S. Momani , Bifurcation Analysis, Parameter Identification and Synchronization of Fractional-Order Newton-Leipnik System using Adaptive Control , International workshop on new trends in science and technology, Ankara, Turkey, 2008.
- [24] M. Shahiri* T. , A. Ranjbar N.*, S.H. Hosseinnia*, R. Ghaderi*, S. Momani, Projective Synchronization of Oscillation and Chaos in Fractional-Order Anti-Controlled Rigid Body System , International workshop on new trends in science and technology, Ankara, Turkey, 2008.
- [25] Mohammad Saleh Tavazoei, Mohammad Haeri ,Synchronization of chaotic fractional-order systems via active sliding mode controller, *Physica A*: 387, 57–70,2008.
- [26] A. Arnéodo, P. Couillet and C. Tresser. Possible new strange attractors with spiral structure. *Communication in Mathematical Physics*. 79(4), 573-579,1981.
- [27] P. Couillet, C. Tresser and A. Arnéodo. Transition to stochasticity for a class of forced oscillators. *Physics. Letter A*, 72, 268,1979.
- [28] M. S.Tavazoei, M. Haeri. Chaos control via a simple fractional-order controller, *Physics Letters A*, 372 (6), 798-80, 2008.
- [29] E.W. Bai, K.E. Lonngren, Synchronization of two Lorenz systems using active control, *Chaos, Solitons Fractals* 8 ,51–58, 1997.
- [30] H.N. Agiza, M.T. Yassen, Synchronization of Rossler and Chen chaotic dynamical systems using active control, *Phys. Lett. A*, 278, 191–197, (2001).
- [31] M.Ch. Ho, Y.Ch. Hung, Synchronization of two different systems by using generalized active control, *Phys. Lett. A* 301 ,424–428, (2002).
- [32] H. Zhang, X.K. Ma, W.Z. Liu, Synchronization of chaotic systems with parametric uncertainty using active sliding mode control,*Chaos, Solitons Fractals* 21 , 1249–1257, (2004).
- [33] H.T. Yau, Design of adaptive sliding mode controller for chaos synchronization with uncertainties, *Chaos, Solitons Fractals* 22 ,341–347, (2004).
- [34] J.S.H. Lin, J.J. Yan, T.L. Liao, Chaotic synchronization via adaptive sliding mode observers subject to input nonlinearity, *Chaos, Solitons Fractals* 24 (2005) 371–381.
- [35] M. Haeri, A.A. Emadzadeh, Synchronizing different chaotic systems using active sliding mode control, *Chaos, Solitons Fractals* 31 (1) ,119–129, 2007.
- [36] M.S. Tavazoei, M. Haeri, Determination of active sliding mode controller parameters in synchronizing different chaotic systems, *Chaos, Solitons Fractals* 32 ,583–591, 2007.
- [37] M. Haeri, M.S. Tavazoei, M.R. Naseh, Synchronization of uncertain chaotic systems using active sliding mode control, *Chaos, Solitons Fractals* 33 , 1230–1239, 2007.

Dynamical behaviors, linear feedback control and synchronization of the fractional order Liu system

Ahmed E. Matouk^{a,b}

Abstract— In this paper, some dynamical behaviors of the fractional order Liu system are investigated. It is found that chaos exists in this system with order less than 3. Chaos control and synchronization are achieved by using linear control technique. Simulation results are used to visualize and illustrate the effectiveness of the proposed control and synchronization methods.

Index Terms—fractional order Liu system, stability conditions, chaos, chaos control, synchronization, linear control technique

I. INTRODUCTION

Although fractional calculus is such an old topic that it was introduced in the early 17th century, it has been extensively studied in the last decade by scientists, engineers and physicists [1-2].

There are many definitions of fractional order derivatives. The Riemann-Liouville definition [3] which is given by

$$D^\alpha f(t) = \frac{d^l}{dt^l} J^{l-\alpha} f(t), \quad (1)$$

where J^θ is the θ -order Riemann-Liouville integral operator which is given as

$$J^\theta u(t) = \frac{1}{\Gamma(\theta)} \int_0^t (t-\tau)^{\theta-1} u(\tau) d\tau, \quad \theta > 0. \quad (2)$$

Another one is the Caputo definition of fractional derivatives [4], which is often used in real applications:

$$D_*^\alpha f(t) = J^{l-\alpha} f^{(l)}(t), \quad \alpha > 0, \quad (3)$$

where $f^{(l)}$ represents the l -order derivative of $f(t)$ and $l = [\alpha]$, this means that l is the first integer which is not less

^a Mathematics Department, Faculty of Science, Mansoura University, Mansoura, 35516, Egypt.

^b Mathematics Department, Faculty of science, Hail University, Hail, 2440, Saudi Arabia.
E.mail: amatouk@hotmail.com

than α . The operator D_*^α is called the “Caputo differential operator of order α ”. In [5], F. Ben Adda studied the geometric and physical interpretation of the fractional derivative.

On the other hand, studying chaos in fractional order dynamical systems is very interesting topic and has much increasing attention in the past few years. Chaotic attractors have been found in the following fractional order systems, Lorenz [6], Chua [7], Chen [8] and Lü [9]. Chaos in fractional order autonomous systems has also interesting phenomena that is it can occur for orders less than three and this can't happen in their integer order counterparts according to Poincaré-Bendixon theorem. Chaos control and synchronization in integer order differential systems are well understood [10-11], but they are still in the beginning in the case of fractional order chaotic systems. Recently, many papers about chaos control and synchronization in fractional order chaotic systems have been published by authors [12-13].

In this work, I study stability, chaos, control and synchronization in the Liu system [14] with same fractional order. I use the Routh-Hurwitz conditions given in [15] to study the stability conditions in this system. It is found that the lowest order for chaos to exist in such system is 2.55. The sufficient conditions for chaos control are derived analytically using linear feedback control technique. Conditions for achieving chaos synchronization via linear control method are studied using the classical Laplace transformation theory. The analytical results are verified by numerical simulations.

II. THE FRACTIONAL ORDER LIU SYSTEM

The fractional order Liu system with same fractional order is given as follows

$$D_*^\alpha x = a(y-x), \quad D_*^\alpha y = bx - kxz, \quad D_*^\alpha z = -cz + hx^2, \quad (4)$$

where α is the fractional order and $\alpha \in (0, 1]$. The parameters a, c, k, h are all positive real parameters and $b \in \mathbb{R}$. When $\alpha = 1$ system (4) is the original integer order Liu system which exhibits chaotic behaviors at the parameter values $a = 10, b = 40, c = 2.5, h = 4$ and $k = 1$.

The equilibrium points of system (4) are $E_0 = (0, 0, 0)$,

$$E_1 = \left(\sqrt{\frac{bc}{hk}}, \sqrt{\frac{bc}{hk}}, \frac{b}{k} \right) \text{ and } E_2 = \left(-\sqrt{\frac{bc}{hk}}, -\sqrt{\frac{bc}{hk}}, \frac{b}{k} \right).$$

III. CONDITIONS FOR STABILITY IN 3-DIMENSIONAL FRACTIONAL ORDER SYSTEMS

The stability conditions of fractional order systems given in [15] are shown in the following:

Consider the 3-dimensional fractional order system

$$D^\alpha x(t) = f(x, y, z), \quad D^\alpha y(t) = g(x, y, z), \quad D^\alpha z(t) = h(x, y, z), \quad (5)$$

where $\alpha \in (0, 1]$ and $(\bar{x}, \bar{y}, \bar{z})$ is an equilibrium solution of (5). The eigenvalues equation of the equilibrium solution $(\bar{x}, \bar{y}, \bar{z})$ is given as

$$P(\lambda) = \lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0, \quad (6)$$

whose discriminant $D(P)$ is given by

$$D(P) = 18a_1a_2a_3 + (a_1a_2)^2 - 4a_3(a_1)^3 - 4(a_2)^3 - 27(a_3)^2. \quad (7)$$

Then $(\bar{x}, \bar{y}, \bar{z})$ is locally asymptotically stable if all the roots of equation (6) satisfy the condition $|\arg(\lambda)| > \alpha\pi/2$.

Consequently, we have the following stability conditions:

- (i) If $D(P) > 0$, then the necessary and sufficient conditions for the equilibrium point $(\bar{x}, \bar{y}, \bar{z})$ to be locally asymptotically stable is

$$a_1 > 0, \quad a_3 > 0, \quad a_1a_2 - a_3 > 0.$$

- (ii) If $D(P) < 0$, $a_1 \geq 0$, $a_2 \geq 0$, $a_3 > 0$ then $(\bar{x}, \bar{y}, \bar{z})$ is locally asymptotically stable for $\alpha < 2/3$. However, if $D(P) < 0$, $a_1 < 0$, $a_2 < 0$, $\alpha > 2/3$ then all roots of equation (6) satisfy the condition $|\arg(\lambda)| < \alpha\pi/2$.
- (iii) If $D(P) < 0$, $a_1 > 0$, $a_2 > 0$, $a_1a_2 - a_3 = 0$ then $(\bar{x}, \bar{y}, \bar{z})$ is locally asymptotically stable for all $\alpha \in (0, 1)$.
- (iv) $a_3 > 0$, is the necessary condition for the equilibrium point $(\bar{x}, \bar{y}, \bar{z})$ to be locally asymptotically stable.

A. Stability conditions of the equilibrium point E_0

The eigenvalues equation of the equilibrium point E_0 is given by

$$\lambda^3 + (a+c)\lambda^2 + (ac-ab)\lambda - abc = 0. \quad (8)$$

When $b < 0$ and $D(P) > 0$ then using the stability condition (i), the equilibrium point E_0 is locally asymptotically stable for all $\alpha \in (0, 1]$. However if $D(P) < 0$ then E_0 is locally asymptotically stable for $\alpha < 2/3$ or if $D(P) < 0$ and $b = \frac{c(c+a)}{a}$ then E_0 is locally asymptotically

stable for all $\alpha \in (0, 1)$ (using the stability conditions (ii) and (iii) respectively).

When $b > 0$ then $a_3 < 0$, and by using the condition (iv) we can easily see that E_0 is unstable.

B. Stability conditions of the equilibrium points E_1 and E_2

When $b > 0$, two other equilibrium points E_1 and E_2 appear, and they have the following eigenvalues equation:

$$\lambda^3 + (a+c)\lambda^2 + ac\lambda + 2abc = 0. \quad (9)$$

If $D(P) > 0$ then the necessary and sufficient condition for the equilibrium points E_1 (E_2) to be locally asymptotically stable for all $\alpha \in (0, 1]$ is $b < \frac{a+c}{2}$ (from the stability condition

- (i)). However, if $D(P) < 0$ then E_1 (E_2) are locally asymptotically stable for $\alpha < 2/3$ or if $D(P) < 0$ and $b = \frac{(a+c)}{2}$ then E_1 (E_2) are locally asymptotically stable for all $\alpha \in (0, 1)$ (using the stability conditions (ii) and (iii) respectively). Now, the following Lemma is proved:

Lemma 1 For $\alpha < 2/3$, system (4) undergoes a pitchfork bifurcation at $b = 0$. Moreover, when $b < 0$ the unique equilibrium point E_0 is locally asymptotically stable, and when $b > 0$, E_0 becomes unstable and two other equilibrium points E_1 and E_2 appear and they are locally asymptotically stable (near $b = 0$).

IV. CHAOS IN FRACTIONAL ORDER LIU SYSTEM

An efficient method for solving fractional order differential equations is the predictor-correctors scheme or more precisely, **PECE** (Predict, Evaluate, Correct, Evaluate) technique which has been investigated in [16], and represents a generalization of the Adams-Bashforth-Moulton algorithm. It is used throughout this paper.

For the parameter values $a = 10$, $b = 40$, $c = 2.5$, $h = 4$ and $k = 1$, it follows that $D(P) < 0$ and therefore the equilibrium points E_1 (E_2) are locally asymptotically stable for $\alpha < 2/3$. For this choice of the parameter values, the integer order form of system (4) exhibits chaotic behavior and has the three equilibria $E_0 = (0, 0, 0)$, $E_1 = (5, 5, 40)$ and $E_2 = (-5, -5, 40)$. The eigenvalues of these equilibrium points are given as follows:

$$\text{For } E_0: \quad \lambda_1 = -25.6155, \quad \lambda_2 = 15.6155, \quad \lambda_3 = -2.5000.$$

$$\text{For } E_1: \quad \lambda_1 = -17.5614, \quad \lambda_{2,3} = 2.5307 \pm 10.3673I.$$

$$\text{For } E_2: \quad \lambda_1 = -17.5614, \quad \lambda_{2,3} = 2.5307 \pm 10.3673I, \quad \text{where } I = \sqrt{-1}.$$

Now, according to [17] the equilibrium point $E_0 = (0,0,0)$ is saddle point of index 1, however the other equilibrium points E_1 and E_2 are saddle points of index 2. Thus, the necessary condition for the fractional order Liu system (4) to remain chaotic is $\alpha > \frac{2}{\pi} \arctan\left(\frac{|\text{Im}(\lambda_{2,3})|}{\text{Re}(\lambda_{2,3})}\right)$. Consequently, the maximum fractional order α for which the fractional order Liu system (4) demonstrates chaos for the above given parameters is $\alpha \approx 0.85$. Since the order of the fractional order chaotic system is the sum of the orders of all involved derivatives, hence, we show that the lowest order for the fractional order Liu system given by equations (4) to show chaos is 2.55.

Simulations are performed for $\alpha = 0.9$ using the above-mentioned parameter values at which system (4) shows Lorenz-like attractor. When $\alpha = 0.85$, system (4) is still chaotic. However below the value $\alpha = 0.85$, system (4) becomes non chaotic.

V. CHAOS CONTROL OF FRACTIONAL ORDER LIU SYSTEM

A 3-dimensional fractional order chaotic system is described as

$$D_*^\alpha X = F(X), \quad (10)$$

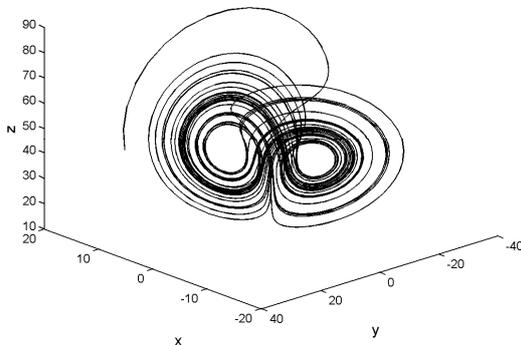


Fig. 1. 3-D plot of the fractional order Liu attractor in the x-y-z space using fractional order $\alpha = 0.9$.

where $X \in R^3$. The controlled system is given as

$$D_*^\alpha X = F(X) - K(X - X_e), \quad (11)$$

where $K = \text{diag}(k_1, k_2, k_3)$, $k_1, k_2, k_3 \geq 0$ and $X_e = (x_e, y_e, z_e)$ is an equilibrium point of (10). Now, by selecting the appropriate feedback control gains k_1, k_2, k_3 such that the eigenvalues of the linearized equation of the controlled system (11) satisfy one of the above-mentioned Routh-Hurwitz conditions, then the trajectories of the controlled system (11) asymptotically approaches the unstable equilibrium point X_e

in the sense that $\lim_{t \rightarrow \infty} \|X - X_e\| = 0$, where $\|\cdot\|$ is the Euclidean norm.

Now, consider the controlled fractional order Liu system which is given by

$$\begin{aligned} D_*^\alpha x &= a(y-x) - k_1(x-x_e), D_*^\alpha y = bx - kx - k_2(y-y_e), \\ D_*^\alpha z &= -cz + hx^2 - k_3(z-z_e), \end{aligned} \quad (12)$$

where k_1, k_2 and k_3 are all positive feedback gains. By a suitable choice of these feedback gains according to the stability conditions mentioned above, one can drive the system's trajectory to any of the three unstable equilibrium points E_0, E_1 and E_2 .

A. Stabilizing the equilibrium point E_0

The eigenvalues equation of the controlled system (12) at $E_0 = (0,0,0)$ is given as

$$\begin{aligned} \lambda^3 + (s_1 + s_2 + k_2)\lambda^2 + (s_1s_2 + s_1k_2 + s_2k_2 - ab)\lambda \\ + s_1(s_2k_2 - ab) = 0, \end{aligned} \quad (13)$$

where $s_1 = c + k_3 > 0$ and $s_2 = a + k_1 > 0$.

By choosing the feedback gains k_1, k_2 and k_3 such that $D(P) > 0$, then using the stability condition (i), we find that the necessary and sufficient condition for the equilibrium solution $E_0 = (0,0,0)$ of the controlled system (12) to be locally asymptotically stable is $k_2 > \frac{ab}{a+k_1}$.

B. Stabilizing the equilibrium points E_1 and E_2

The eigenvalues equation of the controlled system (12) at the equilibria $E_1(E_2)$ is given by

$$\begin{aligned} \lambda^3 + (s_1 + s_2 + k_2)\lambda^2 + (s_1s_2 + s_1k_2 + s_2k_2)\lambda \\ + s_1s_2k_2 + 2abc = 0. \end{aligned} \quad (14)$$

By choosing the feedback gains k_1, k_2 and k_3 such that $D(P) > 0$, the stability condition (i) ensures that the equilibrium solutions $E_1(E_2)$ of the controlled system (13) are locally asymptotically stable if and only if

$$k_2 > -\frac{g}{2} + \sqrt{\frac{(k_1 - k_3 + a - c)^2}{4} + \frac{2abc}{g}}, \quad (15)$$

where $g = k_1 + k_3 + a + c$.

C. Numerical results

Simulation procedures are coded and executed using PECE method. The system (12) is then numerically integrated with parameter values; $a = 10$, $b = 40$, $c = 2.5$, $h = 4$ and $k = 1$. The feedback control gains $(k_1, k_2, k_3) = (1, 40, 1)$ satisfy the conditions of the stability of the equilibrium point E_0 . The simulation results show that this controller stabilize the fractional order Liu system to this equilibrium point with order $\alpha = 0.9$ (see figure 2). However, the feedback gains $(k_1, k_2, k_3) = (2, 120, 1)$ satisfy the stability conditions of the equilibrium points $E_1 = (5, 5, 40)$ and $E_2 = (-5, -5, 40)$. Figures 3a and 3b show the trajectories of the controlled fractional order Liu system with order $\alpha = 0.9$ converge to the equilibrium points E_1 and E_2 respectively.

VI. SYNCHRONIZATION VIA LINEAR CONTROL METHOD

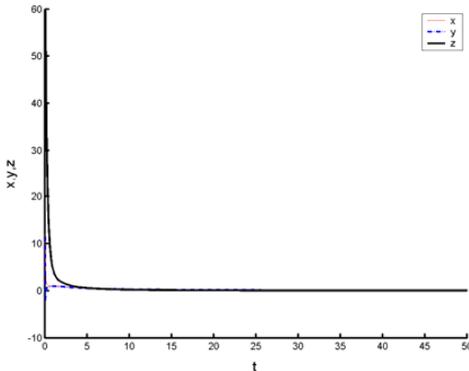


Fig. 2. Shows the trajectories of the controlled system (12) are stabilized to the equilibrium point $E_0 = (0,0,0)$ after the controllers $k_1 = 1$, $k_2 = 40$, $k_3 = 1$ are activated.

In this section, our goal is to study chaos synchronization in

the fractional order Liu system (4) by applying linear control technique. Let the drive and response systems be given as follows:

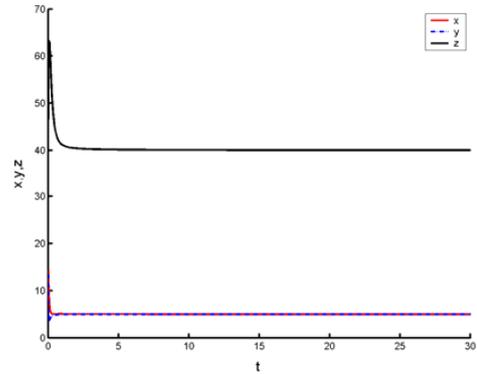
$$D_*^\alpha x_1 = a(y_1 - x_1), D_*^\alpha y_1 = bx_1 - kx_1z_1, D_*^\alpha z_1 = -cz_1 + hx_1^2, \quad (16)$$

and,

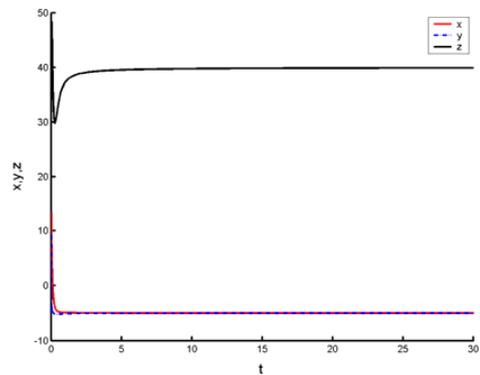
$$D_*^\alpha x_2 = a(y_2 - x_2), D_*^\alpha y_2 = bx_2 - kx_2z_2 - \rho(y_2 - y_1), \quad (17)$$

$$D_*^\alpha z_2 = -cz_2 + hx_2^2,$$

where ρ is feedback control gain and $\rho > 0$. Assume that



(a)



(b)

Fig. 3. The trajectories of the controlled system (12) are stabilized to the equilibrium point (a) $E_1 = (5, 5, 40)$, (b) $E_2 = (-5, -5, 40)$ after the controllers $k_1 = 2$, $k_2 = 120$, $k_3 = 1$ are activated.

$e_1 = x_2 - x_1$, $e_2 = y_2 - y_1$, $e_3 = z_2 - z_1$. By subtracting (16) from (17) we obtain

$$D_*^\alpha e_1 = a(e_2 - e_1),$$

$$D_*^\alpha e_2 = (b - \rho)e_2 - kz_1e_1 - kx_1e_3 - ke_1e_3, \quad (18)$$

$$D_*^\alpha e_3 = -ce_3 + h(x_2 + x_1)e_1.$$

By taking the Laplace transform in both sides of (18), letting $E_i(s) = L\{e_i(t)\}$ where $(i = 1, 2, 3)$, we obtain

$$s^\alpha E_1(s) - s^{\alpha-1}e_1(0) = a(E_2(s) - E_1(s)),$$

$$s^\alpha E_2(s) - s^{\alpha-1}e_2(0) = (b - \rho)E_2(s) - L\{kz_1e_1\} - L\{kx_1e_3\} - kE_1(s)E_3(s),$$

$$s^\alpha E_3(s) - s^{\alpha-1}e_3(0) = -cE_3(s) + L\{hx_2e_1\} + L\{hx_1e_1\}. \quad (19)$$

Theorem 1 If $b \neq \rho$, $E_1(s) \leq \xi$ and $E_2(s) \leq \xi$, the drive and response systems will be synchronized under suitable choice of the feedback control gain ρ .

Proof. By rewriting equation (19) as follows

$$\begin{aligned}
E_1(s) &= \frac{aE_2(s)}{s^\alpha + a} + \frac{s^{\alpha-1}e_1(0)}{s^\alpha + a}, \\
E_2(s) &= -\frac{kL\{z_1e_1\}}{s^\alpha - b + \rho} - \frac{kL\{x_1e_3\}}{s^\alpha - b + \rho} - \frac{kE_1(s)E_3(s)}{s^\alpha - b + \rho} + \frac{s^{\alpha-1}e_2(0)}{s^\alpha - b + \rho}, \\
E_3(s) &= \frac{hL\{x_2e_1\}}{s^\alpha + c} + \frac{hL\{x_1e_1\}}{s^\alpha + c} + \frac{s^{\alpha-1}e_3(0)}{s^\alpha + c}.
\end{aligned} \tag{20}$$

Using the Final-value theorem of the Laplace transformation, it follows that

$$\begin{aligned}
\lim_{t \rightarrow \infty} e_1(t) &= \lim_{s \rightarrow 0^+} sE_1(s) = \lim_{s \rightarrow 0^+} sE_2(s) = \lim_{t \rightarrow \infty} e_2(t), \\
\lim_{t \rightarrow \infty} e_2(t) &= \lim_{s \rightarrow 0^+} sE_2(s) = \frac{k}{b - \rho} \lim_{s \rightarrow 0^+} sL\{x_1e_3\} \\
&\quad + \frac{k}{b - \rho} \lim_{s \rightarrow 0^+} sL\{z_1e_1\} + \frac{k}{b - \rho} \lim_{t \rightarrow \infty} e_1(t) \cdot \lim_{t \rightarrow \infty} e_3(t), \\
\lim_{t \rightarrow \infty} e_3(t) &= \lim_{s \rightarrow 0^+} sE_3(s) = \frac{h}{c} \lim_{s \rightarrow 0^+} sL\{x_2e_1\} + \frac{h}{c} \lim_{s \rightarrow 0^+} sL\{x_1e_1\}.
\end{aligned} \tag{21}$$

If $E_1(s)$, $E_2(s)$ are bounded and $b \neq \rho$, then $\lim_{t \rightarrow \infty} e_1(t) = \lim_{t \rightarrow \infty} e_2(t) = 0$. Now, owing to the attractiveness of the attractor, there exists $\eta > 0$ such that $|x_i(t)| \leq \eta < \infty$, $|y_i(t)| \leq \eta < \infty$ and $|z_i(t)| \leq \eta < \infty$ where $(i = 1, 2)$.

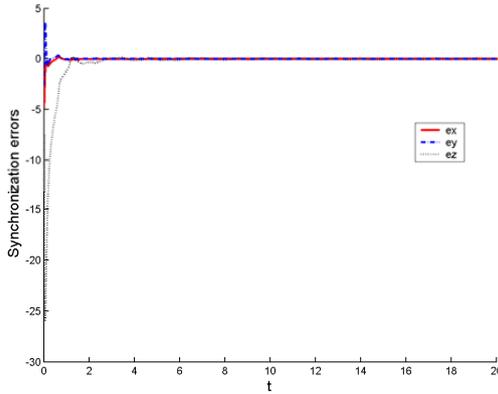


Fig. 4. Synchronization errors of the drive and response systems (16) and (17) with order $\alpha = 0.9$ and using the feedback gain $\rho = 100$.

Therefore, $\lim_{t \rightarrow \infty} e_3(t) = 0$. Finally, we get

$$\lim_{t \rightarrow \infty} e_j(t) = 0, \quad j = 1, 2, 3. \tag{22}$$

Consequently, the synchronization between the drive and response systems is achieved.

A. Numerical results

The drive and response systems (16) and (17) are integrated numerically using the PECE scheme with the initial values $x_1(0) = 15$, $y_1(0) = 20$, $z_1(0) = 29$, $x_2(0) = 10$, $y_2(0) = 15$,

$z_2(0) = 25$ and fractional order $\alpha = 0.9$. From figure 4 it is clear that the drive and response systems are synchronized when using $\rho = 100$.

VII. CONCLUSION

Some dynamical behaviors of the fractional order Liu system have been analyzed. Some stability conditions have been used to study the local stability of the equilibria. Chaos has been shown to be existed for this system with order less than 3. Analytical conditions for feedback control of the fractional Liu system have been derived. Chaos synchronization has also been achieved analytically and numerically using linear control method.

ACKNOWLEDGMENT

The author wishes to thank Professors. E. Ahmed and A.S. Hegazi for discussion and help.

REFERENCES

- [1] P.L. Butzer and U. Westphal, "An introduction to fractional calculus," World Scientific, Singapore, 2000.
- [2] S.M. Kenneth and R. Bertram, "An introduction to the fractional calculus and fractional differential equations," Wiley-Interscience, New York, 1993.
- [3] I. Podlubny, "Fractional differential equations," Academic Press, New York, 1999.
- [4] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," Geophys. J. R. Astron. Soc. 13 (1967) 529-539.
- [5] F. Ben Adda, "Geometric interpretation of the fractional derivative," J. Fract. Calc. 11 (1997) 21-52.
- [6] I. Grigorenko and E. Grigorenko, "Chaotic dynamics of the fractional Lorenz system," Phys. Rev. Lett. 91 (2003) 034101.
- [7] T.T. Hartley, C.F. Lorenzo and H.K. Qammer, "Chaos in a fractional order Chua's system," IEEE Trans. CAS-I 42 (1995) 485-490.
- [8] C. Li and G. Peng, "Chaos in Chen's system with a fractional order," Chaos, Solitons & Fractals 22 (2004) 443-450.
- [9] J.G. Lu, "Chaotic dynamics of the fractional order Lü system and its synchronization," Physics Lett. A 354 (2006) 305-311.
- [10] M. Lakshmanan and K. Murali, "Chaos in nonlinear oscillators: controlling and synchronization," World Scientific, Singapore, 1996.
- [11] A.E. Matouk and H.N. Agiza, "Bifurcations, chaos and synchronization in ADVP circuit with parallel resistor," J. Math. Analysis and Applications 341 (2008) 259-269.
- [12] W.M. Ahmad and A.M. Harb, "On nonlinear control design for autonomous chaotic systems of integer and fractional orders," Chaos, Solitons & Fractals 18 (2003) 693-701.
- [13] T. Zhou and C. Li, "Synchronization in fractional order differential systems," Physica D 212 (2005) 111-125.
- [14] C. Liu, T. Liu, L. Liu and K. Liu, "A new chaotic attractor," Chaos, Solitons & Fractals 22 (2004) 1031-1038.
- [15] E. Ahmed, A.M.A. El-Sayed and H.A.A. El-Saka, "On some Routh-Hurwitz conditions for fractional order differential equations and their applications in Lorenz, Rössler, Chua and Chen systems," Physics Lett. A 358 (2006) 1-4.
- [16] K. Diethelm, N.J. Ford and A.D. Freed, "A predictor-corrector approach for the numerical solution of fractional differential equations," Nonlinear Dynam. 29 (2002) 3-22.
- [17] M.S. Tavazoei and M. Haeri, "A necessary condition for double scroll attractor existence in fractional-order systems," Physics Lett. A 367 (2007) 102-113.

Chaotic Synchronization of Fractional-Order Chua's System with Time-Varying Delays

Shangbo Zhou, Xiaoran Lin, and Hua Li

Abstract—Chaos in fractional-order Chua's system with time-varying delays is illustrated by presenting its waveform graphs, states portraits and bifurcation graphs. Chaotic synchronization system for such system is constructed. the bifurcation graph with respect to the linear coupled parameter k is presented, and the numerical experiments are presented. The study shows that the chaos in such fractional-order Chua's system with time-varying delays can be synchronized. Furthermore, several different coupled systems are constructed, and the synchronization characteristics are compared.

Index Terms—Chua's system, fractional-order, varying time delays, chaotic synchronization.

I. INTRODUCTION

THERE is an increase in the number of applications where fractional calculus has been used. The real objects are generally fractional. In the past three decades, more interest has been devoted to fractional-order circuits and systems (FO-CAS), fractional-order signal processing (FO-SP) and fractional-order control systems (FO-CS) fields, among which complex behaviors such as bifurcations and chaos in electric and electronic circuits, networks, as well as nonlinear dynamic systems have attracted more attention in research [1-8].

Chaos, as a nonlinear phenomenon, has been widely researched and reports in both of theoretical and practical investigations, have appeared a huge amount in number [9-17]. Recently, Chaos in fractional-order system also attracts many researchers to investigate. Based on Chua's System, Tom T.

Manuscript received April 13, 2009. This work was supported in part by National Natural Science Foundation of China under Grant No.60873200, and Natural Science Foundation of CQ CSTC under Grant NO.2007BB2161.

Shangbo Zhou is with the Department of Computer Science and Engineering, Chongqing University, Chongqing, P.R. China 400044 (shbzhou@cqu.edu.cn)

Xiaoran Lin is with the Department of Computer Science and Engineering, Chongqing University, Chongqing, P.R. China, 400044 (lxl1236@126.com)

Hua Li is with the Department of Mathematics and Computer Science University of LethbridgeCanada T1K 3M4 (huali@cs.uleth.ca)

Harthey(1995) et al introduced the fractional-order Chua's system, studied the effect of fractional derivatives on the dynamics system. The bifurcation diagram of fractional-order Chua's system also was shown [9]. Wajdi M. Ahmad and J.C. Sprott(2003) discussed the chaotic behavior in a fractional jerk model, which is used to determine the time derivative of acceleration of an object [5]. In ref.[18], Li(2004) discussed the chaos in Chen's system and its control.

Since the pioneering works by Pecora and Carroll, various effective methods for chaos synchronization have been reported, especially chaotic synchronization of Chua and Chen circuit system [19-31]. In ref. [30], C. Cruz-Hernández and N. Romero-Haros(2008) applied the Generalized Hamiltonian forms and observer approach to synchronize time-delay-feedback Chua's circuits to transmit encrypted confidential information, and had enhanced the level of encryption security. Researching results shown that chaos have promising applications in secure communication. In ref. [31], a stochastic extended fractional Kalman filter is used for state reconstruction in a noisy environment. The chaotic communication scheme proposed by Arman Kiani-B et al is totally different from the traditional cryptosystems, due to employing different chaos states for synchronization and encryption. Chen proposed a new synchronization called "generalized projective synchronization" [26].

A time-delayed Chua's system can be described as following [30]:

$$\begin{aligned}\frac{dx}{dt} &= \alpha(y - x - f(x)), \\ \frac{dy}{dt} &= x - y + z, \\ \frac{dz}{dt} &= -\beta y - \gamma z - \beta \varepsilon \sin(\sigma x(t - \tau)),\end{aligned}\tag{1}$$

where $\alpha, \beta, \delta > 0$, and

$$f(x) = bx + \frac{1}{2}(a-b)(|x+1| - |x-1|), \quad a, b < 0.$$

In this paper, the time delay is considered being varying and the fractional-order Chua's system we will investigate is given as follows:

$$\begin{aligned} \frac{d^{\delta_1} x}{dt^{\delta_1}} &= \alpha(y - x - f(x)), \\ \frac{d^{\delta_2} y}{dt^{\delta_2}} &= x - y + z, \\ \frac{d^{\delta_3} z}{dt^{\delta_3}} &= -\beta y - \gamma z - \beta \varepsilon \sin(\sigma x(t - \tau(t))), \end{aligned} \quad (2)$$

where $\tau(t) > 0$ is delays with time varying. The chaotic phenomena and the chaotic synchronization of Eq. (1) will be studied in this paper.

II. CHAOTIC PHENOMIUN

Now we discuss chaotic phenomena in a fractional-order Chua's system with varying time delay. The fractional-order Chua's system is given as Eq. (2). If $\varepsilon \sigma \neq 0$, then system (1) is an explicit time-delayed system. We have illustrated that the time delay have affected the behavior of a dynamics system significantly [24].

By the definition of fractional-order derivative :

$${}_a D_t^\alpha f(t) = \lim_{h \rightarrow 0} h^{-\alpha} \sum_{j=0}^{\lfloor t-a/h \rfloor} (-1)^j \binom{\alpha}{j} f(t - jh), \quad (3)$$

we take calculus step h such that τ/h is an integer number. Then the discrete form of system (2) can be written as follows:

$$\begin{aligned} h^{-\delta_1} \sum_{j=0}^m \omega_j^{(\delta_1)} x_{m-j} &= \alpha(y_m - x_m - f(x_m)), \\ h^{-\delta_2} \sum_{j=0}^m \omega_j^{(\delta_2)} y_{m-j} &= x_m - y_m + z_m, \\ h^{-\delta_3} \sum_{j=0}^m \omega_j^{(\delta_3)} z_{m-j} &= -\beta y_m - \gamma z_m - \beta \varepsilon \sin(\sigma x_{m\tau}), \end{aligned} \quad (4)$$

where $x_{m\tau}$ is the value of x at $t + \tau(t)$.

From (4), we have

$$\begin{aligned} x_m &= \{h^{\delta_1} [\alpha(y_m - x_m - f(x_m))] - \sum_{j=1}^m \omega_j^{(\delta_1)} x_{m-j}\} / (1 + h^{\delta_1}), \\ y_m &= \{h^{\delta_2} [x_m - y_m + z_m] - \sum_{j=1}^m \omega_j^{(\delta_2)} y_{m-j}\} / (1 + h^{\delta_2}), \\ z_m &= \{h^{\delta_3} [-\beta y_m - \gamma z_m - \beta \varepsilon \sin(\sigma x_{m\tau})] \\ &\quad - \sum_{j=1}^m \omega_j^{(\delta_3)} z_{m-j}\} / (1 + h^{\delta_3}), \end{aligned} \quad (5)$$

$m = 1, 2, \dots$

Formula (5) are implicit nonlinear algebraic equations respect to x_m , y_m and z_m , respectively. From (5), we can construct iteration algorithms to solve them as follows:

$$\begin{aligned} x_m^{(l)} &= \{h^{\delta_1} [\alpha(y_m^{(l-1)} - x_m^{(l-1)} - f(x_m^{(l-1)}))] \\ &\quad - \sum_{j=1}^m \omega_j^{(\delta_1)} x_{m-j}^{(l-1)}\} / (1 + h^{\delta_1}), \\ y_m^{(l)} &= \{h^{\delta_2} [x_m^{(l-1)} - y_m^{(l-1)} + z_m^{(l-1)}] - \sum_{j=1}^m \omega_j^{(\delta_2)} y_{m-j}^{(l-1)}\} / (1 + h^{\delta_2}), \\ z_m^{(l)} &= \{h^{\delta_3} [-\beta y_m^{(l-1)} - \gamma z_m^{(l-1)} - \beta \varepsilon \sin(\sigma x_{m\tau}^{(l-1)})] \\ &\quad - \sum_{j=1}^m \omega_j^{(\delta_3)} z_{m-j}^{(l-1)}\} / (1 + h^{\delta_3}), \end{aligned} \quad (6)$$

$$l = 1, 2, \dots, \quad m = 1, 2, \dots,$$

where l is the iteration number. When $|x_m^{(l)} - x_m^{(l-1)}| + |y_m^{(l)} - y_m^{(l-1)}| + |z_m^{(l)} - z_m^{(l-1)}| < \delta$ (given error, e.g. $\delta = 10^{-6}$), we would obtain the solution of the Eq.(2) at time t as $x_m^{(l)}, y_m^{(l)}, z_m^{(l)}$. If the eigenvalues λ_i ($i=1,2,3$) of the Jacobi matrix

$$\begin{bmatrix} -\alpha(1 + f'(x_m^{(l-1)})) / h_{\delta_1} & \alpha / h_{\delta_1} & 0 \\ 1 / h_{\delta_2} & -1 / h_{\delta_2} & 1 / h_{\delta_2} \\ 0 & -\beta / h_{\delta_3} & -\gamma / h_{\delta_3} \end{bmatrix}$$

($h_{\delta_1} = h^{\delta_1} / (1 + h^{\delta_1}), h_{\delta_2} = h^{\delta_2} / (1 + h^{\delta_2}), h_{\delta_3} = h^{\delta_3} / (1 + h^{\delta_3})$)

satisfy $|\lambda_i| < 1$, then iteration (6) is convergent. So just the calculus step h being taken smaller enough, the iteration (6) will be convergent.

We set the calculus step as $h=0.005$ for our numerical simulation. The phase portraits and the bifurcation diagram with $\alpha = 10.725$, $\beta = 10.593$, $\gamma = 0.268$, $\delta_1 = 0.93$, $\delta_2 = 0.99$, $\delta_3 = 0.92$, $\varepsilon = 0.5$, $\sigma = 0.05$, are shown as in Fig.1 and Fig.2.

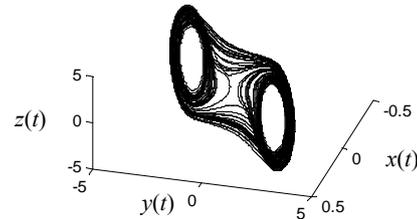


Fig. 1. phase portrait of $(x(t), y(t), z(t))$

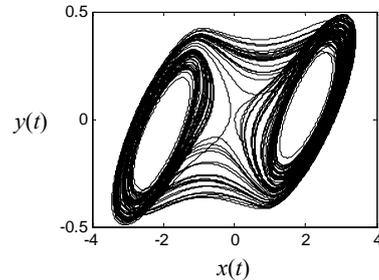


Fig 2. phase portrait of $(x(t), y(t))$

Simulation results imply that the system exhibits chaotic phenomena, which also indicating complex behaviors are found in time-delayed system.

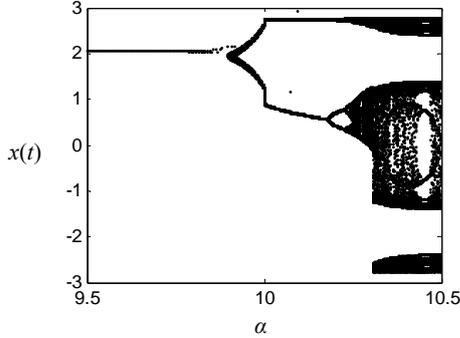


Fig. 3. The bifurcation diagram for system (1)

The bifurcation diagram which across $y=0$ of system (1) is as in Fig.3, and which of system (2) as α is used as the bifurcation parameter is shown in Fig. 4. Comparing Fig.3 with Fig.4, it is clearly that the dynamic behaviors of time-delayed Chua's system is more complex than that of the ordinary one. We can get that system (2) exhibits chaotic phenomena at $\alpha=10$, while system(1)at about $\alpha=10.3$.

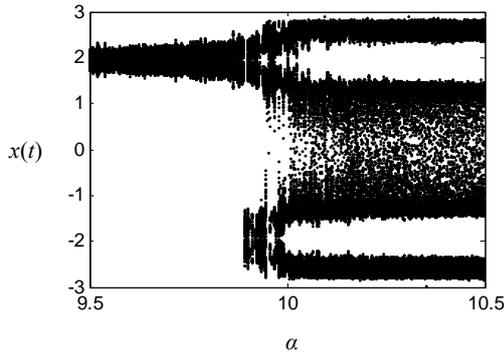


Fig. 4. the bifurcation diagram for system (2)

III. CHAOTIC SYNCHRONIZATION OF FRACTIONAL-ORDER CHUA'S SYSTEM

A. Synchronization Model

Now, we employ (2) as the drive system, and the response system is as follows:

$$\begin{aligned} \frac{d^{\delta_1} x_1}{dt^{\delta_1}} &= \alpha(y_1 - x - f(x)) + k_1(x - x_1), \\ \frac{d^{\delta_2} y_1}{dt^{\delta_2}} &= x_1 - y_1 + z_1 + k_2(y - y_1), \\ \frac{d^{\delta_3} z_1}{dt^{\delta_3}} &= -\beta y_1 - \gamma z_1 \\ &\quad - \beta \varepsilon \sin(\sigma x(t - \tau(t))) + k_3(z - z_1), \end{aligned} \quad (7)$$

Defining the synchronization error as $e_x(t)=x(t)-x_1(t)$, $e_y(t)=y(t)-y_1(t)$, $e_z(t)=z(t)-z_1(t)$, then we can obtain the following error system:

$$\begin{aligned} \frac{d^{\delta_1} e_x}{dt^{\delta_1}} &= -(\alpha + k_1)e_x + \alpha e_y, \\ \frac{d^{\delta_2} e_y}{dt^{\delta_2}} &= e_x - (1 + k_2)e_y + e_z, \\ \frac{d^{\delta_3} e_z}{dt^{\delta_3}} &= -\beta e_y - (\gamma + k_3)e_z, \end{aligned} \quad (8)$$

The Jacobi matrix of system (8) is

$$J = \begin{bmatrix} -(\alpha + k_1) & \alpha & 0 \\ 1 & -(1 + k_2) & 1 \\ 0 & -\beta & -(\gamma + k_3) \end{bmatrix}. \quad (9)$$

Suppose M is the lowest common multiple of the denominators u_i s of δ_i s, where $\delta_i = v_i/u_i$, $(u_i, v_i) = 1$, $u_i, v_i \in \mathbb{Z}^+$, $i=1,2,3$, then if all the roots λ_i s of the equation

$$\det \begin{bmatrix} \lambda^{M\delta_1} + (\alpha + k_1) & -\alpha & 0 \\ -1 & \lambda^{M\delta_2} + (1 + k_2) & -1 \\ 0 & \beta & \lambda^{M\delta_3} + (\gamma + k_3) \end{bmatrix} = 0$$

satisfies $|\arg(\lambda_i)| > \pi/2M$, the coupled system is synchronized [32].

B. Numerical Simulations

For calculus step h , we can not guarantee that $\tau(t)/h$ is an integer number. So for time t , we take approximate values of $x_\tau = x(t - \tau(t))$ as follows:

$$\begin{aligned} x_\tau &= x_{m+[\tau/h]} \frac{\tau/h - [\tau/h]}{h} \\ &\quad + x_{m+[\tau/h]+1} \frac{1 + [\tau/h] - \tau/h}{h}, \end{aligned}$$

where $[\tau/h]$ stands for the largest integer number which does not bigger than τ/h .

Then the discrete form of the response system can be written as follows:

$$\begin{aligned} h^{-\delta_1} \sum_{j=0}^m \omega_j^{(\delta_1)} x_{1,m-j} &= \alpha(y_{1,m} - x_{1,m} - f(x_m)) \\ &\quad + k_1(x_m - x_{1,m}), \\ h^{-\delta_2} \sum_{j=0}^m \omega_j^{(\delta_2)} y_{m-j} &= x_m - y_m + z_m + k_2(y_m - y_{1,m}), \\ h^{-\delta_3} \sum_{j=0}^m \omega_j^{(\delta_3)} z_{m-j} &= -\beta y_m - \gamma z_m - \beta \varepsilon \sin(\sigma x_{m\tau}) \\ &\quad + k_3(z_m - z_{1,m}). \end{aligned} \quad (10)$$

We set the calculus step as $h=0.001$ for our numerical simulation. Taking $\alpha = 10.725$, $\beta = 10.593$, $\gamma = 0.268$, $\delta_1 = 0.93$, $\delta_2 = 0.99$, $\delta_3 = 0.92$, $\tau = 0.5 + \sin^2(4.3t)$, and the activation function as $f(x) = bx + 0.5(a + b)(|x + 1| - |x - 1|)$, then

if $a=-1.1726$, $b=-0.7872$, the original system exhibit chaos phenomena (see Fig.1 to 4). Setting $k_1 =7$ and 10, the waveform diagrams are shown as in Fig.7 and 8, respectively.

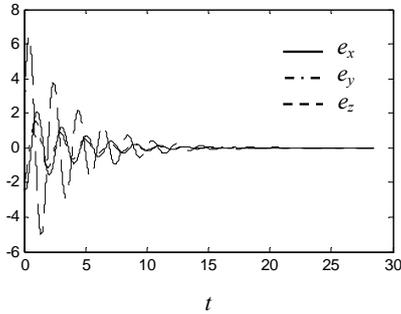


Fig. 7. Waveform diagram of system (7) with $k_1=7$, $k_2=0$, $k_3=0$

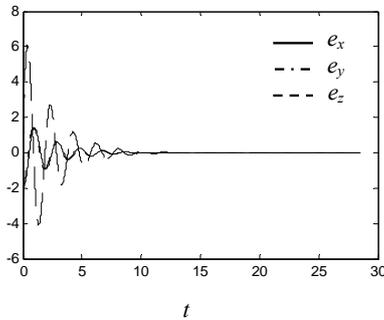


Fig. 8. Waveform diagram of system (7) with $k_1=10$, $k_2=0$, $k_3=0$

If we take the response system as follows:

$$\begin{aligned} \frac{d^{\delta_1} x_1}{dt^{\delta_1}} &= \alpha(y_1 - x - f(x)) + k_1(x - x_1), \\ \frac{d^{\delta_2} y_1}{dt^{\delta_2}} &= x_1 - y_1 + z_1 + k_2(y - y_1), \\ \frac{d^{\delta_3} z_1}{dt^{\delta_3}} &= -\beta y_1 - \gamma z_1 \\ &\quad - \beta \varepsilon \sin(\sigma x_1(t - \tau(t))) + k_3(z - z_1), \end{aligned} \quad (11)$$

where $\beta \varepsilon \sin(\sigma x(t - \tau(t))) + k_3(z - z_1)$ is replaced by $\beta \varepsilon \sin(\sigma x_1(t - \tau(t))) + k_3(z - z_1)$, in the response system. Setting $k=12$, the waveform diagrams of the coupled system (14) is shown as in Fig.9.

From Fig. 7 to 9, we see that the convergence situations of system (9) and (14) are almost the same, i.e. the nonlinear terms of $\beta \varepsilon \sin(\sigma x(t - \tau(t))) + k_3(z - z_1)$ or $\beta \varepsilon \sin(\sigma x_1(t - \tau(t))) + k_3(z - z_1)$, affects the coupled system weakly.

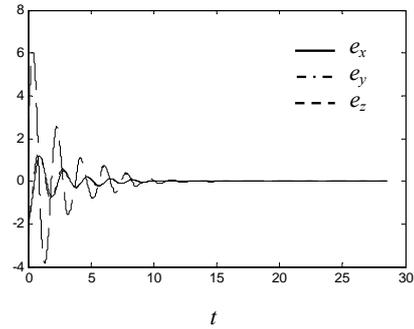


Fig. 9. Waveform diagram of system(11) with $k_1=10$, $k_2=0$, $k_3=0$

Replacing $f(x)$ by $f(x_1)$ in the response equation of the coupled system (7), we obtain following response system:

$$\begin{aligned} \frac{d^{\delta_1} x_1}{dt^{\delta_1}} &= \alpha(y_1 - x - f(x_1)) + k_1(x - x_1), \\ \frac{d^{\delta_2} y_1}{dt^{\delta_2}} &= x_1 - y_1 + z_1 + k_2(y - y_1), \\ \frac{d^{\delta_3} z_1}{dt^{\delta_3}} &= -\beta y_1 - \gamma z_1 \\ &\quad - \beta \varepsilon \sin(\sigma x(t - \tau(t))) + k_3(z - z_1). \end{aligned} \quad (12)$$

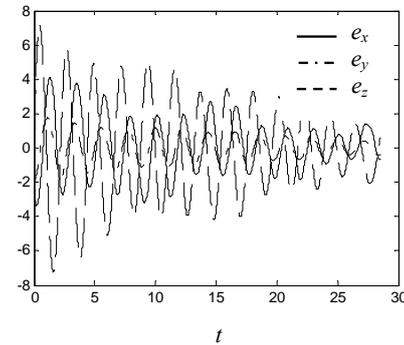


Fig. 10. Waveform diagram of system(11) with $k_1=14$, $k_2=0$, $k_3=0$

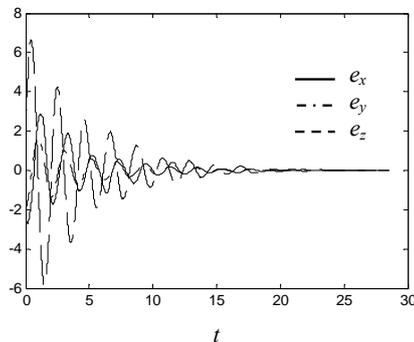


Fig. 11. Waveform diagram of system(11) with $k_1=16$, $k_2=0$, $k_3=0$

Setting $k=14$ and 16 , the waveform diagrams are shown as in Fig.10 and 11, respectively. From the simulation results, we can see that the coupled system is not convergent when $t < 30$ with $k < 14$. By doing some simulation, we come out that when the parameter $k > 16$, the synchronization system is convergent.

Furthermore, $x(t-\tau(t))$ and x is replaced by $x_1(t-\tau)$, x_1 in the response equation of the coupled system (12), coupled system be changed into as Eq.(13). The waveform diagrams of the coupled system with $k=7$ is shown as in Fig. 12.

$$\begin{aligned} \frac{d^{\delta_1} x_1}{dt^{\delta_1}} &= \alpha(y_1 - x_1 - f(x_1)) + k_1(x - x_1), \\ \frac{d^{\delta_2} y_1}{dt^{\delta_2}} &= x_1 - y_1 + z_1 + k_2(y - y_1), \\ \frac{d^{\delta_3} z_1}{dt^{\delta_3}} &= -\beta y_1 - \gamma z_1 - \beta \varepsilon \sin(\sigma x_1(t - \tau(t))) \\ &\quad + k_3(z - z_1), \end{aligned} \quad (13)$$

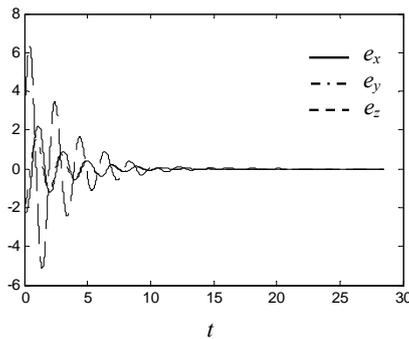


Fig. 12. Waveform diagram of system(13) with $k_1=7$, $k_2=0$, $k_3=0$

From above discussion and simulation results, we see that, for system (2), when the synchronization is linear, the parameter k is rather large for system to convergence; If the synchronization is nonlinear as system (9), k is smaller. But nonlinear system is more complex than linear one, and the system is hard to be complemented when the parameter k is large. The simulations results also show that the delay terms somewhat affect the behaviors of the system. Fig. 13. is the bifurcation diagram for synchronization of system (13). The system achieve the synchronization at $k_1=7$.

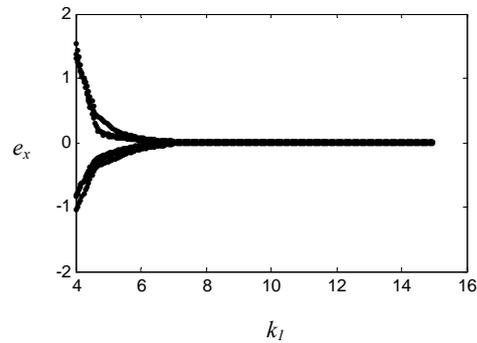


Fig. 13. The bifurcation diagram of system (13)

IV. CONCLUSIONS

In this paper, chaotic phenomena in a time-varying delayed fractional-order Chua's system are discussed. A numerical simulation algorithm for fractional-order dynamics model with explicit delays is presented. By comparing several synchronization models, we conclude that the chaos in such fractional-order system could be synchronized by several ways. How to depress limit of parameter k_1 in the linear coupled system and multi-delays coupling ways will be considered in our further researching. Chaotic synchronization can be studied by Laplace transformation theory. But the characteristic equation is a high order one and the eigenvalue is hard to be obtained. So how to obtain stability condition for a fractional-order system should be investigated in the further work.

REFERENCES

- [1] Siegmund Kempe and Ingo Schaffer, et al, "Fractional Calculus via Functional Calculus: Theory and Applications," *Nonlinear Dynamics*, vol.29, pp. 99–127, 2002.
- [2] YangQuan Chen, Dingy'u Xue and Huifang Dou, "Fractional Calculus and Biomimetic Control," *Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics* Shenyang, China, pp.901-905, 2004.
- [3] YangQuan Chen, Blas M. Vinagre, and Igor Podlubny, "Continued fraction expansion approaches to discretizing fractional order derivatives - an expository review", *Nonlinear Dynamics (Kluwer)*, vol. 9, no. 1, pp. to appear, 2004.
- [4] D. Ingman, J. Suzdalnitsky, "Control of damping oscillations by fractional differential operator with time-dependent order," *Comput. Methods Appl. Mech. Engrg.*, vol.193, pp.5585–5595, 2004.
- [5] Wajdi M. Ahmad, J.C. Sprott, "Chaos in fractional-order autonomous nonlinear systems," *Chaos, Solitons and Fractals*, vol.16, pp.339–351, 2003.
- [6] Wajdi M. Ahmad, Ahmad M. Harb, "On nonlinear control design for autonomous chaotic systems of integer and fractional orders," *Chaos, Solitons and Fractals*, vol.18, pp. 693–701, 2003.
- [7] Wajdi M. Ahmad, Reyad El-Khazali, Yousef Al-Assaf, "Stabilization of generalized fractional order chaotic systems using state feedback control," *Chaos, Solitons and Fractals*, vol.22, pp.141–150, 2004.
- [8] Fernando B. M. Duarte, "Chaotic Phenomena and Fractional-Order Dynamics in the Trajectory Control of Redundant Manipulators," *Nonlinear Dynamics*, vol. 29, pp. 315–342. , 2002.
- [9] Tom T. Hartley, Carl F. Lorenzo, and Helen Killory Qammer, "Chaos in a Fractional Order Chua's System," *IEEE Trans. On Circuit and System — I: Fundamental Theory and Applications*, vol.42, no.8, pp.485-490, August 1995.

- [10] Marco Gilli. "Strange Attractors in Delayed Cellular Neural Networks," IEEE trans. on CAS-1, vol.40, no.11, pp.849-853, 1993
- [11] K. Murali, M. Lakshmanan, and L. O. Chua. "The Simplest Dissipative Nonautonomous Chaotic Circuit," IEEE trans. on CAS-1, vol.41, no.6, pp.462-463, 1994.
- [12] Stuart Nimmo and Allan Kevans, "The Effects of Continuously Varying the Fractional Differential Order of Chaotic Nonlinear Systems," Chaos, Solitons and Fractals, vol.10, no.7, 1999.
- [13] P. Arena, R. Caponetto, L. Fortuna and D. Porto, "Bifurcation and Chaos in Noninteger Order Cellular Neural Network," International Journal of Bifurcation and Chaos, vol. 8, no. 7 pp.1527-1539, 1998.
- [14] P. Arena, L. Fortuna, and D. Porto, "Chaotic behavior in noninteger-order cellular neural networks," Physical Review E, vol.61, no.1, pp. 777-781, Jan. 2000.
- [15] Yousef Al-Assaf, Reyad El-Khazali, Wajdi Ahmad, "Identification of fractional chaotic system parameters," Chaos, Solitons & Fractals, vol.22, pp. 897-905, 2004.
- [16] Wajdi M. Ahmad, "Generation and control of multi-scroll chaotic attractors in fractional order systems," Chaos, Solitons & Fractals, vol.25, pp.727-735, 2005.
- [17] Wei-Ching Chen, "Nonlinear dynamics and chaos in a fractional-order financial system," Chaos, Solitons & Fractals, vol.36, pp. 1305-1314, 2008
- [18] Li CG, Chen G."Chaos in the fractional order Chen system and its control," Chaos, Solitons & Fractals, vol.22, pp.549-554, 2004
- [19] Lu JG. "Chaotic dynamics and synchronization of fractional-order Chua's circuits with a piecewise-linear nonlinearity," *Int J Modern Phys B*, vol.19, no.20, pp.3249-3259, 2005.
- [20] C.P.Li, W.H.Deng, D. Xu. "Chaos synchronization of the Chua system with a fractional order," *Physica A*, vol.360, pp.171-185, 2006.
- [21] Junwei Wang, Xiaohua Xiong, Yanbin Zhang, "Extending synchronization scheme to chaotic fractional-order Chen systems," *Physica A*, vol.370, pp.279-285, 2006.
- [22] Shangbo Zhou, Xiaofeng Liao, Juebang Yu, Kwok-Wo Wong. "Chaos and its Synchronization in Two-Neuron Systems with Discrete Delay," *Chaos, Solitons and Fractals*, vol.21, no.1, pp.133-142, 2004.
- [23] Shangbo Zhou, Hua Li, Zhengzhou Zhu, "Chaos control and synchronization in a fractional neural network system," *Chaos, Solitons and Fractals*, vol.36, no.4, pp. 973-984, 2008
- [24] Hao Zhu, Shangbo Zhou, Weiwei Zhang, "Chaos and Synchronization of Time-Delayed Fractional Neuron Network," *The 9th International Conference for Young Computer Scientists*, pp.2937-2941, 2008
- [25] G.H. Erjaee, Shaher Momani, "Phase synchronization in fractional differential chaotic systems", *Physics Letters A*, vol.372, pp. 2350-2354, 2008
- [26] Guojun Peng, Yaolin Jiang, Fang Chen, "Generalized projective synchronization of fractional order chaotic systems," *Physica A*, vol.387, pp.3738-3746, 2008
- [27] Ivo Petras, "A note on the fractional-order Chua's system," *Chaos, Solitons and Fractals* 38, pp. 140-147, 2008
- [28] C.P. Lia, W.H. Deng, D. Xu, "Chaos synchronization of the Chua system", *Physica A*, vol.360, pp. 171-185, 2006
- [29] Mohammad Saleh Tavazoei, Mohammad Haeri, "Synchronization of chaotic fractional-order systems via activesliding mode controller," *Physica A*, vol.387, pp. 57-70, 2006
- [30] C. Cruz-Hernández, N. Romero-Haros, "Communicating via synchronized time-delay Chua's circuits," *Communications in Nonlinear Science and Numerical Simulation*, vol.13, pp.645-659, 2006
- [31] Arman Kiani-B, Kia Fallahi, Naser Pariz, Henry Leung, "A chaotic secure communication scheme using fractional chaotic systems based on an extended fractional Kalman filter," *Communications in Nonlinear Science and Numerical Simulation*, vol.14, pp. 863-879, 2009
- [32] Weihua Deng, Changpin Li, Jinhua Lu, "Stability analysis of linear fractional differential system with multiple time delays," *Nonlinear Dyn*, vol.48, pp.409-416, 2007

Shangbo Zhou was born in Guangxi Province, China, in 1963. He received the B.S degree from the Guangxi university for Nationalities, Nanning, in 1985, the M.S. degree from the University of Sichuan, in 1991, and the Ph.D degree from University of Electronic Science and Technology of China, in 2003. He is now an professor with the Computer Department of Chongqing University. His research interests include nonlinear dynamics, artificial neural network and information security.

Xiaoran Lin was born in HeBei Province, China, in 1983. She received the B.S degree from the Hebei University of Science and Technology, Shijiazhuang, in 2005. She is currently pursuing the M.S. degree with the Computer Department of Chongqing University. Her research interests include nonlinear dynamics, and artificial neural network.

Hua Li received his B.E. and M.S. degrees from Beijing Polytechnic University and Peking University. He received his Ph.D. from University of Regina. Currently, he works as an assistant professor at Department of Mathematics and Computer Science, University of Lethbridge, Canada. His research interests include parallel systems, reconfigurable computing, fault-tolerant, VLSI design, and information and network security. He is a member of IEEE.

Solution of the Fractional Diffusion Equation with Absorbent Term and External Force

Subir Das*

Department of Applied Mathematics, Institute of Technology
Banaras Hindu University, Varanasi - 221 005, INDIA

Abstract—In the present paper, the approximate analytical solutions of general diffusion equation with fractional time derivative in the presence of an absorbent term and a linear external force are obtained with the help of powerful Homotopy Perturbation Method (HPM). By using initial value, the approximate analytical solutions of the equation are derived. The results are deduced for different particular cases. The present method is extremely simple, concise and highly efficient as a mathematical tool in comparison with the other existing techniques.

I. INTRODUCTION

THE HPM is an approach for finding the approximate analytical solution of linear and nonlinear problems. The method was proposed by He ([1], [2]) and was successfully applied to solve linear and nonlinear equations by He ([3] - [13]). The fundamentals of the method can be found, for example, in He ([14], [15]). The basic difference of this method from the other perturbation techniques is that it does not require small parameters in the equation which overcomes the limitations of traditional perturbation techniques.

We focus our attention to find the solution of the equation (Chot et al. [16])

$$\frac{\partial^\beta}{\partial t^\beta} u(x, t) = D \frac{\partial^2}{\partial x^2} u(x, t) - \frac{\partial}{\partial x} [F(x) u(x, t)] - \int_0^t \alpha(t - \xi) u(x, \xi) d\xi \quad (1)$$

where, D is a diffusion coefficient, $F(x)$ is an external force, $\alpha(t)$ is a time-dependent absorbent term which may be related to a reaction diffusion process.

The diffusion equations have been widely studied due to their various applications in Physics and engineering, but the study related to diffusion equations with nonlinear terms and fractional time derivatives are few in number. Lenzi et al. [17] presented some classes of solutions of a general nonlinear fractional diffusion equation with absorptions. The similar study was made by Assis et al. [18]. Recently Das [19] has used

*corresponding author email: subir_das08@hotmail.com (S. Das).

Variational Iteration Method to find the analytical solution of a fractional diffusion equation of order α ($0 < \alpha \leq 1$) only in the presence of external force. Schot et al. [16] has given an approximate solution of the equation with absorbent term and a linear external force in terms of Fox H-function. Zahran [20] has given a closed form solution in Fox H-function of the generalized fractional reaction-diffusion equation subject to an external linear force field to describe the transport processes in disorder systems.

In this paper the Homotopy Perturbation Method is used to solve the fractional diffusion equation problem in the presence of both linear external force and an absorbent term. Using the initial condition, the approximate analytical expressions of $u(x, t)$ for different Brownian motions are obtained. The effect of external force and absorbent term in the solution is obtained for different particular cases. The elegance of this method can be attributed to its simplistic approach in seeking the approximate analytical solution of the problem.

II. SOLUTION OF THE PROBLEM

Our aim is to solve the analytical fractional diffusion equation (1) for $D = 1$ and $F(x) = -kx$ i.e., the equation now becomes

$$\frac{\partial^\beta u(x, t)}{\partial t^\beta} = \frac{\partial^2 u(x, t)}{\partial x^2} + k \frac{\partial}{\partial x} (xu(x, t)) - \int_0^t \alpha(t - \xi) \cdot u(x, \xi) d\xi \quad (2)$$

with initial condition

$$u(x, 0) = f(x) \quad (3)$$

Equation (9) can be written in operator form as

$$D_t^\beta u(x, t) = D_{xx} u(x, t) + kx D_x (u(x, t)) + k u(x, t) - \int_0^t \alpha(t - \xi) \cdot u(x, \xi) d\xi \quad (4)$$

where, $D_t^\beta \equiv \frac{\partial^\beta}{\partial t^\beta}$

According to the homotopy perturbation method, we construct the following homotopy

$$D_t^\beta u(x,t) = p [D_{xx} u(x,t) + k x D_x(u(x,t)) + k u(x,t) - \int_0^t \alpha(t-\xi) \cdot u(x,\xi) d\xi] \quad (5)$$

where the homotopy parameter p is considered as a small parameter ($p \in [0,1]$). Now applying the classical perturbation technique, we can assume that the solution of equation (2) can be expressed as a power series in p as given below:

$$u(x,t) = u_0(x,t) + p u_1(x,t) + p^2 u_2(x,t) + p^3 u_3(x,t) + p^4 u_4(x,t) + \dots \quad (6)$$

When $p = 1$, equation (5) corresponds equation (4) and (6) becomes the approximate solution of (4) i.e., of equation (2). The convergence of the method has been proved in [2]. Substituting equation (6) for equation (5), and equating the terms with the identical powers of p , we can obtain a series of equations:

$$p^0 : D_t^\beta u_0(x,t) = 0 \quad (7)$$

$$p^1 : D_t^\beta u_1(x,t) = D_{xx} u_0(x,t) + k x D_x(u_0(x,t)) + k u_0(x,t) - \int_0^t \alpha(t-\xi) \cdot u_0(x,\xi) d\xi \quad (8)$$

$$p^2 : D_t^\beta u_2(x,t) = D_{xx} u_1(x,t) + k x D_x(u_1(x,t)) + k u_1(x,t) - \int_0^t \alpha(t-\xi) \cdot u_1(x,\xi) d\xi \quad (9)$$

$$p^3 : D_t^\beta u_3(x,t) = D_{xx} u_2(x,t) + k x D_x(u_2(x,t)) + k u_2(x,t) - \int_0^t \alpha(t-\xi) \cdot u_2(x,\xi) d\xi \quad (10)$$

$$p^4 : D_t^\beta u_4(x,t) = D_{xx} u_3(x,t) + k x D_x(u_3(x,t)) + k u_3(x,t) - \int_0^t \alpha(t-\xi) \cdot u_3(x,\xi) d\xi \quad (11)$$

and so on.

The method is based on applying the operator J^β (the inverse of Caputo operator D_t^β) on both sides of the equations (7) – (11), we obtain the solutions of $u_i(x,t)$, $i \geq 0$ for different expressions of $\alpha(t)$.

Here we will discuss when the Absorbent term is $\alpha(t) = \frac{\alpha t^{\beta-1}}{\Gamma(\beta)}$, $0 < \beta \leq 1$,

In this case, the equations (7) – (11) give rise to

$$u_0(x,t) = f(x) \quad (12)$$

$$u_1(x,t) = \phi_1(x) \frac{t^\beta}{\Gamma(\beta+1)} - \alpha f(x) \frac{t^{2\beta}}{\Gamma(2\beta+1)} \quad (13)$$

$$u_2(x,t) = \phi_2(x) \frac{t^{2\beta}}{\Gamma(2\beta+1)} - 2\alpha \phi_1(x) \frac{t^{3\beta}}{\Gamma(3\beta+1)} + \alpha^2 f(x) \frac{t^{4\beta}}{\Gamma(4\beta+1)} \quad (14)$$

$$u_3(x,t) = \phi_3(x) \frac{t^{3\beta}}{\Gamma(3\beta+1)} - 3\alpha \phi_2(x) \frac{t^{4\beta}}{\Gamma(4\beta+1)} + 3\alpha^2 \phi_1(x) \frac{t^{5\beta}}{\Gamma(5\beta+1)} - \alpha^3 f(x) \frac{t^{6\beta}}{\Gamma(6\beta+1)} \quad (15)$$

$$u_4(x,t) = \phi_4(x) \frac{t^{4\beta}}{\Gamma(4\beta+1)} - 4\alpha \phi_3(x) \frac{t^{5\beta}}{\Gamma(5\beta+1)} + 6\alpha^2 \phi_2(x) \frac{t^{6\beta}}{\Gamma(6\beta+1)} - 4\alpha^3 \phi_1(x) \frac{t^{7\beta}}{\Gamma(7\beta+1)} + \alpha^4 f(x) \frac{t^{8\beta}}{\Gamma(8\beta+1)} \quad (16)$$

where, $\phi_1(x) = f''(x) + k x f'(x) + k f(x)$,

$\phi_{r+1}(x) = \phi_r(x) + k x \phi_r(x) + k \phi_r(x)$,

and $f^{(r)}(x) = \frac{\partial^r}{\partial x^r} (f(x))$, $r \geq 1$

Proceeding in this manner the components u_n , $n \geq 0$ of the Homotopy Perturbation Method can be completely obtained, and the series solutions are thus entirely determined.

Finally we approximate the analytical solution of $u(x, t)$ by the truncated series

$$u(x, t) = \lim_{N \rightarrow \infty} \Phi_N(x, t) \quad (17)$$

where $\Phi_N(x, t) = \sum_{n=0}^{N-1} u_n(x, t)$.

The above series solutions generally converge very rapidly.

III. PARTICULAR CASES

Case I: If $f(x) = x$, $\alpha = 0$, $k = 1$ i.e., in the presence of only external force, the expression of the displacement becomes,

$$\begin{aligned} u(x, t) &= x \left[1 + \frac{2t^\beta}{\Gamma(\beta+1)} + \frac{4t^{2\beta}}{\Gamma(2\beta+1)} + \frac{8t^{3\beta}}{\Gamma(3\beta+1)} \right] \\ &\quad + x \left[\frac{16t^{4\beta}}{\Gamma(4\beta+1)} + \dots \right] \\ &= x \sum_{r=0}^{\infty} \frac{2^r t^{r\beta}}{\Gamma(r\beta+1)} \\ &= x E_\beta(2t^\beta) \end{aligned} \quad (18)$$

Case II: If $f(x) = x$, $\alpha = 1$, $k = 0$ i.e., in the presence of the absorbent term,

$$\begin{aligned} u(x, t) &= x \left[1 - \frac{t^{2\beta}}{\Gamma(2\beta+1)} + \frac{t^{4\beta}}{\Gamma(4\beta+1)} - \frac{t^{6\beta}}{\Gamma(6\beta+1)} \right] \\ &\quad + x \left[\frac{t^{8\beta}}{\Gamma(8\beta+1)} - \dots \right] \\ &= x \sum_{r=0}^{\infty} \frac{(-1)^r t^{2r\beta}}{\Gamma(2r\beta+1)} \\ &= x E_{2\beta}(-t^{2\beta}) \end{aligned} \quad (19)$$

Case III: If $f(x) = x$, $\alpha = 1$, $k = 1$ i.e., in the presence of both the linear external force and absorbent term,

$$u(x, t) = x \left[1 + \frac{2t^\beta}{\Gamma(\beta+1)} + \frac{3t^{2\beta}}{\Gamma(2\beta+1)} + \frac{4t^{3\beta}}{\Gamma(3\beta+1)} \right]$$

$$\begin{aligned} &+ x \left[\frac{5t^{4\beta}}{\Gamma(4\beta+1)} + \dots \right] \\ &= x \sum_{r=0}^{\infty} \frac{(r+1)t^{r\beta}}{\Gamma(r\beta+1)} \\ &= x E_\beta(Kt^\beta) \end{aligned} \quad (20)$$

where, $K^r = (r+1)$

IV. CONCLUSION

There are two important goals that have been achieved for this study. First one is employing the powerful HPM to investigate the general diffusion equation for different particular situations. HPM is a powerful mathematical tool which reduces the nonlinear problems to a set of ordinary differential equations to get the approximate analytical solution easily. Moreover it does not require small parameters in the equations which overcome the limitations of traditional perturbation techniques. This method is very effective, convenient, supplies quantitatively reliable results.

Another important point of this study is to derive the expressions of $u(x, t)$ in the presence of external force, source term and also both the terms simultaneously for the fractional diffusion equation. The author strongly believes that the approximate analytical expressions of displacement for different particular cases discussed in this article will provide significant change from the usual approach to the engineers and physicists working in this area of research.

REFERENCES

- [1]. J. H. He, *Comput. Methods in Appl. Mech. and Engng.* **178**, 257 (1999).
- [2]. J. H. He, *Int. J. of Nonlin. Mech.* **35**, 37 (2000).
- [3]. J. H. He, *Phys. Lett. A* **347**, 228 (2005).
- [4]. J. H. He, *Chaos, Solitons & Fractals* **26**, 695 (2005).
- [5]. J. H. He, *Chaos, Solitons & Fractals* **26**, 827 (2005).
- [6]. J. H. He, *Int. J. of Nonlin. Sci. and Numer. Simul.* **6**, 207 (2005).
- [7]. J. H. He, *Int. J. of Modern Phys. B* **20**, 1141 (2006).
- [8]. J. H. He, *Phys. Lett. A* **350**, 87 (2006).
- [9]. M. El-Shahed, *Int. J. of Nonlin. Sci. and Numer. Simul.* **6**, 163 (2005).
- [10]. A. M. Siddiqui, R. Mahmood, Q. K. Ghori, *Int. J. of Nonlin. Sci. and Numer. Simul.* **7**, 7 (2006).
- [11]. A. M. Siddiqui, R. Mahmood, Q. K. Ghori, *Int. J. of Nonlin. Sci. and Numer. Simul.* **7**, 15 (2006).
- [12]. J. H. He, *Comp. Methods in Appl. Mech. and Engng.* **167**, 57 (1998).
- [13]. S. Momani, Z. Odibat, *Comput. Math. Appld.* **54**, 910 (2007).

- [14]. J. H. He, Topolog. Methods in Nonlin. Anal. **31**, 205 (2008).
- [15]. J. H. He, Int. J. of Modern Physics B **22**, 3487 (2008) .
- [16]. A. Schot, M. K. Lenzi, L. R. Evangelista, L. C. Malacarne, R. S. Mendes, E. K. Lenzi, Phys. Lett. A **366**, 346 (2007).
- [17]. E. K. Lenzi , R. S. Mendes , K. S. Fa , L. S. Mendes , L. R. Silva , L. S. Lecena, J. Math. Phys. **46**, 083506(2005).
- [18]. P. C. Assis, L. R. da Silva, E. K. Lenzi, L. C. Malakarne, R. S. Mendes, J. Math. Phys. **46**, 123303(2005).
- [19]. S. Das , Chaos, Solitons & Fractals. In press (2008).
- [20]. M. A. Zahran, Appld. Math. Model. **33**, 3088 (2008).

Solution of Vibration equation by Homotopy Analysis Method

Ms. S. Chakraborty

Abstract— In this paper the author has solved the vibration equation for large membrane with the help of a powerful mathematical tool called Homotopy Analysis Method .By using initial values, the explicit solutions of the equations for different particular cases have been derived. The examples prove that the method is extremely effective due to its simplistic approach and performance. The numerical results so obtained are discussed in Section 3 and depicted graphically.

Index Terms—, Homotopy analysis method, Initial value problem ,Vibration equation .

I. INTRODUCTION

Homotopy Analysis Method (HAM) was first proposed by Liao [1], by employing the basic ideas of homotopy in topology to produce an analytical method for solving various nonlinear problems. This method has been successfully applied by (for example, please refer to [2]-[5]) to solve different classes of nonlinear problems.

Recently, Das [6] has used Modified decomposition method (MDM) to find the numerical solution of a vibration equation for large membrane. The main disadvantage of this method is that the solution procedure for calculation of Adomian polynomials is complex and difficult as pointed out by many researchers.

In this article HAM is used to obtain the approximate analytical solutions of the vibration equation for very large membrane. The expressions of the displacement for different time and radii of the membrane and also for various wave velocities of free vibration using the initial conditions are deduced which shows that the proposed method gives much better approximations than those given by non perturbative methods like MDM and other existing traditional techniques and does it in a simplistic manner too.

II. SOLUTION OF THE PROBLEM

The vibration equation of a very large membrane is governed by the equation

Author is with the Department of Electrical Engineering, B.P.Poddar Institute of Management & Technology, Kolkata,Poddar Vihar, Kolkata - 700052, W.B. , INDIA (phone: +91-33-25739609/10; fax: +91-33-2573-9401; e-mail: sudipta_ch1@rediffmail.com).

$$\frac{\partial^2 u}{\partial t^2} = c^2 \left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} \right), \quad r \geq 0, t \geq 0 \quad (1)$$

with the initial conditions

$$u(r, 0) = f(r) \quad \text{and} \quad \frac{\partial u(r, 0)}{\partial t} = c g(r) \quad , \quad (2)$$

where $u(r, t)$ represents displacement of the particle at the point r and at time t , c is the wave velocity of free vibration. To solve equation (1) by means of HAM, we choose the initial approximation

$$u_0(r, t) = f(r) + c t g(r),$$

and the linear operator

$$L[\phi(r, t; q)] = \frac{\partial^2 \phi(r, t; q)}{\partial t^2}, \quad (3)$$

with the property

$$L[C_1 + C_2 t] = 0 \quad , \quad (4)$$

where C_1 and C_2 are integral constants. Furthermore, equation (1) suggests that we define the equation of nonlinear operator as

$$N[\phi(r, t; q)] = \frac{\partial^2 \phi(r, t; q)}{\partial t^2} - c^2 \left(\frac{\partial^2 \phi(r, t; q)}{\partial r^2} + \frac{1}{r} \frac{\partial \phi(r, t; q)}{\partial r} \right) \quad , \quad (5)$$

Now, we construct the zeroth-order deformation equation

$$(1 - q)L[\phi(r, t; q) - Z_0(r, t)] = q \hbar N[\phi(r, t; q)] \quad , \quad (6)$$

Obviously, when $q = 0$ and $q = 1$,

$$\phi(r, t; 0) = Z_0(r, t) = u_0(r, t) \quad \text{and}$$

$$\phi(r, t; 1) = u(r, t) \quad . \quad (7)$$

Therefore, as the embedding parameter q increases from 0 to 1, $\phi(r, t; q)$ varies from the initial guess $Z_0(r, t)$ to the solution $Z(r, t)$. Expanding $\phi(r, t; q)$ in Taylor series with respect to q , one can deduce

$$\phi(r, t; q) = Z_0(r, t) + \sum_{m=1}^{\infty} Z_m(r, t) q^m \quad ,$$

where,

$$Z_m(r, t) = \frac{1}{m!} \left. \frac{\partial^m \phi(r, t; q)}{\partial q^m} \right|_{q=0} \quad .$$

If the auxiliary linear operator, the initial guess and the auxiliary parameter \hbar are properly chosen, the above series is convergent at $q = 1$, then one has

$$u(r, t) = Z_0(r, t) + \sum_{m=1}^{\infty} Z_m(r, t),$$

which must be one of the solutions of the original nonlinear equation. Now we define the vector

$$\vec{Z}_n(r, t) = \{Z_0(r, t), Z_1(r, t), \dots, Z_n(r, t)\},$$

Then the m th-order deformation equation is

$$L[Z_m(r, t) - \chi_m Z_{m-1}(r, t)] = \hbar R_m(\vec{Z}_{m-1}(r, t)), \quad (8)$$

with the initial conditions

$$Z_m(r, 0) = 0 \quad \text{and} \quad \frac{\partial}{\partial t} Z_m(r, 0) = 0, \quad (9)$$

where

$$R_m(\vec{Z}_{m-1}(r, \tau)) = (Z_{m-1})_{tt} - c^2 (Z_{m-1})_{rr} - \frac{c^2}{r} (Z_{m-1})_r, \quad (10)$$

Now, the solution of the m^{th} -order deformation equation (8) for $m \geq 1$ becomes

$$Z_m(r, t) = \chi_m Z_{m-1}(r, t) + \hbar \int_0^t \int_0^\xi R_m(\vec{Z}_{m-1}(r, \tau)) d\tau d\xi + C_1 + C_2 t \quad (11)$$

where the integrating constants C_1 and C_2 are determined by the initial conditions (9).

We now successively obtain

$$\begin{aligned} Z_1(r, t) &= c^2 \hbar \left(f''(r) + \frac{1}{r} f'(r) \right) t \\ &\quad + c^3 \hbar \left(g''(r) + \frac{1}{r} g'(r) \right) \frac{t^2}{2!} \\ Z_2(r, t) &= -c^2 \hbar (1 + \hbar) \left(f''(r) + \frac{1}{r} f'(r) \right) \frac{t^2}{2!} \\ &\quad - c^3 \hbar (1 + \hbar) \left(g''(r) + \frac{1}{r} g'(r) \right) \frac{t^3}{3!} \\ &\quad + c^4 \hbar^2 \left(\begin{array}{l} f^{(4)}(r) + \frac{2}{r} f^{(3)}(r) \\ -\frac{1}{r^2} f''(r) + \frac{1}{r^3} f'(r) \end{array} \right) \frac{t^4}{4!} \\ &\quad + c^5 \hbar^2 \left(\begin{array}{l} g^{(4)}(r) + \frac{2}{r} g^{(3)}(r) \\ -\frac{1}{r^2} g''(r) + \frac{1}{r^3} g'(r) \end{array} \right) \frac{t^5}{5!} \end{aligned}$$

$$\begin{aligned} Z_3(r, t) &= -c^2 \hbar (1 + \hbar)^2 \left(f''(r) + \frac{1}{r} f'(r) \right) \frac{t^2}{2!} \\ &\quad - c^3 \hbar (1 + \hbar)^2 \left(g''(r) + \frac{1}{r} g'(r) \right) \frac{t^3}{3!} \\ &\quad + 2c^4 \hbar^2 (1 + \hbar) \left(\begin{array}{l} f^{(4)}(r) + \frac{2}{r} f^{(3)}(r) \\ -\frac{1}{r^2} f''(r) + \frac{1}{r^3} f'(r) \end{array} \right) \frac{t^4}{4!} \\ &\quad + 2c^5 \hbar^2 (1 + \hbar) \left(\begin{array}{l} g^{(4)}(r) + \frac{2}{r} g^{(3)}(r) \\ -\frac{1}{r^2} g''(r) + \frac{1}{r^3} g'(r) \end{array} \right) \frac{t^5}{5!} \\ &\quad - c^6 \hbar^3 \left(f^{(6)}(r) + \frac{3}{r} f^{(5)}(r) - \frac{3}{r^2} f^{(4)}(r) \right. \\ &\quad \left. + \frac{6}{r^3} f^{(3)}(r) - \frac{9}{r^4} f''(r) + \frac{9}{r^5} f'(r) \right) \frac{t^6}{6!} \\ &\quad - c^7 \hbar^3 \left(g^{(6)}(r) + \frac{3}{r} g^{(5)}(r) - \frac{3}{r^2} g^{(4)}(r) \right. \\ &\quad \left. + \frac{6}{r^3} g^{(3)}(r) - \frac{9}{r^4} g''(r) + \frac{9}{r^5} g'(r) \right) \frac{t^7}{7!} \end{aligned}$$

and so on.

Then the series solution expression by HAM can be written in the form

$$u(r, t) = Z_0(r, t) + Z_1(r, t) + Z_2(r, t) + Z_3(r, t) + \dots, \quad (12)$$

or specially when $\hbar = -1$,

$$\begin{aligned} u(r, t) &= f(r) + ct g(r) + c^2 \left(f''(r) + \frac{1}{r} f'(r) \right) \frac{t^2}{2!} \\ &\quad + c^3 \left(g''(r) + \frac{1}{r} g'(r) \right) \frac{t^3}{3!} \\ &\quad + c^4 \left(f^{(4)}(r) + \frac{2}{r} f^{(3)}(r) - \frac{1}{r^2} f''(r) + \frac{1}{r^3} f'(r) \right) \frac{t^4}{4!} \\ &\quad + c^5 \left(g^{(4)}(r) + \frac{2}{r} g^{(3)}(r) - \frac{1}{r^2} g''(r) + \frac{1}{r^3} g'(r) \right) \frac{t^5}{5!} \\ &\quad + c^6 \left(f^{(6)}(r) + \frac{3}{r} f^{(5)}(r) - \frac{3}{r^2} f^{(4)}(r) + \frac{6}{r^3} f^{(3)}(r) \right. \\ &\quad \left. - \frac{9}{r^4} f''(r) + \frac{9}{r^5} f'(r) \right) \frac{t^6}{6!} \\ &\quad + c^7 \left(g^{(6)}(r) + \frac{3}{r} g^{(5)}(r) - \frac{3}{r^2} g^{(4)}(r) \right. \\ &\quad \left. + \frac{6}{r^3} g^{(3)}(r) - \frac{9}{r^4} g''(r) + \frac{9}{r^5} g'(r) \right) \frac{t^7}{7!}, \quad (13) \end{aligned}$$

The convergence of the solution series is determined by h pointed out by Liao [1,7], and thus one can get a convergent series solution by chosen value of h .

III. NUMERICAL RESULTS AND DISCUSSION

In this section, numerical values of $u(r, t)$ for various values of radii of the membrane and time are presented for the initial condition $f(r) = r$ and $g(r) = 1$ through Figs. 1-2. It is observed from the Fig. 1 that the $u(r, t)$ decreases with the increase in r and increases with the increase in t at $c = 6$. It is also seen from Fig.2 that the $u(r, t)$ increases with the increase in t and c both at a fixed value of the radius of the membrane (for $r = 20$). The Numerical Calculations and Figures are made using Mathematica software (Version 5.2).

REFERENCES

- [1] S. J. Liao, "On the proposed homotopy analysis technique for nonlinear problems and its applications", Ph.D. Dissertation. Shanghai Jiao Tong University, Shanghai, 1992.
- [2] M. Ayub, A. Rasheed and T. Hayat, "Exact flow of a third grade fluid past a porous plate using homotopy analysis method", *Int. J. Eng. Sci.* 41 pp. 2091-2103, 2003.
- [3] T. Hayat, M. Khan and M. Ayub, "On the explicit analytic solutions of an Oldroyd 6-constant fluid", *Int. J. Eng. Sci.* 42 pp.123-135,2004.
- [4] T. Hayat, M. Khan and S. Asghar, "Homotopy analysis of MHD flows of an Oldroyd 8-constant fluid", *Acta Mech.* 168, pp. 213-232,2004.
- [5] T. Hayat, M. Khan and S. Asghar, "Magnetohydrodynamic flow of an Oldroyd 6-constant fluid", *Appl. Math. Comput.* 155, pp.417-425, 2004.
- [6] S. Das, "A numerical solution of the vibration equation using modified decomposition method", *Journal of Sound and Vibration* 320 pp.576-583,2009.
- [7] S. J. Liao, "Series solutions of unsteady boundary layer flows over a stretching flat plate", *Stud. Appl. Math.* 117, pp. 239-264,2006.

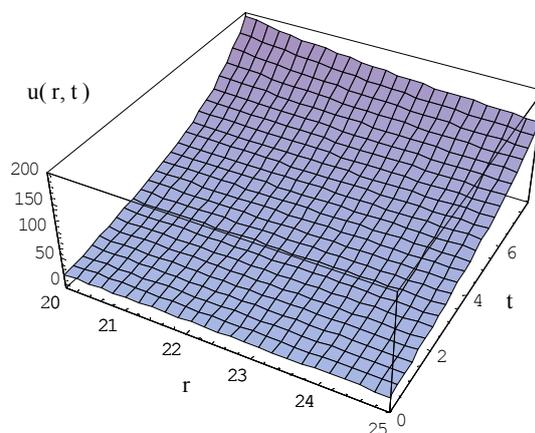


Fig. 1. Plot of $u(r, t)$ with respect to r and t at $c = 6$

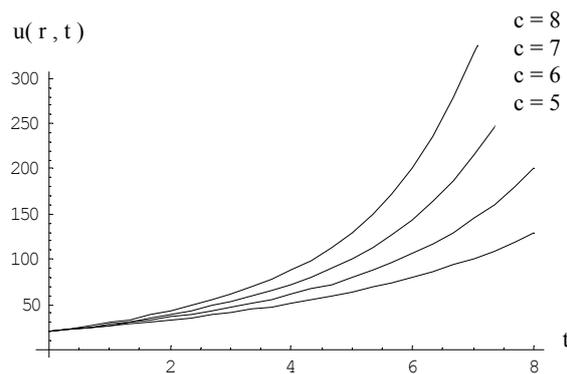


Fig. 2. Plot of $u(r, t)$ vs. t for different values of c at $r = 20$

Transportation and Logistics

Organizing Committee

- Adnan YASSINE
University of Le Havre, France
adnan.yassine@univ-lehavre.fr

Description

Authors are invited to present original work relevant to logistics, including but not limited to:

- Combinatorial Optimization, Operational Research;
- Evolutionary Algorithms, Methods Heuristics;
- Optimization of the global supply chain (organization, Traceability, Security, ...);
- Information Technology;
- Multi-agents System, Self-organization;
- Problems of transport.

Contents

Conception and methodologies to delimitate hinterlands and an application in a container terminal	
Nélio D. PIZZOLATO , Luiz F. SCAVARDA , Rodrigo PAIVA	273
Normative location of neighborhood's shopping centers: a case study in the district of Barra da Tijuca, Rio de Janeiro, Brazil	
Márcio ROZENTAL , Nélio D. PIZZOLATO	281
Dynamical handling of straddle carriers activities on a container terminal in uncertain environment - a swarm intelligence approach -	
Stefan BALEV , Frédéric GUINAND , Gaëtan LESAUVAGE , Damien OLIVIER	290
A multiagent urban traffic simulation Part I: dealing with the ordinary	
Pierrick TRANOUEZ , Patrice LANGLOIS , Éric DAUDÉ	297
Space-time self-organization for the dynamic VRPTW	
Besma ZEDDINI , Mahdi ZARGAYOUNA , Adnan YASSINE , Moncef TEMANI	302
French ports reform: reasons and perspectives	
Sidi Mohamed OULD MOHAMED MOCTAR	308
Industrial dynamics as applied to raw material supply process modeling: the case of a small company from the building industry	
Charles Henri FREDOUET	321

Conception and Methodologies to Delimitate Hinterlands and an Application in a Container Terminal

Nélio D. Pizzolato, Luiz F. Scavarda and Rodrigo Paiva, Industrial Engineering Department, Pontificia Universidade Católica de Rio de Janeiro – Rua Marquês de São Vicente, 225 sala 950L – 22453-900, Rio de Janeiro - Brazil Email: lf.scavarda@puc-rio.br, ndp@puc-rio.br and rtp@hotmail.com

Abstract— The process of containerization and the development of intermodal transportation networks have caused increased competition in the port sector and have interfered in areas of influence of ports (hinterlands) worldwide. These hinterlands are no longer captive to a particular port to be of multiple influence of two or more ports. This paper presents a study whose objective is to characterize the port dynamics in the light of containerization, to map the discussion about the concept of hinterlands and to present methodologies for its delimitation. This conceptual and methodological contextualization is illustrated with an application to the case of the container terminal of the port of Rio Grande (Brazil) with support of Geographic Information Systems (GIS). It follows that the delimitation of hinterlands is essential to allow proper planning of port development by the Port Authorities and Private Terminals in a world of increasing competition.

Index Terms— Hinterland, ports, maritime transportation, Geographic Information Systems

I. INTRODUCTION

The introduction and development of the container as a means of unitization and efficiency gain in cargo movement has added an additional challenge for the planning of the maritime transportation activity, which dramatically changed the operation both of shipping companies as of ports. For the former, the main consequence was the concentration of the maritime transportation activity in the hands of few companies in view of the massive need of capital for investments in ships larger and larger and more modern as a means of reducing the transport unit cost, and for the latter, the challenges is the proper offer of infrastructure and efficient services, at low cost.

Additionally, ports should implement a strategic development planning in order to meet the challenges of a complex and changing competitive environment in function of the current accelerated technological, economic, political and social development. Also, the precise knowledge of its market, the service levels required by its customers, and the business opportunities that can leverage its cargo movement are also necessary, as well as the knowledge of the relative advantages of its direct competitors, in order to anticipate actions and ensure the expansion of the port activity.

In this context, the objective of this work is to characterize the port dynamics in the light of containerization, in the mapping and discussion about the concept of area of influence, internationally known as hinterland, and in the analysis of methodologies for the delimitation of hinterlands. The paper also includes a case study to illustrate the delimitation of the hinterland in the container terminal at the Brazilian Port of Rio Grande with the support of GIS.

The research method related to conception and methodologies to delimitate hinterlands included a literature review concerning the issue and the performance of in situ unstructured interviews in two consulting companies that have developed projects related to ports in Brazil in which the issue of hinterlands was strategically discussed. The elaboration of the case included a database provided by the Brazilian National Center for Transatlantic Navigation (CNNT) related to the movement of containers in municipalities from the state of Rio Grande do Sul (south of Brazil) and with a geo-referenced graphical basis of this state obtained from the Brazilian Institute of Geography and Statistics (IBGE) for application in the Geographic Information Systems (GIS) ArcView software (version 3.3).

The article is divided into 5 sections, being the first one introductory. Following, Section II presents the concept of hinterland, Section III provides a survey of proposed methodologies for its determination, Section IV presents the case study, and Section V summarizes the conclusions of the work.

II. HINTERLANDS

The concept of hinterland is known since the beginning of the twentieth century and has been discussed both in the academic literature and by market professionals, but there is no standardized definition for the concept. Among the definitions offered in the academic literature, the following should be emphasized: Bird (1971), Takel (1976), Kesic et al. (1998), Van Klink and Winden (1998), Van Klink and Van den Berg (1998), Van Cleef, apud Amoyaw (1999), Degrassi (2001), Yap et al. (2006) and Tan (2007). Analyzing these definitions, it could be observed that, over time, due to various historical and technological contexts, the definitions for hinterland have been improved.

This article adopts the definition used in Slack (1993). According to this author, the area of influence of a port is the port's market area in land from which or to which the port ships and receives cargoes. It is an area where the port sells its services and interacts with its customers. The area of influence can be represented as a market share that a port has in relation to the other ports that serve the same geographical region. This vision links the port as a center that offers logistics services. Slack (1993) defines hinterland as the effective market of a port, or the geo-economic space in which it sells its services and interacts with its customers. This definition seeks to treat the port as a logistic link within a competitive market, where it should act commercially and technologically in a pro-active way to ensure its market. Thus, the hinterland is treated as something dynamic and changing, which may be disputed by various ports as they can offer their services in a competitive way in market.

The concept of hinterland has faced criticism in the academic world particularly in relation to two aspects: the growing containerization of general cargo and the development of intermodal transportation networks. McCalla (1999), Haezendonck and Notteboom (2002) and Lacerda (2004), for example, argue that the intensification of the use of containers has increased the geographic performance for the capture of cargoes by ports, making the understanding of traditional definitions of hinterlands difficult. Song (2003) and Lacerda (2004) also mention that intermodality allows containers to travel increasing distances. This facilitates the decentralization of the packaging of cargoes in containers, which could be performed at the origin of goods in factories, or through specialized services, away from areas of ports, thus expanding their hinterlands. Lacerda (2004) also emphasizes the increase of transshipment related to the increased size of container ships, since shipping companies seek to focus cargoes in a main port of each region (hub) in order to obtain large quantities of cargo to transport in the long routes on larger ships. The larger the ships are, the lower the costs of transporting the containers will be, which means that the activity presents economies of scale. In this context, the concept of area of influence of a port has become more complex (Fageda, 2000). Different ports may share the same hinterland, whose borders now depend also on the development of intermodal transportation corridors and not exclusively on the port market. This allows a direct competition between ports away from each other. Additionally, competition between ports is not limited to the increase in their area of influence traditionally defined, but also on their role in transshipment, i.e., to attract movement whose origin or destination is not the port itself or its surrounding area.

Despite the criticism, the concept of hinterland, whether in its traditional sense or in its contemporary sense, remains critical to the port development (Amoyaw, 1999). The delimitation of the port hinterland and the analysis of its nature and extent are basic steps in assessing the development of a port or port terminal.

The definition of hinterland through the static view

characterizes it as a continuous area behind the port. With the development of the concept, functional relations were inserted between the maritime area and its functional locations outside the port such as industrial areas and logistics centers. According to Hayuth (1982) the dynamic concept of a port hinterland makes it possible to identify factors that influence the expansion and reduction of the size of the port hinterland. The possibility of modifying the port hinterland over time presupposes the existence of variables that influence its delimitation. The main factors in this delimitation are:

- The nature of goods: general cargo, containers, bulk, etc. (Takel, 1976; Kesic et al., 1998; Degrassi, 2001);
- The structure of maritime shipping in relation to the types of ships, to the organization of ship owners and to the frequency of scale of ships (Van Klink and Winden, 1998, McCalla, 1999; Degrassi, 2001);
- The structure of the transportation network in land linking the port and its market and consequently the transportation cost associated with this network (Takel, 1976, Slack, 1993; Kesic et al., 1998, McCalla, 1999; Hoyle, 2000; Degrassi, 2001);
- The influence of economic policies (Kesic et al., 1998; Degrassi, 2001);
- The entry of new port competitors in the market (Van Klink and Winden, 1998);
- The labor organization in the port activities (Takel, 1976; Kesic et al., 1998);
- The port infrastructure (Takel, 1976; Kesic et al., 1998);
- The port competitiveness in terms of appropriate fees when compared to other ports (Takel, 1976; Kesic et al., 1998);
- Natural causes such as climate and seasonal factors (Degrassi, 2001);
- Factors and political events (Degrassi, 2001).

III. METHODOLOGIES FOR THE DELIMITATION OF HINTERLANDS

The need to delimit the port hinterland and its analysis to allow the proper port planning, either by port authorities or by port terminals, brings to light the question of how areas of influence of ports should be defined. This section presents several methods proposed for the definition of port hinterlands, identified both in literature and in real cases, and discusses practical situations in which they were conducted. These methods are classified into two groups according to their purposes:

- Methodologies for identifying the current hinterland of a port; and
- Methodologies for defining the potential hinterland of a port.

Both groups of methods can be used in an independent way, depending on the objectives to be achieved. However, the ideal situation is their use together to perform analysis and comparisons between the areas of real and potential influence of a port. Certainly, what the administration of a port should expect is that its potential area of action is actually achieved. The next sub-sections present methodologies for each one of the two

groups offered by the academic literature, specialized literature and two consulting companies, which have implemented methodologies to real cases.

A. Methodologies for identifying the current hinterland of a port

The methodologies for the identification of the port actual hinterland perform an analysis of its effective market. These types of methodologies can be used only for terminals in operation, once it is impossible to calculate the actual hinterland in case there are no historical movement data.

One of the first methodologies in literature for this purpose was proposed by Morgan in 1948, apud Degrassi (2001). This methodology defines the area of influence of a port from the analysis of the movement of cargo in a particular geographical region to / from a particular port. This identification establishes the classification of this region based on its current movements, seeking to classify municipalities according to an ABC curve of importance of participation. Thus, the identified areas can be classified on three levels: primary Hinterland (large movement), secondary Hinterland (intermediate movement) and marginal Hinterland (small movement). It is worth mentioning that the literature offers other classifications for hinterlands, and the contributions of Kesic et al. (1998), Amoyaw (1999), Elliot, apud Degrassi (2001) and Degrassi (2001) are the most important.

The second methodology, proposed by McCalla (1999), consists of the analysis of statistics of the movement of containers in the region of interest. This methodology proposes the collection of data featured in the cargo bills. From these data, the author calculates the market share of each port from the region analyzed in relation to the cities of origin and destination on land and in relation to the foreland region, to where the containers moved by ports were designated or from where they were received, which allows plotting the regions in a map, delimiting the actual ports hinterlands.

Based on the methodology of McCalla, a market survey was conducted by a consulting company to determine the area of influence of a terminal of a port operator in Brazil. To this end, data corresponding to the movement of containers were used and the municipalities of origin and destination of containers moved for exports and imports were surveyed. Then, the participation of the port terminal in the movement of containers of each municipality was calculated, which allowed the identification of three classifications for the municipalities analyzed according to three levels of activity, namely:

- The municipality is area of influence of a single port if at least 70% of its cargo is moved to / from the port;
- The municipality is area of influence of two ports if 80% of its cargo is moved to / from two ports;
- The municipality is area of influence of three ports if 90% of its cargo is moved to / from three ports.

From the determination of the port hinterland, a planning to attract cargoes from municipalities where there is the occurrence of other ports was elaborated.

In another similar research, Neto and Santos (2005)

conducted a study to attract cargo to the Port of Santos (Southeast of Brazil). The main objective of the study was to conduct a detailed and reflected analysis of the main products moved by the port and to define the area of influence for each product of identified commercial interest. In this study, several methodologies for the delimitation of the port of Santos hinterland were proposed. The first sought to establish the importance, from the economic point of view, of products exported and imported through the port of Santos, which represented about 60% of all products exported / imported by this port. The next step was to determine the destination or origin, as the case, of each product per municipality and per unit of the federation. Thus, this methodology allowed a preliminary definition of the area of influence of the port. The study suggested the value of five million dollars as the minimum value of each unit of the federation to be considered as belonging to the area of influence of the port of Santos. The application of this criterion resulted in the area of influence of the port covering 16 states of Brazil of the total of 27 federal units.

Since this first proposal presented a disability, when considering only absolute values as a criterion for defining the area of influence, Neto and Santos (2005) incorporated in the second proposal, all products exported and imported through the port of Santos, and exports and imports per municipality, which allowed the aggregation per state of the country. Combining the results, per state, exports and imports, the values were generated in U.S. dollars from the international trade in each state. Then, participation data from the port of Santos in the international trade of each state were generated. It was defined that, if 10% of the international trade of each state used the port of Santos, this would show that the port has an impact on the economy of the state. For this methodology, the port of Santos hinterland was considered for eight Brazilian states.

The third methodology proposed, called by Neto and Santos (2005) as a hybrid methodology, considers not only the importance of the port in the economy of the states, but also the relevance in terms of monetary value in the trade balance due to the port. In this case, the cutting criterion included the units of the federation that moved, through that port, values greater than or equal to one hundred million dollars in the year under review. Within the area of intersection of the two previous criteria, five states were identified. These states were classified as the primary hinterland of Santos. The study also conceptualized the secondary port hinterland, formed by five states that moved through Santos values exceeding US\$ 100 million, but that did not meet the previous criterion, i.e., the participation of the port of Santos in the international trade of the state higher than 10%.

As critical evaluation, it could be said that the methodology proposed by McCalla (1999) shows similarity to that proposed by Morgan, apud Degrassi (2001), since it uses real movement data for the delimitation of areas of influence of ports. However, McCalla does not consider the amount moved as a key factor in defining which region includes the port hinterland. The methodology proposed can be considered an improvement of that of Morgan, in which the author highlights the need to

analyze the market share of these ports, both on land and in relation to its foreland.

On the other hand, the first methodological proposal developed by Neto and Santos (2005) adopts the absolute value of goods as a delimiting parameter, which consists of an innovation in relation to any other methodology examined. The second proposal considers the importance of the port to the international trade in the region of origin / destination of goods. The third proposal of these authors is the adoption of the two parameters presented in the first two proposals, with the aim of not only inserting a region important for the port in terms of trade in the area of influence, but also to ensure that the port is important in relative terms to the international trade of a particular region.

B. Methodologies for defining the potential Hinterland of a port

There is another set of methodologies primarily aimed at performing a prospective analysis of the potential port market, in other words, they calculate the ideal delimitation of its area of influence using technical, economic, political and social variables. Their results indicate the potential hinterland of a port or port terminal, where the geographic area identified cannot be the same as that of its operation.

Van Klink and Winden (1998) propose a methodology for delimitating the potential area of influence of a port based on the monetary transportation costs. A geographical region will be the port hinterland if the transportation costs for this port are lower than for any other port. If the transportation costs between the region and two or more ports are equal, this would be a region of multiple influence of ports, since it could move its cargo by any of the ports.

A work conducted by the second consultancy company was aimed at studying the capture of cargo from a new port terminal specialized in the movement of containers in the southeastern region of Brazil used the following steps of the methodology proposed by Van Klink and Winden (1998): i) identify the main customers of the ports closest to the new terminal, within a radius of 500 km, with the municipalities of origin / destination of containers moved by these ports, ii) to budget the transportation cost between ports and municipalities of origin / destination of cargoes and also between origins / destinations identified and the new terminal, iii) to compare the transportation costs to identify the locations where the costs up to the new terminal would be more competitive, then marked as potential markets for the terminal, while the localities with similar costs were identified as areas of competition and those with higher costs were discarded, and iv) from this identification, to draw a commercial planning for the attraction of cargoes for the new port.

In another work of similar nature, the Secretariat of Infrastructure of the Brazilian State of Bahia developed a study published in SEINFRA (2004) that, when considering the impact of logistic alternatives in transportation costs in the state, used a methodology to identify the area of influence of the ports of the state. The impact of logistic alternatives has been

previously analyzed based on the comparison of transport costs in the current situation and in the hypothetical situation of supply of alternatives (such as the use of the São Francisco River waterway, improvements to the existing railroad, among others). The area of influence of each port was estimated based on the criterion of minimum cost, i.e., a municipality encompasses the area of influence of a port if the transportation cost to that port is smaller than for any other port. For both alternatives, isocost lines were drawn, and when superposed, show the delimitation of areas of influence of ports from the state of Bahia.

IV. CASE STUDY

The purpose of this section is to illustrate the importance of working with the concept of hinterlands for a port planning, either in regard to the identification of the port market (or port terminal), or in regard to the identification of the market potential of this port. This goal is achieved through an analysis of a real case that consists of the delimitation of the market of the port of Rio Grande in the exportation of containers from municipalities belonging to the state of Rio Grande do Sul. The current market of this port and its potential market are both analyzed, as well as the delimitation of its primary and secondary hinterlands through the analysis of participation of this port in the market of municipalities from the state of Rio Grande do Sul compared to other ports in southern Brazil. In this sense, it will be used the premise that the exportation of container belonging to municipalities from the state of Rio Grande do Sul by Brazilian ports other than ports of Rio Grande, Sao Francisco do Sul, Itajaí and Paranaguá are irrelevant. The consistency of the premise is that the ports above are the most representative in the southern region of Brazil to work with general containerized cargo. Figure 1 presents a map of the region.



Fig. 1. Main ports in the south of Brazil.

To achieve the goal, the following were necessary:

- i) methodologies for the delimitation of the port hinterlands;
- ii) a database for the southern region from the Brazilian National Center for Transatlantic Navigation (CNNT) with the movement of containers from the port studied and its competitors, which allowed identifying the municipality of origin, and,

iii) a geo-referenced graphic base of the region studied obtained from the Brazilian Institute of Geography and Statistics (IBGE, 2005), for application in the Geographic Information Systems (GIS) ArcView software.

The case study reported below is divided into three distinct stages. The first refers to the survey of cargoes exported in containers by municipalities from the state of Rio Grande do Sul and the identification of the market of the port of Rio Grande through the analysis of the amounts moved by using the methodology of Morgan, apud Degrassi (2001).

The second stage identifies the market share of the port of Rio Grande, with the methodology proposed by McCalla (1999), establishing the primary and secondary port hinterlands in function of the relative market share of this port in these municipalities.

In the third stage, which corresponds to the desired identification of the potential hinterland of the Port of Rio Grande, it was chosen for developing a new methodology able to identify municipalities potentially important in terms of attracting cargoes; however, without the need to calculate freights, whose data are not available, but using real movement data from municipalities and port, for confrontation and delimitation of the municipalities to be subject to a more intense commercial activity.

A. Exportation of containers of municipalities from the state of Rio Grande do Sul through the port of Rio Grande

The aim was firstly to check the exportation of containers, via one or more of the four container terminals of ports from Southern Brazil (Rio Grande, Sao Francisco do Sul, Itajaí and Paranaguá), with origin in municipalities belonging to the State of Rio Grande do Sul. This step uses the methodology proposed by Morgan, apud Degrassi (2001). In this sense, database from CNNT (CNNT / Datmar, 2006) was searched to check the movement of containers for exportation, with origin in municipalities from this state, via the four ports in the southern region for 2003. The year is not considered very relevant to the research, since the prime objective is to illustrate the importance of the study that delimitates hinterlands and not the current issue of the terminal itself. The variables searched and separated were:

- Direction of the traffic: exportation;
- Name of the municipality of origin, located in the state of Rio Grande do Sul of cargoes to each of the four ports in the southern region;
- Port (terminal) of destination: Rio Grande, Sao Francisco do Sul, Itajaí and Paranaguá;
- Total amount moved, measured in capacity of Twenty Equivalent Unit (TEU).

From the data obtained in the database query, the municipalities were listed in decreasing order of movement and prioritized into three categories using the ABC classification, according to data presented in Table 1.

The number of municipalities that moved containers in 2003 through one or more of the four ports in southern Brazil was 135, totaling an export movement of 137,388 TEUs. Table 2

TABLE I
CLASSIFICATION OF EXPORTING MUNICIPALITIES

Classification	Volume in TEUs
Large Movement	> 4.961 TEUs
Intermediate Movement	641 – 4.960 TEUs
Small Movement	1 – 640 TEUs

shows this number according to port with the respective percentages.

TABLE II
EXPORTATION OF CONTAINERS FROM THE STATE OF RIO GRANDE DO SUL THROUGH PORTS (CONTAINER TERMINALS) FROM THE SOUTHERN REGION OF THE COUNTRY

Port	No. in TEUs	
Rio Grande	114.479	83.3%
São Francisco do Sul	6.806	5.0%
Itajaí	14.713	10.7%
Paranaguá	1.390	1.0%
Total	137.388	100%

Source: adapted from the CNNT/Datamar database (2006)

Of these 135 municipalities, ten can be classified as of large movement and only ten as of intermediate movement. Figure 2 presents the results presented in GIS (Arcview) for 2003.

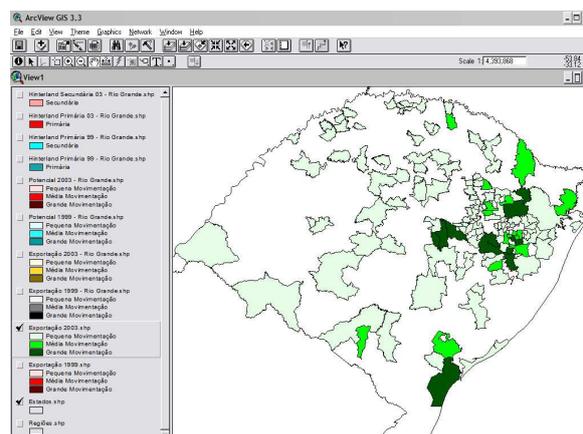


Fig. 2. Movement of containers from municipalities from the state of Rio Grande do Sul through the four ports in the southern region..

The next step was to identify municipalities from that state using the port of Rio Grande in the exportation of containers in order to delimit the market of this port, using methodology proposed by Morgan, apud Degrassi (2001), and the same classification of municipalities as of large, intermediate and small movement presented in Table 1.

The analysis shows that from the 135 municipalities in the state of Rio Grande do Sul using the ports from southern Brazil in 2003 for the exportation of containers, only 7 have not used the Port of Rio Grande.

Only seven municipalities in the state of Rio Grande do Sul that moved containers through the port of Rio Grande were classified as of large movement and eighteen other municipalities were categorized as of intermediate movement, as shown in Figure 3.

B. Port of Rio Grande actual Hinterland

This step aims to delimit the area of influence of the Port of

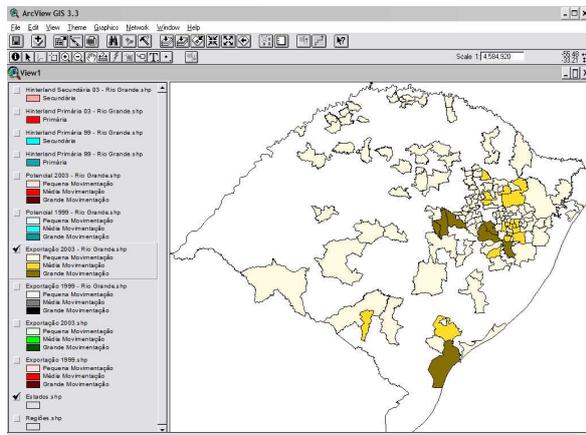


Fig. 3. Municipalities in the state of Rio Grande do Sul with movement of containers through the Port of Rio Grande.

Rio Grande through the analysis of its market share. The elaboration of this phase used methodology proposed by McCalla (1999), using container movement data from these municipalities and calculating the market share of the port of Rio Grande for each of these municipalities. This survey allowed the geo-referenced identification of municipalities that compose the port hinterland and the primary and secondary hinterlands, according to classification proposed by Degraasi (2001), based on the following criteria:

- If the Port of Rio Grande moves 70% or more of the total containers moved by a given municipality in the southern ports, this municipality is its Primary Hinterland;
- If the Port of Rio Grande moves between 40% and 70% of all containers moved by a given municipality in the southern ports, this municipality is its Secondary Hinterland, which can also be understood as area of multiple influence, as described in Slack (1993) and Kesic et al. (1998).

Since all classifications and identifications of municipalities that compose the primary and secondary hinterlands of the port of Rio Grande, as well as the compatibility of databases have been conducted, these areas of influence were plotted in the ArcView software in a geo-referenced way, as shown in Figure 4.

Most municipalities in the state of Rio Grande do Sul that moved cargoes through the port of Rio Grande in 2003 were defined as primary port hinterland, with only 16 municipalities defined as secondary hinterland. However, the municipalities of Porto Alegre (State capital and most important city) is among these 16 municipalities, which in spite of moving much of its cargo through the port of Rio Grande, still has high movement through other ports in the southern region.

The delimitation of the port of Rio Grande hinterland through the study of the market share in the municipalities belonging to the state of Rio Grande do Sul shows the consolidated position of the port in relation to this market. As shown in Table 2, the port moved over 80% of containers exported by these municipalities. Thus, the study confirms this diagnosis at this stage, as the analysis of each municipality alone

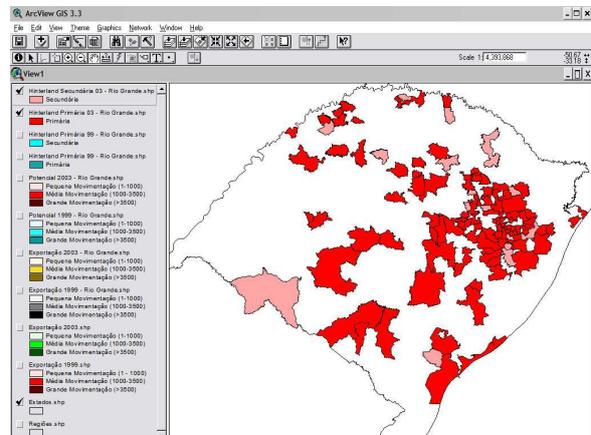


Fig. 4. Primary and Secondary Hinterland of the Porto of Rio Grande.

shows that most municipalities comprises the area of primary influence of the port, exporting more than 70% of their containers through the port of Rio Grande.

C. Potential hinterland of the port of Rio Grande

The objective of this stage is to find out which municipalities are characterized as potential market of the port of Rio Grande, i.e., which municipalities have containers not yet moved through this port and their commercial importance.

Since the methodologies studied for the delimitation of the potential hinterland of a port use the transportation cost between the place of origin of the cargo and port as main variable for analysis, and since this variable is not available in databases obtained, a new methodology for the performance of this stage of the study was elaborated. This third stage of the case study is characterized as a complementary analysis to that performed in stage 1, since to find municipalities with potential movement in the port of Rio Grande, the total movement of containers of these municipalities and also the total cargo of these municipalities already absorbed in the port of Rio Grande were analyzed. Based on this information, the potential market of the port in the state of Rio Grande do Sul could be obtained. The operation performed for defining the potential movement of containers in the port of Rio Grande is shown in Figure 5

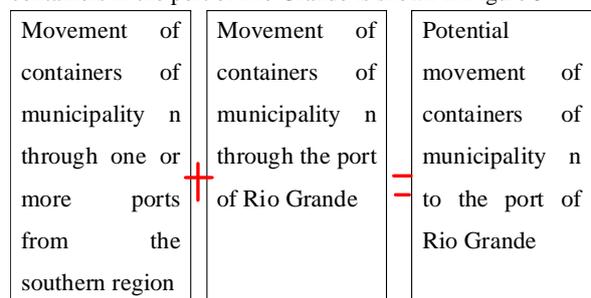


Fig. 5. Definition of the Potential Movement.

This operation allows verifying graphically which municipalities located in the State of Rio Grande do Sul that move containers through the four ports in southern Brazil are

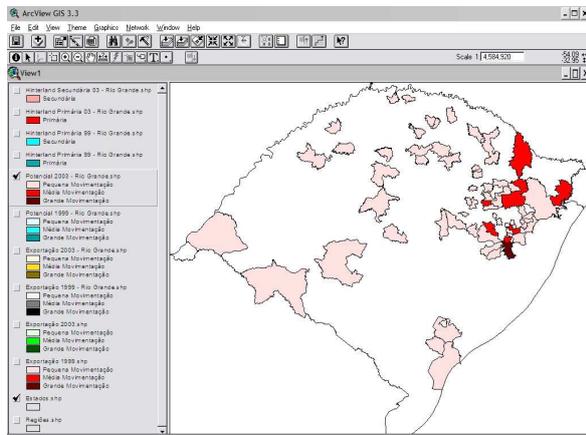


Fig. 6. Municipalities with good movement potential to the Port of Rio Grande.

characterized as a potential market of the Port of Rio Grande.

Once municipalities with potential cargo to the Port of Rio Grande were identified, they were classified into municipalities of large, intermediate and small potential, based on the same criteria in Table 2.

With support from the ArcView software, Figure 6 characterizes municipalities as potential markets of the Port of Rio Grande for 2003.

The analysis of the potential Port of Rio Grande hinterland shows that the municipality to be given more commercial attention is Porto Alegre, since, although already moving cargoes through Rio Grande, it also offers great potential for this port, being the only municipality of great potential in the state, with a possible amount of additional attraction of 9,360 TEUs. Its cargoes through other ports in the southern region of the country: 5,877 TEUs for Itajaí, 3,380 TEUs for São Francisco do Sul and 102 TEUs for Paranaguá.

Yet, due to the high participation of the Port of Rio Grande in the exportation of containers from these municipalities, the potential to attract new cargoes to this port is very limited. However, it should be given special attention to municipalities of large and intermediate potential, and to municipalities from neighboring states.

V. FINAL REMARKS AND CONCLUSIONS

The competitive environment in which ports are contextualized requires the proper planning of port development to provide appropriate infrastructure and services to the current reality. This reality is the result of the use of containers as a transportation tool and the development of intermodal transportation networks, especially in developed countries. These facts, combined with the growing market of shipping firms worldwide and to the transportation logistics of containers, marked by the use of megacarriers that arrive in a few hub ports, cargo hubs, fed by small-size feeder lines, significantly decreased the position of the port as cargo monopolist operator of a particular region. In this context, an important tool for the strategic planning of a port is the

identification of its area of influence, which process also includes the evaluation of the analogous area of influence of its main port competitors.

The present study of methodologies for the delimitation of port hinterlands shows that, despite the discussion in literature on this subject is still scarce; ports and consulting firms in the sector have used this concept and developed new methodologies for delimitation and knowledge, which allows port authorities and port terminals to elaborate port development plans. The present study identified several methodologies for this delimitation, which were classified into methodologies aimed at identifying the current hinterland for a port, and methodologies aimed at delimiting the potential hinterland for a port.

The case conducted at the Container Terminal of the Port of Rio Grande analyzed, for the State of Rio Grande do Sul, the market of this port terminal through a combination of two methodologies found in literature (McCalla, 1999, and Morgan, apud Degrossi, 2001) and with the elaboration of a new one by the authors of this article. The delimitation of hinterlands was visualized in geo-referenced maps, elaborated by means of the use of the ArcView software. Based on this case, it was found that areas of multiple influence are not yet a reality, and the port terminal still has the domain of the container exportation market in the state of Rio Grande do Sul, with market share exceeding 80%. Probably, the analogous analysis of other Brazilian ports in the south such as Itajaí and São Francisco do Sul, and even the port of Santos, could reveal results different from those found for the port of Rio Grande and offer guidelines for their planning and competitive development, showing the importance of the study of areas of multiple influence for these ports. The determination of the terminal hinterland as an analysis tool of its market revealed to be important to identify its area of operation and expansion, as well as its cargo attraction potential, allowing its planning to meet the competition of new markets and mainly for the maintenance of its customers, as a form of sustainable development of the port terminal.

ACKNOWLEDGMENT

This research was supported by the National Council for Technological and Scientific Development, CNPq, Brazil.

REFERENCES

- [1] AMOYAW, B. K. B. The Hinterland Concept and Port Development in Ghana, 1999. 188 f. Dissertação de Mestrado, Faculty of Graduate Studies of the University of Manitoba, Manitoba, 1999.
- [2] BIRD, J. H. Seaports and seaport terminals. London: Hutchinson & Co., 1971.
- [3] CNNT / DATAMAR – Centro Nacional de Navegação Transatlântica. Banco de Dados Movimentação de Contêineres da Região Sul do Brasil, 2006.
- [4] DEGRASSI, S. The Seaport Network Hamburg, 2001. 387 f. Tese de doutorado, Universität Hamburg, Hamburgo, 2001.
- [5] FAGEDA, X. Load Centres in the Mediterranean Port Range, Ports Hub and Ports Gateway. University of Barcelona, Barcelona, 2000.
- [6] HAEZENDONCK, E., NOTTEBOOM, T. The competitive advantage of seaports. In: Huybrechts, M., Meersman, H., Van de Voorde, E., Van Hooydonk, E., Verbeke, A., Winkelmans, W. (Eds), Port Competitiveness, An Economic and Legal Analysis of the Factors

- Determining the Competitiveness of Seaports, De Boeck, Antwerp, pp. 67–87, 2002.
- [7] HAYUTH, Y. Inter-modal Transportation and the Hinterland Concept. IN: TEG - Tijdschrift voor Economische en Sociale Geografie, 1982, p 13-21.
- [8] HOYLE, B. Journal for Maritime Research. ISSN 1469-1957, 2000.
- [9] IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Base de dados municipal, Rio de Janeiro, 2005.
- [10] KESIC, B.; KOMADINA, P.; CISIC, D. Towards the Unified Theory of the Port Gravitational areas Introducing Information Centrality Factors. University of Rijeka, Croatia, 1998.
- [11] LACERDA, S. M. Navegação e Portos no Transporte de Contêineres. IN: Revista do BNDES, Rio de Janeiro, v. 11, n. 22, p 215 – 243, 2004.
- [12] McCALLA, R. J. Global change, local pain: intermodal seaport terminals and their service areas. IN: Journal of Transport Geography, v.7, p. 247-254, 1999.
- [13] NETO, C. A. S. C.; SANTOS, M. B. Atração de Cargas para o Porto de Santos: perspectiva e crescimento sustentável. Brasília: IPEA, 2005.
- [14] SEINFRA – Secretaria de Infra-Estrutura do Governo do Estado da Bahia. Programa Estadual de Logística de Transportes: caminhos para o desenvolvimento. Salvador, 2004. 159 p.
- [15] SLACK, B. Pawns in the game: ports in a global transportation system. Growth and Change, v. 24, p.379-388, 1993.
- [16] SONG, D.W. Port co-operation in concept and practice. Maritime Policy and Management, v. 30, n. 1, 29–44, 2003.
- [17] TAKEL, R. E. Port and Hinterland Relationships. IN: UNCTAD, Manual on Port Management, v. 2, p. 67-73, 1976.
- [18] TAN, T-Y. Port cities and hinterlands: A comparative study of Singapore and Calcutta. Political Geography, v. 26, n. 7, p. 851-865, 2007.
- [19] VAN KLINK, H. A.; WINDEN, W. Towards a new hinterland orientation for Rotterdam: the entrepreneurial port. 38th Congress of the European Regional Science Association, 1998.
- [20] VAN KLINK, H.A., VAN DEN BERG, G.C. Gateways and intermodalism. Journal of Transport Geography 6, 1–9, 1998.
- [21] YAP, W.Y., LAM, J.S.L. Competition dynamics between container ports in East Asia. Transportation Research Part A: Policy and Practice, v. 40, n. 1, p. 35-51, 2006.

Normative location of Neighborhood's Shopping Centers: A case study in the district of Barra da Tijuca, Rio de Janeiro, Brazil

Márcio Rozental and Nélio Domingues Pizzolato, Industrial Engineering Department, Pontifical Catholic University of Rio de Janeiro. Rua Marquês de São Vicente, 225 sala 950L, Brasil Email: Rozental@uol.com.br and ndp@puc-rio.br.

Abstract—The Shopping Center industry has recently shown an impressive but disordered development, creating in some cities overlapping problems relative to the population distribution, since the profile of these ventures has been largely disregarded. Several public facility location studies are currently used mostly directed to planning health services and the basic education network. Predominantly, these studies use the p-median model, associated to exact and heuristic methods for solution, but no similar study for locating shopping centers has been identified in the literature. This article examines the various kinds of shopping centers and studies the location of those called Neighborhood Shoppings as applied to the district of Barra da Tijuca in Rio de Janeiro, Brazil. The study uses local census tract data, to accumulate the population in discrete points, but the geography of the region and the condominium style of occupation has suggested the use of a simpler heuristics that remembers the Voronoi diagrams. The study proposes six neighborhood shopping centers and suggests the points for their location.

Index Terms—Location, Shopping Center, p-median

I. INTRODUCTION

Location is a theme of permanent practical and academic interest that permeates several sciences, such as mathematics, economy, engineering, operational research and so on. Specific studies of location have always been in the interests of human beings as a strategy to preserve safety and welfare. Any historical survey shall identify Archimedes as the first inventor revered by western history. Among his brilliant inventions, the defense of Syracuse from invading Roman ships have required accurate location studies for displaying mechanical levers and mirrors for concentrating solar rays on those moving ships.

Certainly, the old Greek geometers, such as Euclides, pioneered studies for determining centers of gravities and for identifying properties of physical order. However, Drezner et al (2002) affirm that precursory studies about identifying points in the plan endowed with certain properties are usually credit to Fermat (1601-1665) and, with almost equal frequency, to Torricelli (1608-1647).

Concretely, the location analysis applied to industrial

activities, using measures of utility as production or costs, is regularly attributed to Weber (1909) who, looking for minimizing transport costs and reducing costs for the consumer, studied the positioning of an industrial installation, with several inputs and a single marketing consumer.

According to ReVelle and Eiselt (2005), the location analysis refers to the modeling, formulation and solution of a class of problems which could be described as positioning facilities in a space. Based on our experience, studies on the theme are usually made under three perspectives: practical, applied, and academic studies.

Practical studies concern the location of a specific industrial or commercial activity. In this perspective, the factors that should be considered in a preliminary evaluation include: materials availability and location; distance and dimensions of the marketing consumer; means of transport; land availability; climate; pluvial standards, infrastructure, topographical data; power, water and sewerage systems; labor availability and its costs, life conditions, laws and regulations, tax structures, etc. Certainly, a mistaken decision would cause an unrecoverable cost. Evaluations of that nature constitute a favorable knowledge area for specialized consultants whose knowhow is not often made public.

Applied studies would be those in which the complexity of the inter-relationships are of such order that the development of explicit modeling is recommended. The model would emphasize the most relevant elements and ignore a mass of other pertinent information by using what are called simplifying hypotheses. The produced model will be useful only if it is capable of pointing out good solutions. Notable examples of applied studies related to locating public services overflow in the modern society, such as schools, maternities, health centers, leisure areas, firemen, ambulances, post offices, water networks, sewer systems, and so on, inspired in some societies' basic needs, as well as, radars, communication networks, distribution centers, oil platforms, etc, inspired in commercial demand or military defense. In fact, the problem object of this work is framed in this domain of the applied studies, and a minimum sample of this may be found in Galvão et al (2002), Pizzolato et al (2004), Pires et al (2004), Monteiro and Pascoal (2005), and Teixeira and Antunes (2008) who examined,

respectively, the location of perinatal units, of public schools, of condensers in electric networks and the location of an hierarchy of facilities, illustrated with an application to school location.

In the third perspective, academic studies would be associated to the development of methods and processes directed to the resolution of complex models, not necessarily concerned with any explicit application. Typically, theoretical studies focus on the modeling of more complex problems, on the quality of the solution, on the processing time, on the development of computer resources, and so on. It is a fundamental area with remarkable relevance, as noted by the number of articles, congresses and journals dedicated to the topic.

Taking management science as its base, or operational research as a more restricted expression, the available literature on location is unlimited, but the foremost prestige stay with the theoretical studies, followed by the applied, and with less prominence, the practical ones. The first methodologies were concerned with the location of a single installation, but, with computer and modeling resources advances, the methodologies became directed to the development of methods for the simultaneous location of several facilities. The immense literature on the subject turns frustrating any attempt to a general literature review, even if labeled as representative. In any case, we would mention ReVelle and Eiselt (2005) and Reese (2005), that present extensive bibliographical reviews. Besides these references, in any search system, the word location will point out an unlimited number of thesis, dissertations, scientific articles, editorials, and books, highlighting the importance of the theme for administration, economy, engineering, computer science, etc.

The present work aims to study the location of Neighborhood's Shopping Centers, which are designed to support shoppers living around limited geographical areas that, according to our investigations, would never have been the object of published research. Certainly, the shopping center, as a general concept, grew in the United States, and its location seems to follow the geography of the highways, while the practical perspective as discussed above, would emphasize customer's needs, family income, accessibility, traffic flow in neighboring highways, etc. However, this intuitive concept applies to locating one single venture which might ignore the existence of several kinds of shopping centers and does not address the question of locating a network of shoppings.

This study is organized as follows: section 2 presents the several types of shopping centers and highlights the importance of the Neighborhood Shopping Centers; section 3 summarizes some of the technical procedures for public facility location; section 4 describes the Barra da Tijuca neighborhood, highlighting its physical, demographic and urban characteristics, and proposes the ideal location based on a heuristic methodology; while section 5 synthesizes the conclusions of the study.

II. SHOPPING CENTERS

A. *Development of Trade*

A possible definition for a shopping center, not fully endorsed by the present work, as justified further on, may be found in the wikipedia: "a shopping center is a business establishment specifically built to cover a center of purchases that presents a diversified market, feeding area, leisure area, parking and high comfort level services, such as: air conditioned area, lifting stairs, elevators, safety, etc. The kind of anchor stores, the amount of stores and the fact of existing owned stores with rented stores also characterize that category."

Going a step back along the history, the early merchants initially used to walk towards the clients or would check their own conveniences to choose his business point without much worries about his neighborhood. With the growth of the cities, however, he noticed the advantages of establishing at certain places which presented a high concentration of activities. Grouped, the merchants could attract more buyers, jointly benefiting consumers and their own trade. As an old example of this relationship are the medieval fairs, that were usually located in large cities of easy access.

The evolution determined permanent constructions in order to shelter the local trade. In the beginning, this grouping of stores was not planned to work in an integrated way, operating disorderly, just differentiated from traditional trade because they were located inside appropriate places. Later, those constructions developed to assume the form of galleries, usually located in the center of the cities, a place that would propitiate commercial growth, forming streets and internal yards that allowed the access to the stores. The galleries, although appropriately designed, still did not offer an organization or planning to promote the operation in an integrated way. (Blay & Sucupira 1962, p.81-95)

As the trade begins to develop in organizational and administrative terms, the great merchants began to build their own commercial facilities in large buildings, introducing the department stores that, in a same place, have a great diversification of products. The consumer begins to have more comfort in their purchases, avoiding long displacements and the visit to many different and distant stores in order to acquire what he wanted.

At the beginning of the twentieth century two scenarios appeared in the capitalist world, specially in the United States, that would be at the origin of the current expansion of the Shopping Centers, namely the intense urbanization and the widespread use of the automobile. The traditional place where the trade was located begins to experience heavy traffic on their streets, due to the urban growth. The progressive deterioration of the urban centers and the consumers' preference for parking their cars in places near to the trade, brought new challenges for assisting the clients. In this way, the purchase centers have appeared, and have developed until the creation of the shopping centers, a place that aims to offer comfort, parking easiness, safety, leisure and several types of services (Hirschfeldt, 1986, p.15-19) .

The precise origin of the shopping centers is certainly controversial, considering the kind of business model that is known today, but the first building with characteristics close to the current ones began in the USA in 1907, by Edward H. Bouton, showing architectural uniformity, centralized administration by the entrepreneur, congregation of stores of different branches and a parking for carriages. In the 1920's appeared another similar enterprise, called the Country Club Plaza in Kansas City, USA, that presented an unified administrative policy and place for parking the automobiles.

In 1931 appears in Dallas, USA, the first enterprise in the same style of the current shopping centers. It was the Highland Park Village that, in agreement with the Urban Land Institute (ULI), became a standard for shopping centers for America. That enterprise had a centralized administrative control, and, in agreement with the local needs, offered to their users a certain number of vacancies for parking.

At the end of the Second World War, the shopping centers industry has shown great progress, as demonstrated by the Lijnbaan Shopping in Rotterdam - Holland. In the fifties, the malls have appeared, which have as their main difference the consumers' circulation inside the building, different from the previous conceptions, where the shop windows faced the street. On October 7, 1956, the Southdale Center in Edina, Minnesota – USA was built, having an all closed construction due to the rigorous winter; its architectural model was adopted by all future shopping centers, not only in the architecture, but in the vacancies for parking which turned until now to be a pattern. Regarding the open mall, its pioneer was the King of Prussia, in Philadelphia, built in 1958, whose success was attributed to the perfect alignment of the architecture content with the place. In the sixties, architectural variations and innovations appear in the market, with a great expansion in the European countries.

Certainly, the great urban agglomerates of Brazil are not equivalent to those in North America, because the car is not accessible to all, and the congested roads do not stimulate long displacements for the routine shopping. However, Barra da Tijuca, in Rio de Janeiro, is an area of fast urbanization, with a town planning project addressed to the high middle classes while most of the dwellers own at least one car. The area is geographically isolated from the remaining of the city by the ocean, by mountains, forests, and lagoons. Its occupation has started at the beginning of the seventies, after the construction of a fine-looking system of tunnels and viaducts coasting the sea. In order to organize the urban development of the area, Lucio Costa, the same town planner responsible for the project of Brasília was called. A Master Plan of neighborhood occupation has taken place, influenced by the perspective of increasing use of private cars as a mean of transport. The highway design included a longitudinal road with six tracks in each direction and an avenue coasting the beach, with two tracks in each direction. The growth of the district was felt in an accelerated way, becoming the area in the city of Rio de Janeiro in which more buildings are built in every year. The planned occupation has reserved housing areas for great condominiums and

business areas in which magnificent shopping centers of several types, named by many as consumption temples, have been continuously built. However, the explosive population growth, the general increase in the middle classes income, the widespread access to automobiles, the attraction that beaches and the mentioned consumption temples exercise on the residents of other parts of the city, are bringing a progressive saturation in their highways, not only the internal ones but also the access roads, tunnels, and viaducts. These difficulties are discouraging long displacements for routine activities and encouraging the use of the referred Neighborhood Shopping Centers.

For ABRASCE (Brazilian Association of Shopping Centers), the definition of a Shopping Center requires six characteristics, as follows:

- 1) It is constituted by a set of planned stores, operating in an integrated way under one unique centralized administration;
- 2) It is composed of shops for exploring diversified or specialized branches of trade and services;
- 3) The tenant shopkeepers are subject to standardized contract rules, and, for most of the stores, such contracts shall state a variable rental clause according to the monthly revenue of the shopkeepers;
- 4) It has anchor stores, and special structural or marketing features, that operate as an attraction force to the Shopping Center in order to ensure the continuous influx of consumers essential for the proper performance of the venture;
- 5) It provides parking space consistent with the area the of shops and the corresponding influx of vehicles to the shopping center; and
- 6) It is under administrative control by individuals or groups of proven reputation and recognized entrepreneurship.

The ALSHOP (Brazilian Association of Shopping Centers' Shopkeepers) is another organization involved in the same matter which does not make distinction among rented and sold shops. The ALSHOP produces statistics based on questionnaires sent to many enterprises, while ABRASCE only works with numbers supplied by their associates. That differentiation explains the disagreement among the data supplied by both societies, as shown by Table 1.

Table 1: Source ABRASCE and ALSHOP, Brazil 2007

Item	ABRASCE	ALSHOP
Nº of Shoppings	333	577
Revenue	R\$ 44,0 billions	R\$ 53,4 billions
Employment	524,090	2,500,000

B. Types of Shopping Centers

In the popular understanding, there are many types of shopping centers, according to the kind of shoppers or the kind of merchandise sold. So, the terminology currently used includes: shopping center outlet; thematic shopping center; discount shopping center; wholesale shopping center etc.

However, in the line of the present study, we prefer to adopt the proposal of Hirschfeldt (1986), ULI that takes into account the physical size and the population attracted to shopping. According to the cited author (institution?), the shopping centers can be classified according to three patterns: Neighborhood, Community and Regional. Basically, the difference among them is linked to the type of anchor store, the physical area, and the diversification of the commercial project. Illustration 1, complemented by Table 1, indicates the area of influence of each of those models. Illustration 1 displays the influence upon the neighboring population, implies that a Regional Shopping Center reaches the resident population up to 25 km; a Community Shopping Center attracts the population up to 8 km, while the Neighborhood Shopping Center reaches the population up to 5 km apart.

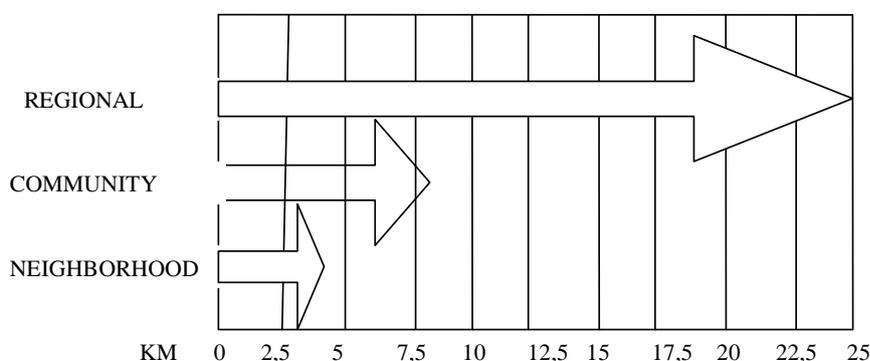


Illustration 1-Source:U.L.I. (1977)

Table 2 complements Illustration 1, specifying some of the general characteristics that each type of shopping center should have. Thus, for each type of shopping, Table 2 lists the kind of anchor stores, the rentable gross area, including higher and lower limits, the minimum area of the land, and the necessary support population, which is an element of particular interest. Certainly, for the largest shoppings a much larger support area is expected.

a) Neighborhood Shopping Center: More than 2/3 of the total number of Shopping Centers existing in the USA fall in this category. The attended population nears approximately 10,000 to 50,000 inhabitants. The total number of shoppings of this kind are responsible for a revenue of 75% to 90% of the total volume of sales in USA. The displacement time of a typical customer is, at most, seven minutes by car, and the distance is no more than 5 km of the residence. The RGA (Rentable Gross Area) of a Neighborhood Shopping Center is about 3,000 m² to 10,000 m²;

b) Shopping Center of Community: Enterprise in which the RGA is between 10,000 m² to 30,000 m². It attracts clients located up to 8 km of distance, serving a population of 50,000 to 250,000 inhabitants. The anchor store is usually a junior

department store and a supermarket. This kind of shopping includes convenience goods, personal services, feeding area, with a large diversification of products and activities;

c) Regional Shopping Center: Constitute the most common type of shopping center in Brazil. Its RGA is between 30,000 to 100,000 m². The area of influence is around 15 to 25 km, and may reach more according the easiness of access, the quality and the amount of stores and anchors stores. The served population will be larger than 250,000 inhabitants. This kind of shopping is anchored by complete department stores, i.e., stores with a built area of approximately 9.000 m², that offer different products, such as clothes, furniture, utensils, etc.

Table 2 - Shopping Centers' characteristics

Types of Shoppings	Anchor Stores	Rentable Gross Area- RGA (m ²)	Limits of RGA (m ²)	Minimum area of the lands (m ²)	Necessary population
Neighborhood	Supermarket	5,000	3,000 a 10,000	12,000	10,000 a 50,000
Community	Variety and Department Stores	15,000	10,000 a 30,000	40,000	50,000 a 250,000
Regional	Complete department store (One or more)	40,000	30,000 a 200,000	120,000 a 200,000	More than 250,000

Source: .L.I.(1977)

In fact, the district under study has experienced a disordered proliferation of multiple types of shopping centers which

resulted in the high vacancy index observed in many of those ventures. In this way, the present work proposes a normative study for planning the location of such shoppings.

C. The Shopping Centers Industry in Brazil

The common judgment considers the luxurious Iguatemi, inaugurated in Sao Paulo in 1966, as the first shopping built in the country. The numbers generated by the shopping Centers' industry, as shown in Table 3, based on the data of ABRASCE, the most conservative entity, show that this industry answers, in the year 2007, for 18% (eighteen percent) of the national sales, excluded the sales of automobiles.

Table 3 - Shoppings Centers' Characteristics, Brazil 2007

Total # of Shoppings:	346
Operating	333
Under construction	13
Rentable Gross area (m2)	7,452,171
Occupied area (m2)	55,658,000
Places for cars	440,000
Satelite stores	52,712
Anchor stores	1,300
Cinema / Theater	1,315
Employment level (in thousands)	524,090
Revenue (in billions R\$)	44,0
Sales in comparison to National Wholesale (Except Automotive sales)	18%

Source: ABRASCE, 2006.

Location Model

One of the main objective of this work is concerned with locating Neighborhood Shopping Centers. One of the most popular location model is the p-median model, which chooses the p medians or locations in order to minimize the weighted sum of distances from the population centered in every node to its closest service location. These vertices concentrate the population that lives in a relatively small space, the census tract, as defined by the National Census Bureau which offers reliable information, and revises the problem to a discrete location. The p-median model is used in cases in which each user frequently travels the distance between his/her home and the place that offers the needed services. This is typically the case of the daily home-school-home travel of the student but also the eventual but successive displacement home-shopping for attending the daily needs. Assuming that all nodes or vertices can be elected as medians, the p-median model can be modeled as the following binary integer-programming problem:

$$\text{Min } Z = \sum_{i=1}^n \sum_{j=1}^n w_i d_{ij} x_{ij}$$

$$\text{Subject to } \sum_{j=1}^n x_{ij} = 1; \quad i \in N \quad (1)$$

$$\sum_{j=1}^n x_{jj} = p \quad (2)$$

$$x_{ij} \leq x_{jj}; \quad i, j \in N \quad (3)$$

$$x_{ij} \in \{0,1\}; \quad i, j \in N \quad (4)$$

where:

[dij]n×n is the symmetric matrix of distances, with dii = 0, \forall i;

[xij]n×n is the allocation matrix, with xij = 1 if the vertice i is allocated to the vertice j, and xij=0 otherwise;

xjj = 1 if the vertice j is a median and xjj = 0, otherwise;

p is the number of service positions, or medians to be located;

N = {1, ..., n} is the set of vertices; and

wi represents the weight of vertice i;

The objective function indicates the minimization of the weighted distances between the customers' population and the places that offer the service; the restriction (1) indicates that each vertice i is allocated to only one vertice j; the restriction (2) establishes that only p vertices offer the proposed service; the restriction (3) says that the customer only goes to one vertice that must be a vertice or location that offers the service; and the restriction (4) imposes binary decisions.

III. A CASE STUDY: LOCATION OF NEIGHBORHOOD SHOPPING CENTERS AT BARRA DA TIJUCA – RIO DE JANEIRO, BRAZIL

Since the late seventies, Barra da Tijuca, a district of Rio de Janeiro, presents an explosive population expansion, with an ever increasing annual number of licenses granted for constructing new housing units. In 2005, for instance, the district concentrated nearly 50% of the newly built apartments in the whole city. On the other hand, the district has approximately the form of a rectangle, that remained for a long time isolated from the rest of the town, because of its geographical topography. In fact, in the south there is the ocean; to the east and north there are the Tijuca Mountains and also some lagoons; and to the west an ecological reserve. The urban development of Barra da Tijuca has started in the 60's and comply with the Lucio Costa's Master Plan, the same urbanist responsible for planning of Brasilia, the capital of the country. The design included two longitudinal avenues, one close to the sea and another in its interior, along which most of the shopping centers, commercial centers and supermarkets have being

located. The traditional small trade is restricted to two population streets that existed previously to the Plan, located at the doorway of the district on the east side, at the time only accessible by mountain roads. Illustration 2 shows a 2001 year map of the district, with the lagoons and mountains to the north, the ocean to the south, and the ecological reserve to the east. In the black dots we can notice the location of some commercial ventures, currently but improperly denominated as shopping centers.

The region encompassed by the case study also possesses extensive preserved land, large speculative areas awaiting valorization, and some areas of disordered occupation, most of them established before the Master Plan. Given those different nucleus, some simplifying hypotheses became reasonable, in order to reduce the number of centroids considered in the study.

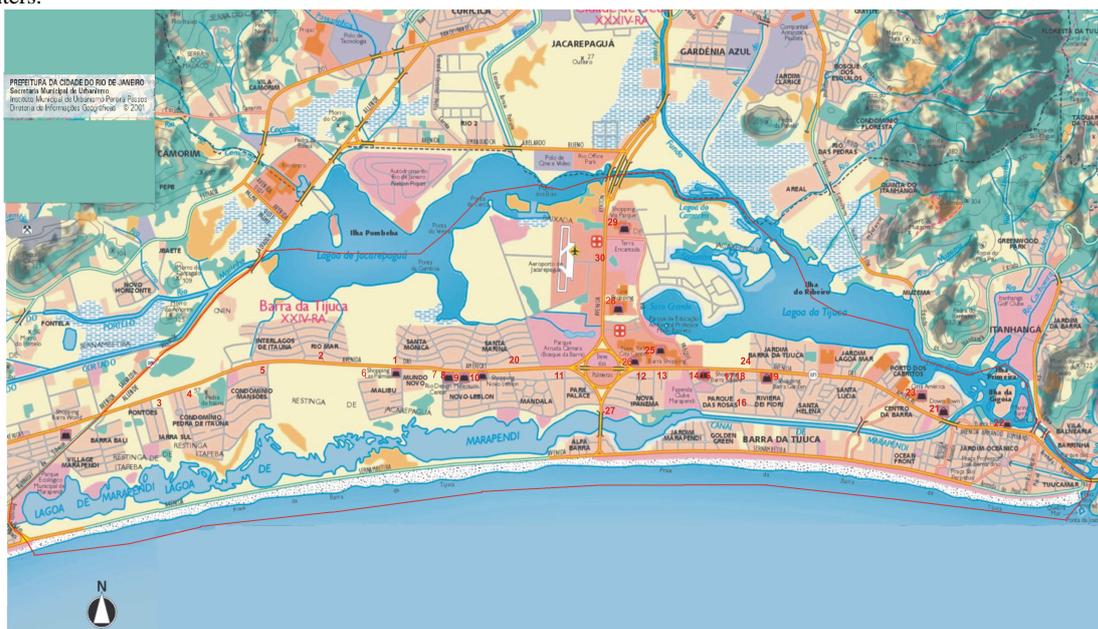
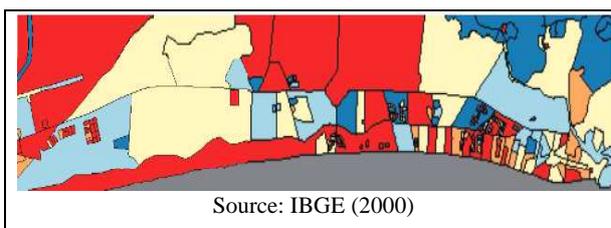


Illustration 2: Map of the district object of the case study showing the location of all shoppings

A. 4.1- Location Methodology.

The location methodology parallels similar studies that apply the p-median model, as those public facilities applications mentioned before, according to the logic of minimizing of the average distance user-installation. The methodology considers each census tract, whose population is periodically counted during the national Census. These tracts are small geographical units defined by the national institute responsible for the demographic census. Illustration 3 portrays the census tracts used in the districts pertaining to the case study. For each census tract, the referred institute makes public several information such as the population, age groups, average income, and so on. Based on both maps and data, the gravity center of each census tract is calculated, and the weighted network becomes available, according to the standard procedures.

Due to the many condominiums composed by residential buildings and houses, some simplifications and adaptations became convenient for the purposes of the present study. Most often, one condominium has inside its limits several other census tracts, constituted by one or more residential buildings.



Source: IBGE (2000)

Illustration 3 –Census section generated by Estatcart

This was done in the following way:

- a) For each census tract in the interior of a large condominium, its center of gravity has been identified;
- b) The centroid of each large condominium was calculated, but disregarding the census tracts that might be inside it;
- c) The center of gravity of a group of centroids has been found through the simplified calculation:

$$x_n = \sum \frac{w_i}{w} x_i \quad \text{and}$$

$$y_n = \sum \frac{w_i}{w} y_i$$

where i varies from 1 up to the number of tracts that are grouped, w_i is the population of each studied area; $w = \sum w_i$

is the total population of all census tracts included in the larger census area; x_i, y_i are the Cartesian coordinates of the gravity center of each area, and x_n, y_n the coordinates of the centroid of the collective census tracts;

d) In many cases small neighboring census tracts have also been consolidated in a single and larger census tract, to form a new centroid, calculated in the same way as before, agglutinating census tracts and the residents in the areas;

After these simplifications, the p-median model was applied. It is important to say that, according to the Census of year 2000, there was 172 census tracts in the district, with a total population of 89,142 people. Given the simplifications described before, those 172 tracts were reduced to 51, as shown in Illustration 4. On the one hand, about 94 tracts had large population aggregated in a relatively small space composed by blocks of buildings, suggesting the reduction of 94 to 16 tracts. On the other hand, among other 71 tracts, 28 were internal to other tracts, due to the many condominiums, resulting in the incorporation of these internal sections to their respective external tracts. So, there was a drastic reduction in the problem size, without any major loss on precision.

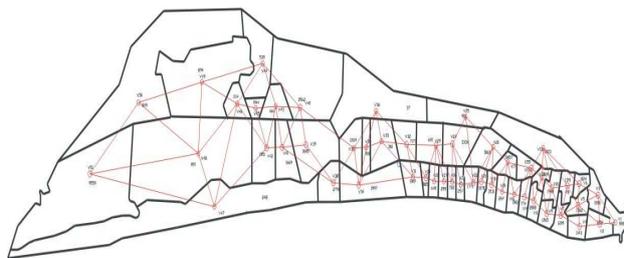


Illustration 4- Centroid Vertices location of each section

In this way, with 51 vertices and 89,142 inhabitants, the purpose of the model was to find the best location for installing some Neighborhood Shopping Centers, to support the commercial demands of the population. Taking into account the recommendations of ULI (Urban Land Institute) that the population support to make possible a Neighborhood Shopping Center would be from 10,000 to 50,000, this study has assumed a value of about 15,000 people, resulting in the proposal of $p = 6$ shopping centers of the kind for the studied area.

It would be the moment to apply the p-median model but, given the typical limitations of Barra da Tijuca, a simpler and

very logic heuristic solution was chosen, that, in a certain way, reminds the diagrams of Voronoi. In words, the idea of this diagram can be so explained: suppose a given location is designed to hold a facility and to satisfy the public demand in its proximity. Its area of influence corresponds to a bubble, that expands according to a growing circle until it touches one of the neighboring circles produced by similar bubbles generated at the same pace by the other proposed locations. The process finishes when all the centroids are reached by one of the bubbles.

In the present case, we started from the east part of the district, which was at the origin of the occupation, and we tried to create connected areas, with about 15,000 inhabitants, but taking into account the local geographical restrictions, as lagoons, bridges, restricted accesses, and so on. Advancing to the left and following the same procedure, the area was divided in six regions, as shown in Illustration 5, in which the positions of the six proposed shopping centers are marked.

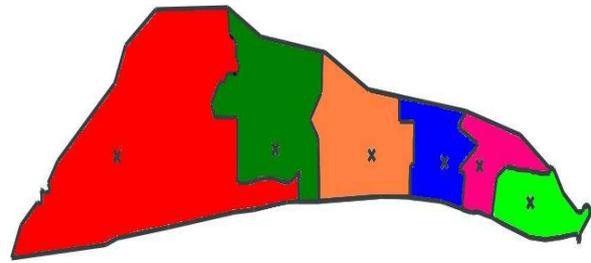


Figure 5 - Subset of sectors, each with approximately 15,000 people

As a location study, the work would be considered complete at this point, however, since the objective was to propose a normative solution seeking public implementation, one more step was given taking into account two relevant items. The first was to verify the zoning for multi-commercial buildings and for commercial areas as defined by the Master Plan. On the other hand, it is important to notice that any shopping center is a generator of traffic, and its success is also linked to the easiness or difficulty for a customers' entrance/exit. So, the location proposal should consider the street map and the flow of vehicles. Therefore, moving from the right to the left in Illustration 5, the first shopping center would be located in a cross road, and the four others along the main longitudinal road that crosses the district.

The Illustration 6 displays the main roads of traffic in the district, including the two longitudinal roads and the most important cross roads. So, the proposal is to locate the shoppings in those main roads as long as the location is close to the main traffic flows. As a final phase of this location study, the subsequent stage has to involve the micro location, in which costs of land as well as their availability near the selected points ought to be chosen, before the construction project is elaborated.

As a critical evaluation of the study, at least three affirmatives might be stated: i) the current situation differs significantly from the present proposal, because the real estate' boom in the area

has impelled constructions of commercial centers of all kinds without any appropriate study, generating a situation of high vacancy levels and frustrations for the investors; ii) the demographic growth, that shall remain intense, may justify in the near future an increasing number of shopping centers; and iii) the use of a population supports inferior, 10,000 people for instance, still within the limits suggested by the ULI, would recommend an increase in the number of Neighborhood Shopping Centers from six to nine.

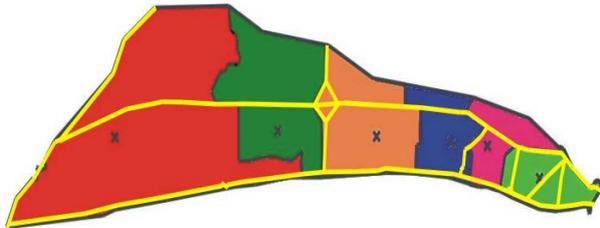


Figure 6 – Location of main highways in the district.

IV. CONCLUSIONS.

The location of shopping centers seems to be an important issue but rarely if ever addressed in the technical literature. There may two reasons for this: one is that shoppings are rarely planned to compose a network but rather are individual decisions about a single venture. On the other hand, by the same token, the technicalities involved in the decision process belong to the knowledge of consultants which prefer not to openly share their findings and knowhow.

This research has produced a brief overview of the development of the shopping industry from the early times of trading up to the present magnificent shopping centers, or temples of consumption. A number of technical studies has been found which categorize the main types of shopping centers and evaluates the shoppers' behavior. This has led to the concept of Neighborhood Shopping Center which satisfies the routine demand of the public living in their proximity.

The case study developed has considered a pleasant area made recently accessible by way of tunnels and roads coasting the ocean. The area is experiencing a very fast inhabitation with dwellers mostly from the upper classes of Rio de Janeiro, tending to rely on cars for their displacements. An obvious consequence has been the real estate valorization which has resulted in a few millionaires in a short period.

The resulting speculation and the perspective of continuous opportunities for capital gains has determined an uncontrolled expansion in all kinds of constructions including housing, business, and shopping areas. However, the present world crisis has shown a number of contradictions that might be attributed to the significant excess of commercial enterprises. Points (a) through (d) highlight some of the consequences from the excess of optimism and the general mistakes made:

a) Enterprises Overlap - a large number of enterprises overlap, suggesting a demand inferior to the one necessary to

sustain the feasibility of the projects.

b) Lack of studies for the implementation of the enterprises. The installed shoppings were in their great majority sold, i. e., the initial entrepreneur sold the units for a third person. For ABRASCE, nor either for ICSC (International Council of Shopping Centers), organ to which the first is affiliated, this sale no longer characterizes the enterprise as an actual shopping.

c) The mistakes listed on items (a) and (b) above result in vacancy for many enterprises. It might be noticed that, while the first installed shops are kept relatively well busy, the ones that followed have several empty units. The existence of empty units led their owners, in order to minimize their losses, to lease such units to any activity, causing a vicious cycle where the failure in the mix of units to be located, carries a lower frequency of consumers, causing new empty units, and so forth. That cycle is eventually compensated by the enterprises in which the units are under the control of a group, which only open concessions for leasing units considered strategic, and eventual losses in these units are compensated by leases of other stores.

d) Inexistence of new releases. Over the recent past, even preceding the international crisis, no new shoppings areas in the district have been initiated, only the construction of residential buildings and commercial offices. Clearly, there is no justification for new releases of shopping centers in the district, suggesting a partial validation of the present study. In face of the certain economical loss, the owners who acquired units in those enterprises expect the following events, in order to compensate for their investments: i) other residential releases and commercial offices are established close to their enterprises; ii) that the accelerated population growth remains, in a way to create new flows of consumers; iii) that the new releases of commercial centers are built but supported by location studies ordered by their entrepreneurs or the public authorities following the normative character proposed in the present study.

Acknowledgments: The second author is grateful to the National Council for Technological and Scientific Development, CN

References

- [1] Alves, S.F.M (2001), Assentamentos Urbanos. Revista da Faculdade de Engenharia e Arquitetura, Vol 3, no.1.
- [2] ALSHOP (2005), Anuário Brasileiro de Shopping Centers: Informações da Indústria de Shopping Centers do Brasil.
- [3] Blay, A; Sucupira, F.E, Do Trabalho à Civilização, Editora Fulgor Ltda, 1962.
- [4] Drezner, Z., K. Klamroth, A. Schöbel and G. O. Wesolowsky (2002), The Weber Problem, In Z. Drezner e H. W. Hamacher, Facility Location: Applications and Theory, Springer Verlag, pp. 1-36.
- [5] Galvão, R.D. Acosta, and L.G.E. Boffey, B. (2002). A Hierarchical Model for the Location of Perinatal Facilities in Municipality of Rio de Janeiro. European Journal of Operational Research, 138, pp. 495-517.
- [6] Hirschfeldt, R. V., Shopping Center – O Templo do Consumo, Brazilian Association of Shopping Centers Shopping Center – ABRASCE – Rio, 1986.
- [7] Monteiro, R. A. F. and A. J. Pascoal, A Model and a Heuristic for the Primary Health Care Coverage Planning Problems in Portugal – Application to “Cova da Beira” Case, Investigação Operativa, Vol. 25, pp. 195-227.

- [8] Pizzolato, N.D., F.B. Barcelos and L.A.N. Lorena, School Location Methodology in Urban Areas of Developing Countries, *International Transactions in Operational Research*; Vol. 11, 2004, pp. 667-681.
- [9] ReVelle; C.S. and H. A. Eiselt (2005) Location Analysis: A Synthesis and Survey. *European Journal of Operational Research*, Vol. 165, pp. 1-19.
- [10] Reese, J. (2005), Methods for Solving the p-Median Problem: An Annotated Bibliography. Department of Mathematics, Trinity University, San Antonio, Texas, EUA.
- [11] Rossi, A. M., and Natacha C. Raiune, Shopping Concept, Brazilian Association of Shopping Centers – Rio, 1988.
- [12] Teixeira, J. C. and A. P. Antunes; A Hierarchical Location Model for Public Facility Planning; *European Journal of Operational Research*; Vol 185, 2008, pp. 92-104
- [13] The Urban Land Institute, U.L.I (1977). *Shopping Centers Development Handbook*, Washington.
- [14] Weber, Alfred (1909). *Über den Standort der Industrien*. Tübingen, Germany. (English Translation by C.J. Friederich, 1957, *Theory of the Location of Industries*. Chicago University Press, Chicago).

Dynamical Handling of Straddle Carriers Activities on a Container Terminal in Uncertain Environment

- A Swarm Intelligence approach -

Stefan Balev, Frédéric Guinand, Gaëtan Lesauvage*, Damien Olivier

Université du Havre, LITIS EA 4108
BP 540, 76058 Le Havre, France

E-mail: {stefan.balev, frederic.guinand, damien.olivier}@univ-lehavre.fr,
gaetanlesauvage@gmail.com

Abstract—The CALAS project consists in a laser measure system allowing to localize precisely straddle carriers in a container terminal. The information given by such a tool makes an optimization possible. In fact, a box terminal is an open system subject to dynamics, in which many events can occur. Among others, they concern container arrivals and departures. Within the terminal, straddle carriers are trucks which are able to carry one container at a time in order to move it through the terminal. We aim to optimize the straddle carrier handling in order to improve the terminal management. Moreover, missions come into the system in an unpredictable way and straddle carriers are handled by humans. They can choose to follow the schedule or not. For these reasons, the exact state of the system is unknown. The optimization process that we try to build must be fail-safe and adaptive. In this context, we propose an approach using a meta-heuristic based on Ant Colony to resolve the problem of assigning missions to straddle carriers. We built a simulator which is able to test and to compare different scheduling policies.

Index Terms—swarm intelligence, colored ant colony system, dynamic graph, multiple criteria optimization, vehicle routing problem, container terminal.

I. SYSTEM DESCRIPTION

The CALAS project aims at localizing precisely handling trucks on a box terminal. It uses a laser localizing system and software which allows to deal with the data sent by laser sensors. This project is the result of a collaboration between *Laser Data Technology Terminal* company and the *Terminaux de Normandie* company. The goal of the CALAS project is to know the state of the terminal in real time, meaning both containers and vehicles location.

A container terminal is divided into three main areas (see Fig. 1). Each part is a set of box rows where containers can be stacked up and these areas are linked by oriented roads. The first area is the quayside. It is beside a channel where ships can tie to the docks. It is an area bound to prepare the ship (un)loading. The second area, the landside, is used to load or unload trucks and trains. The third part is a storing area linking the two others. Containers are moved into this area when a ship, a truck or a train is unloaded, and containers are

moved from this area when a ship, a truck or a train is loaded. Managing a box terminal involves three kinds of tasks:

- Preparing a ship (un)loading;
- Preparing a truck (or a train) (un)loading;
- Optimizing storing area.

In order to accomplish these tasks, containers are moved from one position to another. Such moves are called missions. Each mission is assigned to a straddle carrier.



Fig. 1. Terminal de Normandie, Le Havre, France¹.

The container terminal is an open system subject to dynamics. Though a subset of missions is known before starting the schedule, new missions arise when the schedule has already been established and its execution has started. Moreover, trucks arriving time is not known precisely enough to forecast container delivery. If a truck is late, the straddle carrier which has to load or unload it, could be assigned to another mission instead of staying idle and waiting for the truck. Human behavior also affects the system because straddle carriers are handled by human drivers who can choose to follow the schedule or not.

II. RELATED WORK

In such a system, the turn around time of both vessels and trucks/trains has to be as small as possible. Three different ways have already been used to solve this real problem.

* Corresponding author.

¹source: <http://www.t-n.fr/tn.htm>

First, the analytical approach is based on a study of interrelated factors which have to be taken into account to improve the efficiency of the system. In [1], an integrative decision support system is described. It has been created by studying inter-related decisions made daily in a container terminal. The authors evaluated the system at a terminal in Hong Kong and measured a reduction of 30% of the ships turn around time and the costs of container handling have dropped by 35%.

The second approach is the simulation. It consists in building a simulator which is able to test several methods of optimization. In [2], the authors have used both a genetic algorithm and a neural network system for the regulation of container yard operations. With 2 berths, 64 blocks, a planning period of 24h and a forecast period of 3 days, their simulation had shown a reduction of the total ship waiting time from 64h to 46h. In [3], the authors tried to improve the performance of the Rotterdam's Maasvlakte port area in studying its design. Their simulation gives information about quay length, storage capacity and handling and transport equipment of the terminal. Their results are useful for designing the next terminals.

The last approach is the multi-agent system (MAS). Thurston and Hu [4] aimed at improving the performance of the terminal by a dynamic and cooperative rescheduling of quay cranes and straddle carriers. Here each part of the system is considered as an autonomous agent able to take decisions according to the information of its own environment. Henesey *et al.* [5]–[8] have developed this idea. Their agents try to reach their own goal by searching, coordinating, communicating, and negotiating with other agents. They take their decisions according to a market based mechanism. Like in an auction, they bid for winning a task. Their system allows to test several policies of berthing, stacking or sequencing. They figured out that good decisions about stacking and berth allocation impact positively on the vessel turn around time.

According to these last conclusions, it appears that optimizing the performance of a container terminal means handling the vehicles' moves and their missions allocation. In this context, we deal with a vehicle routing problem.

III. VEHICLE ROUTING PROBLEMS

Vehicle routing problems (VRP) are largely studied and represent practical interest since they appear in many industrial processes. In general, VRP can be formulated as follows. One or many vehicles must start from a depot, visit a set of customers, delivering (or picking-up) some goods, and come back to the depot. The aim is to minimize the vehicles' routes. Many different subproblems belong to the VRP class, such as Capacitated Vehicle Routing Problem (CVRP) or Vehicle Routing Problem with Pickup and Delivery (VRPPD) for instance. Every subproblem contains a little variation of the main one, for example, there can be many depots, or vehicles must respect time windows... We distinguish static and dynamic instances of these problems because the methods to solve them are different.

A. Vehicle Routing Problem with Time Windows (VRPTW)

The Vehicle Routing Problem with Time Windows [9] (VRPTW) consists in visiting a set of cities by a set of

capacitated vehicles, optimizing overall path length. For instance, an Italian factory produces toys. It has to deliver a set of stores spread all over the country and goods are carried by trucks. Trucks capacity is restricted and they all start from the factory depot. Deliveries can only be done during a defined time interval. If a truck comes too early, it will have to wait. A solution to this problem should minimize the global length of the trucks runs.

The Dynamic VRPTW (DVRPTW) includes dynamics of the new orders. For the above example, if the stores can ask for deliveries when an already scheduled plan is running, then this problem belongs to DVRPTW class.

B. Pickup and Delivery Problem (PDP)

According to [10], PDP contains three subclasses:

1) *Many to Many Pickup and Delivery Problems (M-MPDP)* : Here, the vehicles have to pickup many objects to many locations. This kind of problem still relatively neglected because it is not frequently present in real situations.

2) *One to Many to One Pickup and Delivery Problems (1-M-1PDP)* : In this class, there are two different directions for the goods. They are first delivered to the customer. When the customer has done with them, he will ask for bringing back the goods to the depot. These problems may be with single or combined demands. In the first case, each customer asks either for a delivery or a pickup. With combined demands, the same customer can ask for both a delivery and a pickup.

3) *One to One Pickup and Delivery Problems (1-1PDP)* : This is the main subclass of Pickup and Delivery Problem, meaning the most frequently encountered problem in real life. It deals with picking-up one object at one location and delivering it to one destination. The main problem of this class is the Vehicle Routing Problem with Pickup and Delivery (VRPPD). In this problem, we have to compute the best routes for a fleet of vehicles in order to move objects on a graph. Every route has to start and to end at the depot. The difference to a 1-M-1PDP is that here, each object has its own pickup and delivery location.

When the problem deals with people, it is called Dial-A-Ride Problem (DARP). Some particular cases of VRPPD problems like the Stacker Crane Problem (SCP) are also common in practical life. This is a single vehicle with unit capacity problem. In another subproblem, vehicles are allowed to temporarily drop their loads on special locations called transshipment points to be able to answer customers demands faster. This is called Vehicle Routing Problem with Pickup, Delivery and Transshipment.

When some requests are not known in advance the above static problems may become dynamic. Those Dynamic Pickup and Delivery Problems [11]–[13] (DPDP or DVRPPD)

consist in optimizing vehicles routes in order to pickup a load somewhere, then to deliver it to its destination, adapting these routes to the new incoming orders without recomputing from scratch. Most of the time, DVRPPD has to handle time windows (DVRPPDTW). Indeed, to start a mission, vehicles have to wait the beginning of its time window. If it is not respected, the vehicle will have to wait for the right time and, meanwhile this vehicle becomes useless.

As we have just seen, the Vehicle Routing Problem class contains a lot of different subproblems. It becomes very important to exactly identify our own problem.

C. Identification of our problem

In our problem [14], several vehicles (straddle carriers) of unit capacity must accomplish missions (by moving containers within the container terminal). They can also use transshipment location to make the tasks more efficient. A very specific aspect of our system is that the straddle carriers can start from anywhere, *i.e.* they do not have to start from the depot. Moreover, every mission has a time window in which the container must be delivered. If a vehicle comes too early for picking up or delivering a container, it will have to wait the beginning of the missions time window. Furthermore, if a straddle carrier is late, meaning its time window is already closed, in some cases, the mission must be aborted and a new one dealing with the same container will appear into the system.

For all these reasons, our problem belongs to the Dynamic Vehicle Routing Problem with Pickup and Delivery and Time Windows (DVRPPD-TW).

Three interconnected problems must be solved:

- Minimize straddle carriers moves: shortest path problem
- Minimize resources: clustering problem
- Minimize customers delays: scheduling problem

In order to construct a good schedule, the system must integrate the shortest path concept. In the same time, scheduling shortest paths tends to reduce straddle carriers moves. Moreover, we have to define a quality of service level to satisfy customers while lowering operation costs. This is a dynamic large scale problem which requires a real time solution. We propose an on-line algorithm based on Ant Colony Optimization [15], [16] and more precisely on a colored version of this swarm algorithm [17].

IV. ANT COLONY AND STRADDLE CARRIER HANDLING

Ant Colony [15], [16] is a meta-heuristic which makes a solution appear thanks to the run of artificial ants into the solution space. The system is self-regulated. In fact, ants spread pheromone according to the solution quality (positive feedback) but the pheromone tracks evaporate progressively (negative feedback). The positive feedback makes the algorithm converge to a quality solution, and the negative feedback prevents it to trap into a local extremum.

Ant Colony with one colony provides a sorted list of missions to accomplish [18]–[20]. The problem is to set a mission to a specific straddle carrier.

We propose to employ a solution using colored ants [17]. In our model, every straddle carrier represents a colony with its own color. Convergence is assured by the fact that ants are attracted by the pheromone of their own colony and repulsed by the pheromones of foreign colonies. This approach simulates a mechanism of collaboration and competition between colonies and will provide a sorted list of missions for each straddle carrier.

A. Modelling

1) *Graph construction:* Our algorithm uses a graph representation of the problem. In this oriented graph, every vertex represents a mission. We first build a precedence graph. We say that a mission is prior to another if its time window starts before the one of the other mission. Once this precedence graph has been built, a colored node is added to the graph for each straddle carrier. Those vertices are linked to every compatible mission by an arc of the same color. Next, the arcs added during the precedence graph construction are colored according to the compatibility between the straddle carrier and the missions. In fact, if two missions, linked by an arc in the precedence graph, match with the straddle carrier of color c , then we color the edge between them with the color c . If there is already a colored arc between these two nodes, then instead of changing the color of this arc, we add a new one colored with the color c . At the end, if uncolored arcs remain, they are removed from the graph. So we obtain a multi-graph allowing to run our colored ant colony algorithm.

• Straddle Carriers:																							
		<table border="1"> <thead> <tr> <th>Name</th> <th>Color</th> </tr> </thead> <tbody> <tr> <td>s0</td> <td>green</td> </tr> <tr> <td>s1</td> <td>blue</td> </tr> </tbody> </table>		Name	Color	s0	green	s1	blue														
Name	Color																						
s0	green																						
s1	blue																						
• Missions:																							
<table border="1"> <thead> <tr> <th>Name</th> <th>Start</th> <th>End</th> <th>Matching vehicles</th> </tr> </thead> <tbody> <tr> <td>m0</td> <td>5:00</td> <td>6:00</td> <td>s0, s1</td> </tr> <tr> <td>m1</td> <td>5:30</td> <td>6:00</td> <td>s0</td> </tr> <tr> <td>m2</td> <td>7:00</td> <td>9:00</td> <td>s0</td> </tr> <tr> <td>m3</td> <td>6:00</td> <td>7:30</td> <td>s0, s1</td> </tr> </tbody> </table>				Name	Start	End	Matching vehicles	m0	5:00	6:00	s0, s1	m1	5:30	6:00	s0	m2	7:00	9:00	s0	m3	6:00	7:30	s0, s1
Name	Start	End	Matching vehicles																				
m0	5:00	6:00	s0, s1																				
m1	5:30	6:00	s0																				
m2	7:00	9:00	s0																				
m3	6:00	7:30	s0, s1																				

Fig. 2. Example of a simple instance of our problem

2) *Example:* Consider a simple instance of our problem where two straddle carriers have to execute four missions. The compatibility between these vehicles and the missions are as in Fig. 2. So we first build the precedence graph (see Fig. 3). Then we add the straddle carriers nodes (see Fig. 4). Finally we color the arcs as described above. The Fig. 5 shows the multi-graph obtained using this procedure.

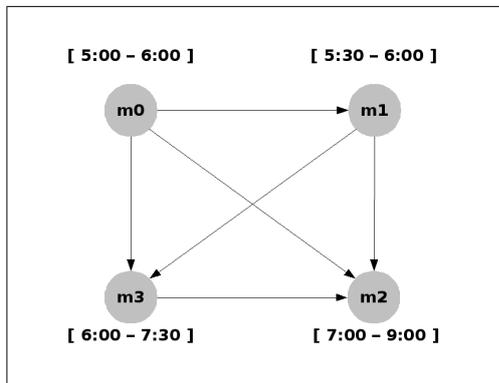


Fig. 3. Precedence graph of the problem described in 2

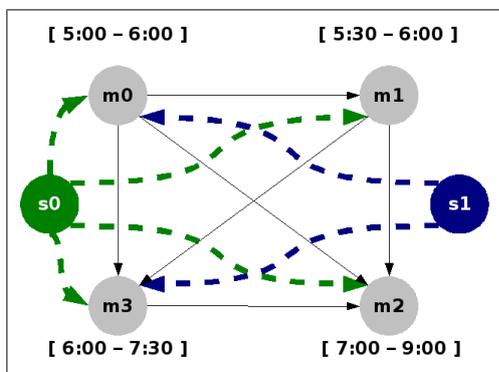


Fig. 4. The vehicles nodes are added to the precedence graph

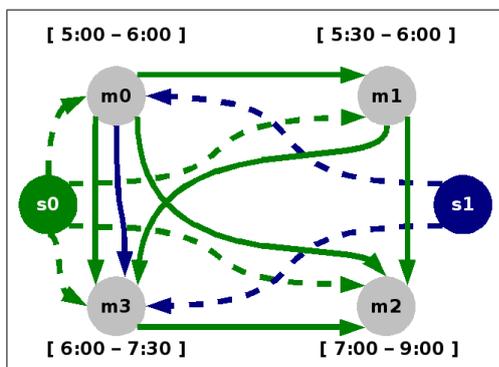


Fig. 5. Mission graph for 2 straddle carriers and 4 missions

3) *Arcs weighting*: We introduce arc weights which influence the ants when moving in the graph. The weight of an arc measures how efficient it is to assign the two missions connected by the arc to the same carrier. This part of our model provides flexibility and allows to test different weighting policies.

Our policy takes into account both the cost of the mission execution and the time windows proximity. Indeed, if two missions have time windows which are too close, and if they are assigned to the same straddle carrier, then the execution

of the first mission will cause the overrunning of the time window of the second mission. It is really important to prevent these phenomena by modelling the linking penalty. We also define a concept of priority. The more the end of the time window is close, the more the priority is high. The weighting function of the arcs takes into account also the distance between the delivery location of the first mission and the pickup location of the second one.

B. Colored Ant Colony Algorithm

In this algorithm, each straddle carrier has a corresponding colony of the same color. Each colony starts from the node representing its straddle carrier. Then, the ants move in the graph using only the arcs of their color. When an ant is in a node, it chooses the next node to visit according to three factors:

- the pheromone rate of its own color
- the pheromone rate of foreign colors
- the weight of the arc

The ant is attracted by the pheromone of its color and repulsed by the pheromone of different colors. Once an ant have reached the chosen node, it spreads pheromone according to the quality of this choice. When a straddle carrier of color c asks for a new mission, it chooses the mission which has the highest rate of pheromone of color c . The overall description of algorithm is shown on Fig. 6.

```

1: for all colony  $c$  do
2:   for all ant of colony  $c$  do
3:     choose an unvisited destination
4:     move towards it according to the ant speed
5:     spread pheromone
6:   end for
7: end for
8: evaporation

```

Fig. 6. Colored Ant System main algorithm

Our ant colony approach is relevant for solving the considered problem because of its dynamic nature, the large size of the solution space and the real time constraint.

The main asset of ant colony is to provide an anytime solution. It is an on-line algorithm which adapts easily to the changing environment. Indeed, ants reinforce the pheromone rates to get closer to the best solution. At the same time, evaporation process provides a feedback control of the algorithm by preventing it to get stuck into a local optimum and allowing dynamic events to be handled.

Ant colony deals with many parameters such as evaporation, solution evaluation, ants quantity and speed, dynamic events, *etc...* Here is the major weakness of this metaheuristic. Solution quality strongly depends on these interdependent settings. We have tried to make these parameters self-adaptive. We use a local method to adapt some of these parameters on-line.

C. Division of labor

As there are several distinct colonies and each ant has only a local vision of its environment, there is no way to use a pheromone spreading process based on a global characteristic. In fact, in this architecture, a colony cannot compare the quality of its own solution to the solutions of the other colonies. So, we must use the same pheromone spreading process for each colony. However, we are able to adapt the quantity of pheromone spread by ants of a colony according to the corresponding vehicle skill for a task. Indeed, we observed that we can reduce the serving time of mission by specializing the vehicles into a kind of missions.

We can increase the quantity of pheromone spread by a given vehicle for tasks concerning a specific area in the container terminal and decrease the quantity spread for the tasks located into the other areas. At the same time, we do the opposite for all other vehicles. In this way, we try to specialize the vehicle in a kind of tasks and we are able to regulate these quantities by taking into account both the preference and the distance criteria.

This regulation keeps the benefit of allowing a vehicle to take a mission for which it is not specialized. It is really important in some cases where the number of missions is high because this regulation prevents the system from having unused vehicles in an area of the terminal and unaffected missions, close of the end of their time window, in an other area.

This original approach has a limit. In fact, the time needed by the adaptive system for affecting a vehicle to a mission which does not belong to its specialization may be considerable. For this reason, the system may become less responsive than with no specialization.

D. Reducing resources

Always in a cost lowering purpose, we try to decrease the number of straddle carriers in the system. Our current solution to the entire problem tends to distribute the missions upon all the vehicles. So, every vehicle has almost the same activity rate. But if this rate is under a defined lower threshold, it is possible to conclude that a vehicle could be removed. Otherwise, if the rate is greater than the upper threshold, it is possible to say that a new vehicle should be added to the fleet.

The thresholds must be computed by taking into account several facts. First, it has to deal with the quality of service. Indeed, the system must answer the requests before the end of their time windows. Furthermore, if a vehicle is ready to serve several missions before the beginning of their time windows, it means that this vehicle is maybe superfluous, and the thresholds must be modified accordingly. On the other hand, the target rate has to deal with other criteria like the covered distance of a vehicle per mission or the ratio between the number of vehicles and the number of missions, and it has to set these criteria against the penalties of transcended time windows. Measuring the time of inactivity of every

straddle carrier may also lead the optimization. Concerning this last criterion, we must interrelate the time of inactivity with the penalties of transcended time windows.

So, as for the missions arrivals into the system, the number of vehicles is subject to dynamicity. A vehicle can break down and then must be sent to the maintenance. In function of the failure seriousness, we can estimate the time needed to repair the vehicle and so make it available for routing. We take a rate of fault into account for optimizing the number of vehicles into the system because if this number is as low as possible without transcending some time windows, it will become too low if one vehicle of the fleet breaks down.

V. SIMULATOR

The simulator has two main parts. The first one is the terminal simulation (see Fig. 7), and the second one is the Colored Ant Colony Optimization System (see Fig. 8). The first part contains an implementation of the terminal structure and components. Roads and crossroads provide the network of the terminal on which straddle carriers will be able to go. Some of these roads may contain containers. Quay crane locations are represented by these special roads, as well as the trucks handling locations. This terminal is built at the very beginning of the simulation. A scenario file is read to set the terminal configuration. The second part of the simulator contains the algorithmic view of the simulation, *i.e.* the dynamic mission graph. In this way, it shows how the missions are chosen by the vehicles. This part of the simulator uses GraphStream² toolkit which allows to handle dynamic graphs easily [21].

The simulator uses a discrete time engine which has to iterate every object of the simulation on every time step. During the simulation, the scenario file is read and some dynamic events are sent back to both terminal and Ant Colony views. In this way, the system can simulate the dynamicity of the incoming missions and of the vehicles availability.

In order to have relevant tests and results, we have to define several levels of dynamicity. In [22], Allan Larsen points out two main ways to measure the degree of dynamicity.

First, the degree of dynamism (*dod*) [23] is the ratio between the number of dynamical requests and the total number of requests. The main weakness of this measure is that it does not take into account the arrival time of these requests into the system. Indeed, with *dod* if the requests come into the system at the beginning of the day, the system is as dynamic as if they come late in the day. Yet, the later these requests are known, the shorter is the delivery delay. This lateness impacts on the performance of the system.

For this reason, Larsen *et al.* in [22] defined the effective degree of dynamism (*edod*) by the following formula:

²<http://graphstream-project.org/>

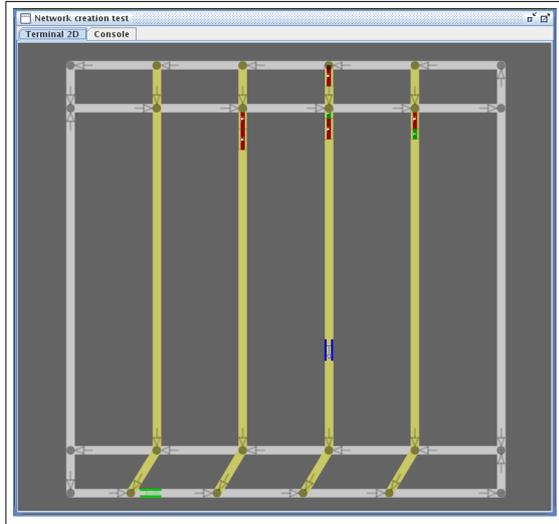


Fig. 7. Terminal view in the simulator

$$edod = \frac{\sum_{i=1}^{\eta_d} \frac{t_i}{T}}{\eta_d + \eta_s} \quad (1)$$

Here, η_s and η_d are respectively the number of static and dynamic requests, t_i is the arrival time of request i (with $0 < t_i < T$) and T is the time of the simulation end. This measure takes into account the average of the incoming time of the requests into the system. The more the dynamical requests come late, the more $edod$ will be high. If $edod = 0$, then the system is totally static. Else, if $edod = 1$, then the system is purely dynamic.

Every straddle carrier on the terminal simulation receives a schedule from the Colored Ant Colony System. Then they act in function of it and move to their pick-up location. Once they have picked-up their container, they move to the delivery location to achieve their mission. At the same time, the mission graph is dynamically updated and the colored ants keep colonizing it.

Simulator gives information about each mission like its length, container, straddle carrier, pickup and delivery time windows, *etc.* and about other parts of the terminal like the state of the roads for instance.

VI. PRELIMINARY RESULTS

As we are still collecting real data from our partners, we are just able to test the relevance of our modelling and of our algorithm on simulated data. For this purpose, we have first run a simulation with a static context, which means that every mission is known at the very beginning of the simulation and that the resources are always available. In a second time, we have added dynamic events such as new incoming missions. For each simulation we have measured the global time needed for achieving all the missions, the number of overrun time windows and the global overrun time. We consider that a time window has been overrun if the non respect of this time

	Static	Half Dynamic	Dynamic
dod	0	0.5	1
$edod$	0	0.25	1
End time	22693	22276	22693
Number of overrun tw	3	5	7
Overrun time penalty	6467	8477	12485

Fig. 9. Results of simulations

window represents a penalty for the container terminal. Indeed, if we overrun the time window of a mission in which we have to move a container from or toward a truck for instance, then the truck will ask for a compensation.

Figure 9 shows the results of three instances containing 12 missions and 3 straddle carriers. The only difference between these instances is their degree of dynamicity. Indeed, we have only changed the arrival time of these missions into the system to make them more or less dynamic.

As we can see in Fig. 9, our algorithm seems to act as expected. It means that the more the missions are known in advance, the better is the performance. The worst case occurs when the mission is known at the very beginning of its time window. These are preliminary results and we have not tested all the parameters of the ant algorithm yet.

VII. CONCLUSION

The problem considered in this paper belongs to the Dynamic Pickup and Delivery Problem with Time Windows class. However, it does not exactly fit. So it is an original unsolved problem. We propose to solve it using swarm intelligence method. An Ant Colony System is being developed. It uses colored ants and a graph modelling in order to plan a schedule. Moreover, we are trying to minimize the number of vehicles into the fleet in order to both maintain a sufficient quality of service and reduce costs. A simulator able to reproduce the behavior of such a system and to handle dynamic events is being developed. The preliminary results confirm that our algorithm is able to handle dynamicity and we are actually collecting data in order to compare the performance of our system into a container terminal environment with the current scheduling methods used in a terminal of the seaport of Le Havre in France.

REFERENCES

- [1] K. G. Murty, J. Liu, Y.-w. Wan, and R. Linn, "A decision support system for operations in a container terminal," *Decis. Support Syst.*, vol. 39, no. 3, pp. 309–332, 2005.
- [2] C. Jin, X. Liu, and P. Gao, "An intelligent simulation method based on artificial neural network for container yard operation," *Lecture Notes in Computer Science*, vol. 3174, pp. 904–911, 2004.
- [3] J. A. Ottjes, H. P. M. Veeke, M. B. Duinkerken, J. C. Rijsenbrij, and G. Lodewijks, "Simulation of a multiterminal system for container handling," *Container Terminals and Cargo Systems*, vol. 2, pp. 15–36, 2006.
- [4] T. Thurston and H. Hu, "Distributed agent architecture for port automation," in *COMPSAC '02: Proceedings of the 26th International Computer Software and Applications Conference on Prolonging Software Life: Development and Redevelopment*. Washington, DC, USA: IEEE Computer Society, 2002, pp. 81–90.

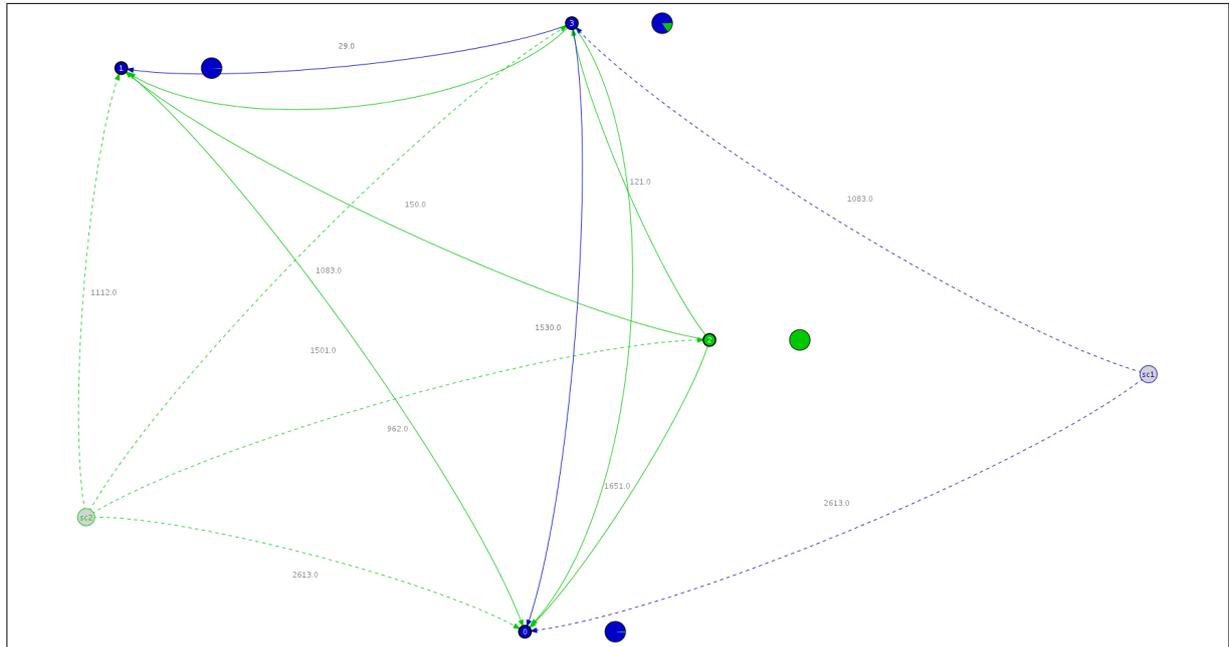


Fig. 8. Mission graph view in the simulator

- [5] L. Henesey and J. Törnquist, "Enemy at the gates: Introduction of multi-agents in a terminal information community," *C.A. Brebbia, G. Sciutto (Eds) Maritime Engineering and Ports III*, pp. 23–32, 2002.
- [6] L. Henesey, F. Wernstedt, and P. Davidsson, "A market-based approach to container port terminal management," 2002.
- [7] —, "Market-driven control in container terminal management," in *Second International EuroConference on computer applications and information technology in the maritime industries*, Hamburg, Germany, 2003.
- [8] L. Henesey, "Enhancing container terminal performance: A multi agent systems approach," Ph.D. dissertation, School of Engineering - Dept. of Systems and Software Engineering, Ronneby, 2004.
- [9] L. Bianchi, "Notes on dynamic vehicle routing - the state of the art," *Technical Report, IDSIA*, 2000.
- [10] G. Berbeglia, J.-F. Cordeau, I. Gribkovskaia, and G. Laporte, "Static pickup and delivery problems: A classification scheme and survey," *TOP*, 2007.
- [11] S. Mitrovic-Minic, "The dynamic pickup and delivery problem with time windows," Ph.D. dissertation, Simon Fraser University, 2001.
- [12] S. Mitrovic-Minic and G. Laporte, "Waiting strategies for the dynamic pickup and delivery problem with time windows," *Transportation Research Part B*, 2004.
- [13] S. Mitrovic-Minic, "Pickup and delivery problem with time windows : a survey," *Technical Report 1998-12, SCS Simon Fraser University*, 1998.
- [14] G. Lesauvage, "Gestion dynamique des activités des chariots cavaliers sur un terminal à conteneurs en environnement incertain," Master's thesis, Université du Havre, 2008.
- [15] M. Dorigo, "Learning and natural algorithms," Ph.D. dissertation, Politecnico di Milano, Italie, 1992.
- [16] M. Dorigo and L. Gambardella, "Ant colony system : A cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, 1997.
- [17] C. Bertelle, A. Dutot, F. Guinand, and D. Olivier, "Distribution of agent based simulation with colored ant algorithm," in *14th European Simulation Symposium*.
- [18] R. Montemanni, L. Gambardella, A. Rizzoli, and A. Donati, "A new algorithm for a dynamic vehicle routing problem based on ant colony system," *Technical Report, TR-23-02, IDSIA, Galleria 2, Manno, 6928, Switzerland*, 2004.
- [19] B. Bullnheimer, R. F. Hartl, and C. Strauss, "Applying the ant system to the vehicle routing problem," in *In Proceedings of the 2nd International Conference on Metaheuristics - MIC97*.
- [20] B. Bullnheimer, R. Hartl, and C. Strauss, "An improved ant system algorithm for the vehicle routing problem," *Annals of Operations Research*, 1999.
- [21] A. Dutot, F. Guinand, D. Olivier, and Y. Pigné, "Graphstream: A tool for bridging the gap between complex systems and dynamic graphs," in *EPNACS: Emergent Properties in Natural and Artificial Complex Systems*, 2007.
- [22] A. Larsen, "The vehicle routing problem," Ph.D. dissertation, Department of Mathematical Modelling, Technical University of Denmark, 2000.
- [23] K. Lund, O. B. G. Madsen, and J. M. Rygaard, "Vehicle routing problems with varying degrees of dynamism," *Technical report, IMM, The Department of Mathematical Modelling, Technical University of Denmark*, 1996.

A multiagent urban traffic simulation

Part I: dealing with the ordinary

Pierrick Tranouez, Patrice Langlois, Éric Daudé

Abstract— We describe in this article a multiagent urban traffic simulation, as we believe individual-based modeling is necessary to encompass the complex influence the actions of an individual vehicle can have on the overall flow of vehicles. We first describe how we build a graph description of the network from purely geometric data, ESRI shapefiles. We then explain how we include traffic related data to this graph. We go on after that with the model of the vehicle agents: origin and destination, driving behavior, multiple lanes, crossroads, and interactions with the other vehicles in day-to-day, “ordinary” traffic. We conclude with the presentation of the resulting simulation of this model on the Rouen agglomeration.

Index Terms— multiagent systems, traffic simulation, geomatics, multiscale

I. INTRODUCTION

IF traffic modeling is nearing a century of age, most of these models belong to Operational Research problems – finding an optimal solution balancing various constraints. In these models, roads and road users were abstracted and aggregated, so as to become a flow problem that could then be optimized. They can answer interesting questions in urban or public transport planning [1].

Sometimes considering average response to a problem is not enough for the scientific problem at hand. We are interested in a dynamic modeling of urban traffic. In this kind of problem, the actions of a few can have a definite impact on the global traffic. An accident implicating half a dozen vehicles in a strategic crossroads of a town can create a traffic jam wave that can affect thousand vehicles. This is the kind of complex phenomenon we would like to be able to model and simulate. Classic OR tools aren’t well suited to the task.

Although we aren’t the first to make this statement [2,3], models that tried to alleviate this too-large-scale limitation, have mainly tried to use cellular automata for the task. They added some level of individual-based components to their modeling, but still failed to encompass all that could be needed. Cellular automata are eulerian methods – intelligence is in one place, rules describe the behavior of bits of space.

The authors would like to thank the GRR SER and the region Haute-Normandie for the funding of the MOSAIIC program from which this work stems from.

P. Tranouez is with Litis, Rouen University. (e-mail: Pierrick.Tranouez@univ-rouen.fr)

Patrice Langlois and Éric Daudé are with UMR IDEES, Rouen University (email: Patrice.Langlois@univ-rouen.fr and Eric.Daude@univ-rouen.fr)

Values linked to the cells seem to simulate the entities of the modeled system, the same way alternatively lit crystals in an LCD display can give the illusion an object moves around a screen. This contrasts with lagrangian descriptions, where entities of an environment are distinguished, and their spatial coordinates are but one of their describing characteristics. Unlike what can be easily simulated in a CA, lagrangian entities have a *trajectory*: even in a discretized space a la CA, they can for example act according to something that happened n time steps and m space steps before or away, or according to a plan. This can’t *practically*¹ be done in a CA. Multiagent systems belong to this latter category of modeling. As we try to build a model with a grain fine to the level of geometrically correct individual vehicle behavior, from which at least town-quarters-level flow disturbance can arise, we believe this technique is the right one for the task.

II. FROM GEOMETRY TO TOPOLOGY

A. Geographical databases

A Geographical Information System is a system designed for creating, storing, analyzing and managing spatial data and associated attributes. Although it contains a relational database, it needs to go beyond what is needed for classical alphanumeric databases to manage geometrical information, which is continuous by nature, as opposed to the discreteness of usual databases. Indeed for example the database cannot contain all the points of two segments in order to compute a possible intersection: other storing and managing methods must be used for the geometric data of the system.

A geographic database is generally comprised of layers or coverage overlapping on a same spatial domain. Each layer contains homogeneous spatial features such as the limits of a city, the course of a river, the geometry of a road etc. Each feature is described in two different ways. First the geometric and optionally topological information is stored in different binary files in the base. Second the record description is a line in the record table; it contains different attributes and descriptions of the feature (generally text or numbers).

B. ESRI shapefiles

The first step of the constitution of our system is the constitution of a basic layer of geographic database. This layer is built from the importation of *shapefiles*, a GIS file format popularized by ESRI [4]. In order to build a traffic simulation,

¹ As opposed to theoretically, as cellular automata are universal calculators

we will build our model from data relative to the road network and optionally from other localized information such as living or working areas.

A shapefile is mainly constituted of three files: one contains the attribute table (.dbf), another contains the geometric data (.shp) and the third is an index allowing matching entries of the first with those of the second.

A shapefile contains only the geometric description of objects through a collection of 2D or 3D coordinates that represents, according to the layer type, a cloud of points, open polygon lines (for networks or closed polygon lines to describe the boundary of surfaces). The topological information, which describes in geomatics the relationships between the geometric entities, such as connections of edges with nodes in a graph or the adjacency between zones in a surface partition, is absolutely not present in a shapefile, and must therefore be computed by our application from the raw geometry of the imported data.

To build a realistic representation of the traffic network of an important urban agglomeration able to simulate the circulation of tens of thousand of vehicles, we had to conceive a network layer structure both complex and efficient. Furthermore, the importation of data coming from existing data provider such as IGN, NAVTEQ or Tele Atlas, we had to deal with the way each modeled things in their solutions.

C. Urban network structure specifications

A road network is modeled according to specifications that are in part common to any network and in part dependent on decisions made by the data producer.

1) General specifications

A road network shares the properties of any geographic network. It is constituted of two main geometric entities: *lines*, linear components, comprised of several shape points, and *nodes*, point components that join or terminate lines.

These two entities are joined in an *oriented multigraph* $G = (S, A, f)$ where S is the set of vertices, associated to the geometric nodes, while the set A of edges is associated to the geometric lines, while function $f : A \rightarrow S \times S$ associates to each edge one initial and one final vertex.

Unlike most other geographic information layer, a road network may not be planar: two lines can intersect in their planar projection without modeling an intersection in the real world. This happens when these lines are at different altitudes such as in bridges, tunnels, or motorway embranchments.

Furthermore, geographic graph are *topological graph* differ from usual graph in that they are associated to one geometric representation, called the *embedding* of the graph. Only vertices of degree 3 or more are considered to be true vertex, those of degree 2 being seen as *shape points*, useful for the geometric information they bring, but not “true” connectors. The geometric representation of the graph is always present to the mind of the geographer, which may create misunderstandings with other scientists more used to a more abstract representation of graph, with *planar graph* rather than *plane graph*. As previously said, it is also sometimes extended

to non planar graph: the geometric information in the shapefile represents in that case the projection on a connected compact 2-manifold of a graph embedded in a connected compact 3-manifold (intuitively: a 3D graph is drawn on a surface).

The attribute table associated to the network will contain all the traffic related information, such as the number of lanes, speed limits, sense of travel etc. Nonetheless this information may not be associated to elementary lines or nodes. For example major roads may contain different lines and important roundabout may contain different nodes and lines. We therefore defined the notion of super-nodes that relate to several nodes (and the assorted sub-graph) and super-edges that relate to several edges (and the assorted sub-graph). G is therefore a hypergraph in these conditions. Whether these are met or not depend on modeling decisions made by the data provider.

2) Geographic data based specifications

There are different ways of structuring the geographic information in a shapefile to model a network.

For example NAVTEQ chose in its Navstreets product to create a node for each intersecting link, even if the road they model are not connected. Another layer represents the relative elevation of the entities of this first layer. Both must therefore be used to correctly build the road network in our simulation. Another example is the orientation of the edges, as the links are oriented following another convention (called “Reference nodes”) than what could be used in a shapefile, and the edge must therefore be computed following this convention.

D. Building the topology from the geometry.

Building a topology from the geometric information contained for example in a shapefile depends on the kind of spatial organization we want to represent.

1) Planar mesh

In the case of a surface mesh (ex: limits of countries, of urban areas, of town quarters etc.), we aim at rebuilding the boundaries and the junction nodes between them from closed polygonal chains (aka polylines). The layer we produce is thereafter structured around a planar multigraph of vertices, edges and faces, and with each oriented edge associated to 2 vertices (initial and terminal) and to 2 faces (left and right).

The building algorithm uses a quadtree and a tree connecting each point, in which all the points of the shapefile are organized. Each leaf of the quadtree contains a point P_i and 4 branches for the 4 quadrants of space (NE, SE, SW, NW) surrounding P_i . When a branch is a leaf, it contains a point belonging to right quadrant relatively to its father, and vice versa. This structure allows for a quick detection of the multiplicity of points. For example, a point with a multiplicity of 3 or more will be associated to a vertex, while a point of multiplicity of 2 will be a shape point of an edge. Furthermore, the connection tree allows the quick detection of adjacent points along a polyline, and detecting the superposition of two lines forming the boundaries of two zones, or the succession of angular sectors around a vertex common to three polygons or more.

2) Network

In order to build the structure of a planar network (for example hydrographic or of roads), we do not store faces but the polar order of succession in the edges. Each edge stores the next edge turning left and the prior edge turning right. This structure is known as DCEL, Doubly Connected Edge List [5]. The algorithm to generate this topology uses the same dynamic quadtree structure to build the DCEL.

The road network often exists in 3D, although despite the existence of this possibility, most shapefiles only contain a 2D geometric representation. The data provider must in that case model the altitude differently, and our algorithm must be adapted to this. For example NAVTEQ's Navstreets [6] uses another layer called z-levels that must be consulted to know whether a point corresponds to a node or not.

At the end of this step, we have a topological graph that is structured like the road network, but without its semantics. We will now build from it and from the database part of the shapefile a non-topological graph that models this ontology.

E. From static topology to traffic-oriented network

1) Traffic oriented graph

Our traffic model is individual-based: each vehicle will be modeled as an agent. This implies the creation of an adapted environment for them, in terms both suitable to their ontology and adapted to the geographic data we reaped. For that a graph will be built, a *transport graph* that will contain the necessary structures and values.

This first version of our models is only interested in simulating motor vehicle: pedestrians and bicycle are ignored.

The database contains the sense of travel and the traffic restrictions for each topological edge. One oriented edge is created for each sense of direction allowed for motor vehicles.

Edges and vertices of the transport graph are called *elements*. To each element is associated a *data container* and a *vehicle transporter*.

The *data* associated to an edge are for example its geometric length, its number of lanes, its speed limits etc.

The data associated to a vertex is notably the size of the container of its transporter, depending on the number and the sizes of the edges connected to him.

Transporters are non-mobile agents associated to elements. They handle parts of the collective behavior of the vehicles. They will be described in more depth in the following section.

2) Routes in the graph

Mobile agents will try to reach destinations in the graph. As we intend to simulate a realistic traffic of tens of thousand of vehicles, we want to facilitate their computing of their trajectory. To do that, we build a set of "shortest" path stored in the traffic graph.

We compute a weight on the edges that combines different parts of its data: its length, the speed it can reasonably be driven upon, its estimated width based on the number of lanes etc. to model the attractiveness of this edge. After that we compute Dijkstra's algorithms [7] from each vertex to all the others, which we store in each vertex. This data takes

$(\text{numberOfVertices})^2$ bytes of data, which is important, but allows the computation of a good path by an agent in constant time, which is a good thing as hundreds of agents are generated at all time in the simulation (simulating vehicles entering the road network of the simulated urban agglomeration).

III. MOBILE AGENTS OF THE NETWORK

Our agents are mainly so far car agents, trying to go from one place to another.

A. Strategic behavior

Modeling in details the various detailed trajectories of car users is a research problem in itself [8]. Nonetheless we are not interested in who did what or why, but only in what are the fluxes in our network in typical scenarios. When an agent is injected in the network, a starting point and a destination are randomly chosen.

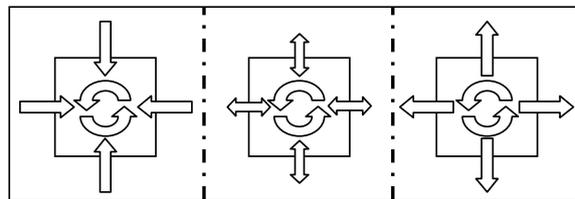


Fig. 1. 3 different scenarios of source/destination pseudo-random choice

This randomness is not necessarily uniform. If we suppose the agglomeration centered on its main town, like the agglomeration of Rouen that we simulated more than others, we can shape different distribution, favoring the likelihood of drawing rather a inner or an outer edge for example. Traffic between 8:00 AM and 9:00 AM for example starts mainly on the border or outside the agglomeration and ends to the same distance to the center (outer edges): we can simulate traffic that do that. When shops close in the town center, we have a traffic that is mainly outer bound, with a more important center generation: we can simulate that. We do not have to know what this car and its driver did in the morning, we don't have to simulate realistically its history, as long as we model the actual traffic fluxes right.

Once the agent knows where it is, and what its destination is, it can use the best paths stored in the traffic network to plan a trajectory. It then drives here, adapting his path through its tactical behavior, and managing its immediate surroundings through its operational behavior.

B. Operational behavior

The planned trajectory of an agent is a succession of edges. Once in an edge the agent tries to drive to its end, the next connection, where it will be able to choose the next planned edge.

When it enters an edge, the agent first chooses a lane if several are available, based on the traffic density in each, with a bias for the rightmost lane. As we have a good geometric description of the lane, the driving behavior is fairly detailed, incorporating the length of the car, its capacity/will to

accelerate and brake, the taste of its driver for long/short safety distance, its taste for following or breaking speed limits etc. All this is incorporated in a driving model inspired by Martin Treiber's Intelligent Driver Model [9]. IDM is a longitudinal traffic model, so we had to expand it to handle multiple lanes and crossroads – the original IDM works for an unlimited one-way, one-lane road – we did not use Treiber's MOBIL lane changing model as it is better adapted to motorways than to urban lane changing decisions.

The data provided by geographic providers does not include right of passage or traffic lights at crossroads. We therefore had to develop our own model aiming at the simulation of crossroads in a heavy traffic.

When a vehicle reaches a crossroad, it slows down and acts according to the fluidity of traffic in the crossroad, in the edge it is currently upon and in the edge it wishes to go to. If they are encumbered, it will more often wait in its way, but it may enter the crossroad and wait here, thus encumbering it (with a more or less strong individual tendency to do so). If the edge it is aiming at has multiple lanes, it will watch both of them, to see if it could fit in one.

C. Tactical behavior

Although vehicles have an original plan, they will adapt it to what they perceive of their environment. When stuck in what they perceive is a jam, they will try to find alternate routes out of it to their destination.

The first method we used is the simpler one. When a vehicle doesn't move enough to its liking – this saturation is variable amongst agents – it tries to take alternate paths as soon as possible, favoring the roads with least dense circulation – although this is not absolute, so as to avoid loops. Once it estimates it's far enough from the jam that sprang this alternate behavior, it resumes using the best path table to find a suitable one to its destination.

The second one is more sophisticated, as it will have uses beyond mere traffic avoidance. Its intelligence is modeled more in the Transporter agents than in the vehicles. Transporters estimate their encumbrance. To do that, they employ direct measure – how many vehicles do they contain over how many vehicles can they contain in average – but also statistics on the proportion of vehicles they contain that are annoyed by the traffic – as described in the first method – and information from the Transporters around them. If based on this they decide they are *encumbered* they also warn the Transporters around them of their perception. This will lower the threshold for them to feel encumbered.

Once *encumbered*, the nodes they are connected to will recompute their best path table, using a huge weight for the encumbered edges. When a vehicle arrives to one of these nodes and wants to go to one of the jammed edges, it is informed of the edge state, and it can recompute a route around it, or take the edge anyway.

This mechanism is also theoretically interesting, as it is an implementation of an emergent property: the interactions of individual behavior affect the behavior of an agent of a higher scale, who alters his behavior, which in turns

transforms the behavior of the lower level vehicles. This reifies the perception an individual driver can have of the state and dynamics of the traffic he is plunged in as a whole.

A Transporter can also be *barred*, because of an accident for example. In that case the same mechanism is used, except that this time circumnavigating is mandatory.

The mechanism of these two states is especially useful in what was the original purpose of our model and its main application: simulating urban important accidents – such as industrial accident – as the modeler can bar the edges it wants as part of his scenario, and see how the traffic adapts to it in simulation real time, as the vehicles discover the evolving road network and fluxes. This is the application that will be developed in the part 2 of our article.

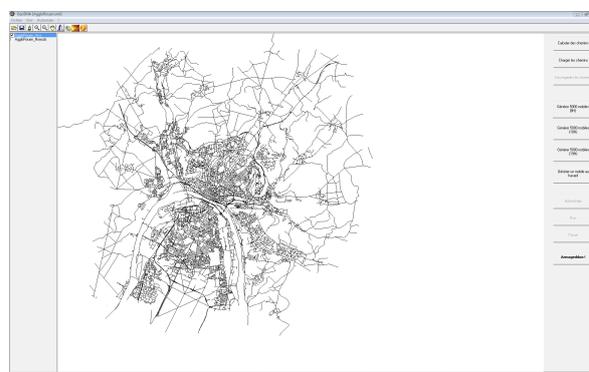


Fig. 2. A screenshot of our simulator, loaded with Rouen agglomeration

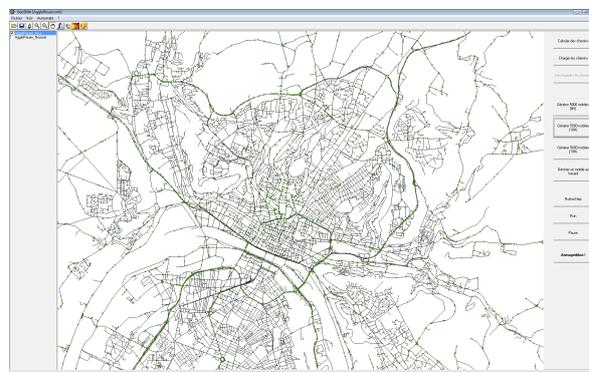


Fig. 3. In green vehicles in an example of running traffic, in a zoom on Rouen itself.

IV. CONCLUSION

This article is more technical than thematic. We tried to write the kind of article we would have liked to read when we started on this work. We have nonetheless done thematic validation.

One of the problems for the validation is that modeling as seldom been taken to such a detail level. This level is necessary because of the multi-level nature of traffic: the decision of one driver can start a jam or jams for thousands of drivers, half a town away, half an hour later, a la butterfly effect. Macro model of fluxes, which dominate the field of

traffic simulation, cannot do that. Their validation for example is often based on the fundamental diagram of traffic flow of a few selected axes, for hourly traffic. We can compute second by second fundamental diagrams of each edge of our network. We can therefore be an order or two of magnitude more precise in our measure, but what to do of all this information? Indeed, if we describe in details the behavior of vehicles, one must not lose sight that they are *not* what we are trying to model, the traffic is what we are trying to model, that is their behavior as a group. We fine-tune individual behavior to have the emerging group behavior right.

What we have ascertained so far is that:

- Most edges comply with the Fundamental Diagram made over 5 minutes of time, most of the time. This remains the case even once the measure-time unused edges are taken out of the count
- We simulated our university home agglomeration of Rouen with up to 50 000 vehicles, and traffic specialists find the results subjectively very satisfying

We tried to compare the results of our simulations with data we had about the traffic of the Rouen agglomeration. The data dated from 2001, while the geographic data we had for the network dated from 2006-2007. The western part of the road network had changed too much during this period for any solid conclusions to be drawn from it, despite superficial resemblances in other parts of the networks. We have contacted the road management of the agglomeration for more recent data.

Finally, as we will describe in further details in part 2 of this article, we have a toolbox to simulate urban industrial accident and its effect on traffic with a realistic level of traffic volume.

REFERENCES

- [1] May, Adolf. Traffic Flow Fundamentals. Prentice Hall, Englewood Cliffs, NJ, 1990.
- [2] <http://transims-open-source.org/>
- [3] <http://www.matsim.org/>
- [4] ESRI Shapefile Technical Description, ESRI White Paper, July 1998
- [5] F. P. Preparata, M.I. Shamos - *Computational Geometry: An introduction*- Springer Verlag, 1985
- [6] NAVTEQ's NAVSTREETS Street Data Reference Manual v2.0 -- 16 June 2006
- [7] E. W. Dijkstra: "A note on two problems in connexion with graphs". In *Numerische Mathematik*, 1 (1959), S. 269–271
- [8] A. Banos, T. Thévenin (2008) "Création de champs de potentiel et simulation d'itinéraires à partir de l'enquête ménages-déplacement" In: *Information Géographique et Dynamique Urbaine*, Volume 1 : Analyse et simulation de la mobilité des personnes Edited by: Marius Thériault et François Des Rosiers. Hermes
- [9] A. Kesting, M. Treiber, and D. Helbing (2007) "General lane-changing model MOBIL for car-following models" *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1999/2007, pp. 86-94. DOI 10.3141/1999-10

Space-Time Self-organization for the Dynamic VRPTW

Besma Zeddini¹ Mahdi Zargayouna² Adnan Yassine¹ Moncef Temani³

¹ LE HAVRE University,
LMAH Laboratory,
25, rue Philippe Lebon,
76600 Le Havre, France.

² INRETS Institute,
GRETIA Laboratory,
2, Rue de la Butte Verte,
93166 Noisy le Grand, France.

³ ENSI School of Computer Science,
LI3 Laboratory,
Campus Universitaire de la Manouba,
2010 La Manouba, Tunisia.

{besma.zeddini, adnan.yassine}@univ-lehavre.fr, zargayouna@inrets.fr, moncef.temani@fst.rnu.tn

Abstract—Vehicle Routing problems are highly complex problems for which different Artificial Intelligence techniques have been used. In this paper, we propose an agent-oriented self-organization model for the dynamic version of the problem with time windows. Our proposal is based on a space-time representation of the Action Zones of the agents, which is able to maintain a good distribution of the vehicles on the environment. This distribution answers the objective of the dynamic problem, since it allows the agents to take their decisions while anticipating future changes in the system's parameters.

Index Terms—Self-organization, Multi-agent Systems, Applications, Planning, Scheduling

I. INTRODUCTION

The deliveries of goods to stores, the routing of school buses, the distribution of newspapers and mail etc. are instantiations of theoretical problems called the Vehicle Routing Problems (VRP). The VRP have been an intensive research area in the last decades because of their large applicability in real life problems. Several constrained variants were proposed in order to meet specific operational applications. Constraints concern vehicle capacity, time restrictions, requests configuration etc. One of the most widely studied problem is the time (and capacity) constrained version: the Vehicle Routing Problems with Time Windows (VRPTW henceforth). VRPTW and their variants (Pickup and Delivery Problem with TW, Dial A Ride Problem with TW ...) are hard combinatorial optimization problems met in many industrial applications. It can be formally stated as follows:

Let $G = (V, E)$ be a graph with node set $V = N \cup 0$ and edge set $E = (ij) | i \in V, j \in V, i \neq j, N = 1, 2, \dots, n$ is the customer set with node 0 is the depot. With each node $i \in V$ is associated a customer demand $q_i (q_0 = 0)$, a service time $s_i (s_0 = 0)$, and a hard service-time window $[e_i, l_i]$ i.e. a vehicle must be at i before l_i but can be at i before e_i and must wait until the service starts. For every edge $(i, j) \in A$, a distance $d_{ij} \geq 0$ and a travel time $t_{ij} \geq 0$ are given. Moreover, the fleet of vehicles is homogeneous and every vehicle is initially located and end its route at a central depot. Each customer demand is assumed to be less than the vehicle capacity Cap . The objective is to find an optimal set of routes (with the minimal cost) such that:

(1) All routes start and end at the depot;

(2) each customer in N is visited exactly once within its time window;

(3) the total of customer demands for each route cannot exceed the vehicle capacity Cap . The performance criteria are in general (following this order):

1. The number of vehicles used,
2. the total distance traveled,
3. the total waiting time.

Since the problem is NP-hard, exact approaches are only of theoretical interest, and heuristics are performed in order to find good solutions, not necessarily optimal, within reasonable computational times. VRPTW can be divided into two sets: static problems and dynamic problems. In the static problems, all the problem data are available before the start of the execution. In order to meet the reactivity requirement of operational applications, the most promising category of problems is the dynamic version where some data could be not initially available, and especially the amount of available customers, before the start of execution. Indeed, operational vehicle routing problems are rarely fully static, and we can reasonably say that today a static system cannot meet the mobility needs of the users. In operational settings, and even if the whole number of customers to be served is known, there is still some elements that makes the problem dynamic. These elements include breakdowns, delays, no-shows, etc. It is thus always useful to consider a problem that is not fully static.

A multi-agent modeling of the dynamic VRPTW is relevant for the following reasons. First, since it's a hard problem, choosing a design allowing for processing distribution can be a solution to propose short answer times to customers requests. Second, with the technological developments, it is reasonable to consider vehicles with onboard calculation capacities. In this context, the problem is, actually, distributed and necessitates an adapted modeling to take profit of the onboard equipments of the vehicles. Finally, the consideration of a multi-agent point of view allows to envision new measures, new heuristics, not envisaged by centralized approaches.

In this paper, we propose a distributed version of an insertion heuristic with a special focus on the insertion cost of a customer in the route of a vehicle. Several multi-agent proposals in the literature have been proposed to distribute insertion heuristics, consisting in inserting the customers fol-

lowing their revealing order, and by choosing the vehicle that has to make the minimal detour to visit the new customer. But Very few proposals in the literature propose new measures of the insertion cost of a customer in the route of a vehicle, instead of the detour. In the present work, we do propose such a new measure, based on a space time representation of *Vehicle* agents' action zones. The objective is to allow the MAS to self-adapt exhibiting an equilibrated distribution of his *Vehicle* agents, and to decrease this way the number of vehicles mobilized to serve the customers.

The remainder of this paper is structured as follows. In section II, we briefly discuss previous proposals for the dynamic VRPTW w.r.t our approach. Section III presents the architecture of the MAS that we propose. In the section IV, we detail the space-time representation of the Action zones of the vehicles and its use as a measure for the insertion decision of the customers. We report our experimental results in section V before to conclude.

II. RELATED WORK

As we said in the introduction, exact approaches cannot meet operational settings, and interested readers in the optimization approaches can refer to [2] for a survey. In fact, most of the proposed solution methods are heuristic or metaheuristic methods. Artificial Intelligence metaheuristics have shown better performances than heuristics in average with benchmarking problems; for instance local search [8], genetic algorithms [5], simulated annealing [1], tabu search [11], ant colony [4] etc. (see [12] for various heuristics with artificial intelligence based techniques). Note that these approaches generally need several parameters, which values are closely dependent of the input data and are set after several system runs. That is why we are trying, in order to assess the impact of our approach, to minimize the necessary parameters to run the system.

These approaches perform well with static problems. However, the final solutions they provide are very constrained (tight spatiotemporal gap between customers in a route), and the insertion of new requests probably leads to new vehicles' creation, or to incoming requests rejection. Generally speaking, two approaches can be envisioned to deal with dynamic requests. We can solve a static problem every time a change in the problem data occurs, which is obviously very expensive and is not realistic in operational settings. The other approach, which is in fact one of the most popular approaches in solving the dynamic versions is insertion methods. Insertion methods are greedy algorithms, meaning that they do not cancel a previous decision to insert a certain request in a specific route. Two versions are possible, sequential and parallel insertion. Parallel in this context means that several routes are created in parallel, in opposition with the sequential version which constructs only one route until no customers can be inserted. In [7], the authors show that parallel insertion procedures outperform sequential approaches.

The ADART [3] system is based on an *a priori* geographical segmentation of the network, by allowing each segment to some vehicles. For the dynamic management of customers

arrival, the communication is established between the customer that has called the service and the onboard computer of the vehicle. Vehicles of the same regional zone negotiate the insertion of the customer, and the one with the minimal cost is chosen. In [13] and in [6], the authors propose a multi-agent architecture. The principle is the same: distribute an insertion heuristic, followed by a post-optimization step. In-Time [6] is a system composed of *Customer* agents and *Vehicle* agents. The *Customer* agent announces himself and all the *Vehicle* agents calculate his insertion cost in their routes. As usual, the *Customer* agent selects the cheapest offer.

From a protocol and an architecture point of view, our system sticks with the systems we have just described, since we propose a distributed version of insertion heuristics. But the traditional insertion cost of a customer in the route of a vehicle, based on the incurred detour of the vehicle, is the measure that is widely used. We propose a new insertion cost measure, focused on the space-time coverage of the vehicles that aims at counterbalancing the myopy of the traditional measures by privileging an insertion process that is future-centered.

III. MULTI-AGENT SYSTEM FOR THE DYNAMIC VRPTW

Our system is composed of a dynamic set of agents which interact to solve the dynamic VRPTW. A solution consists of a series of vehicles routes, each route consist of a sequence of customers with their associated visit time. We define three categories of agents. *Customer* agents, which represent users of the system (persons or goods), *Vehicle* agents, which represent vehicles in the MAS and *Interface* agents which represent an access point to the system (Web server, GUI, simulator, etc). When a user logs in the MAS, the data he provides are verified (existing node, valid time windows, etc.) and, if the data are correct, a *Customer* agent representing him and described by the data he provided is created.

In [16], we have designed, implemented and compared three possible architectures to model the dynamic VRPTW problem: a centralized architecture, a decentralized architecture and a hybrid architecture. We present them briefly in the following sections.

A. Centralized architecture

In this architecture, all the requests are treated by the same "agent". He has all the required information about each vehicle and each customer: the occupancy rate of vehicles, their current positions and the traffic conditions in real time. Having all these information, he assigns to each customer the most appropriate vehicle for the service, i.e the one having the minimal overcost related to the customer insertion. Fig. 1 illustrates this architecture, in which, besides the three agents described above, we add a *Planner* agent which represent the decision-making center, he has in charge the routes computation and vehicles notification of his decision.

The scenario that we have proposed to study is the following: At a given moment, a user interacts with the *Interface* agent, which creates a *Customer* agent to represent him in the system. Once created, the *Customer* agent sends his request to the *Planner* agent which tries to insert him in each vehicle's

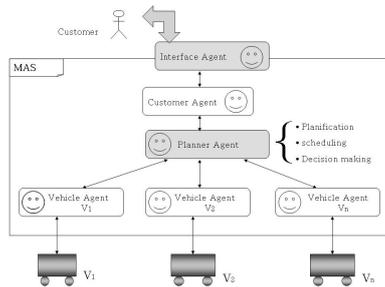


Fig. 1. Centralized architecture

route, and retains the one with the minimal additional cost. If there is no *Vehicle* agent which can insert the customer, a new vehicle is created. Finally, the *Planner* agent sends the current route to each vehicle and informs the *Customer* agent of his vehicle and his visit times. The *Vehicle* agents don't execute any operation, thus, they merely receive their current route and update their information. The centralized approach poses obvious problems. Indeed, sequential treatment of the customer requests slow down the system response time, which goes against the requirement of fast response to dynamic customers. Moreover, the failure of the *Planner* agent leads to a blackout at the global level. Nevertheless, the centralized architecture has the advantage of minimizing communications and updates at the agent level.

B. Decentralized architecture

The decentralized architecture is illustrated in Fig. 2. In this architecture, there is no bottleneck for the routes computation. Each *Vehicle* agent tries to insert the new customer in his route, proposes a cost for its insertion, and the vehicle with the minimal additional cost is selected. At each appearance of a new customer, the *Customer* agent broadcasts his request to all the vehicles in the system. *Vehicle* agents exchange their overcosts via messages. Each *Vehicle* agent compares his own cost with other agents' costs, and stops bidding if the cost that is being offered to him is better than hers. Finally, the winner agent (the *Vehicle* agent with the minimal insertion cost) communicates with the *Customer* agent and both (the *Vehicle* agent and *Customer* agent) update their information. This architecture offers the advantage of a distributed processing and to be fault-tolerant. However, the communication costs explode with this architecture: the number of messages exchanged between *Vehicle* agents is of quadratic complexity.

C. Hybrid architecture

The hybrid architecture (cf. Fig. 3) is a compromise between the centralized and the decentralized approach. A new agent *Dispatcher* is inserted between the *Customer* and *Vehicle* agents and he has the role of dispatching the customer's request, collecting bids from the *Vehicle* agents and choosing the one offering the minimal cost. The process describes a CNP (*Contract Net Protocol*) [9] where, in each occurrence

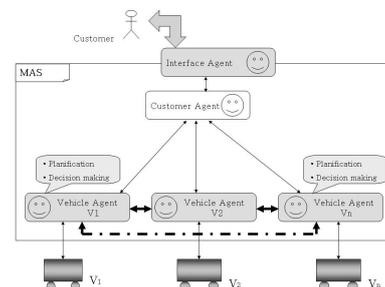


Fig. 2. Decentralized architecture

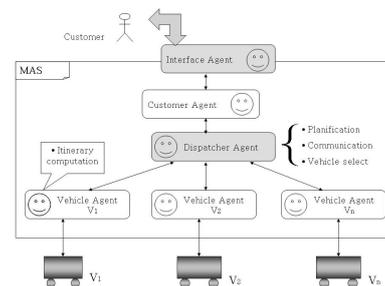


Fig. 3. Hybrid architecture

of a *Customer* agent, the *Dispatcher* agent receives a set of proposals and selects those with a minimal cost.

Our proposal is based on this architecture. In the previous description, we use the additional cost related to the customer insertion to make a decision to include the new customer in the vehicle route. This cost is function of the detour made by the vehicle to integrate the new customer. We propose the variation of *Vehicle* agents' action zones as a cost for customers' insertion, as an alternative to the traditional measure.

IV. SELF-ORGANIZATION MODEL

The self-organization model that we propose has the objective of allowing the *Vehicle* agents to cover a maximal space-time zone of the transportation network. A space-time pair (n, t) - with n a node and t a moment - is said to be "covered" by the *Vehicle* agent v if v can be in the node n at moment t . In the context of the dynamic VRPTW, to maximise the space-time coverage of the *Vehicle* agents is to give them the maximum chances of satisfying the demand of a new customer in the future. This measure breaks with the logic of traditional measures which focus on the increase of the traveled distance, neglecting the impact of the current decision on future insertions.

A. Action Zone of a Vehicle Agent

Following the description provided above, the *Dispatcher* agent chooses between several *Vehicle* agents the one with the minimal proposed insertion cost. The systems that are based on this kind of heuristics - said insertion heuristics - utilise generally the measure used by Solomon [10] as an

insertion cost. This measure consists in inserting the customer that result in a minimal increase of the general cost of the vehicle (function of the detour to be made by the vehicle). This measure is simple and is the most intuitive but unfortunately it suffers from an obvious drawback, because the insertion of the current customer might induce the insertion impossibility of a great number of future customers. Its problem is that it generates vehicle routes that are very constrained in time and space, i.e. routes that offer very few insertion possibilities between every pair of adjacent customers in the route of a vehicle. The appearance of new customers might mobilise new vehicles to serve them. With the modeling of *Vehicle* agents' action zones, we propose a new measure of the insertion cost of a customer in the route of a vehicle, and therefore a new choice criterion between candidate vehicles for the same customer. We propose a measure which objective is to choose the *Vehicle* agent for whom "the decrease in the probability of participating to future insertions is minimal". We use the variation of the action zone of the *Vehicle* agent as an insertion cost of a customer in his route.

B. Intuition of the Action Zones

Consider a *Vehicle* agent v that has an empty route. In order for this agent to be able to insert a new customer c described by n a node, $[e, l]$ a time window, s a service time, and q a quantity, l has to be big enough to allow v to be in n without violating his time constraints. More precisely, the current time t , plus the travel time between the depot and n has to be less or equal to l (cf. Fig. 4).

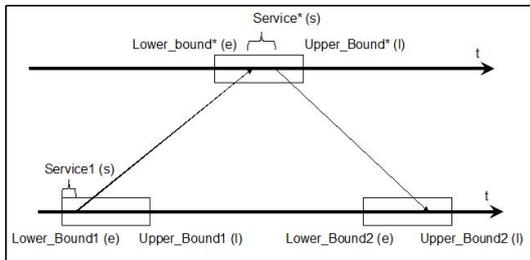


Fig. 4. Feasible insertion

Starting from this observation, we define the Action Zone of a *Vehicle* agent as the number of potential customers that satisfy this constraint. To do so, we define "the physical environment" as a set of pairs $\langle node, time \rangle$, and the Action Zone of a *Vehicle* agent as the number of pairs that remain valid given his current route. When a *Vehicle* agent inserts a customer in his route, his Action Zone is recomputed, since some $\langle node, time \rangle$ pairs become not valid because of his insertion. The associated cost to an offer from a *Vehicle* agent v for the insertion of a *Customer* agent c corresponds to the hypothetical decrease of the action Zone of v following the insertion of c in his route.

The idea is that the chosen *Vehicle* for the insertion of a customer is the one that loses the minimal chance to be candidate for the insertion of future customers. Thus, the

criterion that is maximized by the society of *Vehicle* agents is the sum of their Action Zones, i.e. the capacity that the MAS has to react to the appearance of *Customer* agents, without mobilizing new vehicles.

To illustrate the Action Zones and their dynamics, we present the version of the measure that is related to an Euclidean problem, i.e. where travel times are computed following the Euclidean metric. The following paragraphs detail the measure as well as its dynamics.

C. The Computation of Action Zones

In the Euclidean case, the transportation network is a plane, and the travel times between two points i (described by (x_i, y_i)) and j (described by (x_j, y_j)) is equal to

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Therefore, if a vehicle is in i at the moment t_i , he cannot be in j earlier than $t_i + \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

We can compute at any time, from the current position of a vehicle, the set of triples (x, y, t) where he can be in the future. Indeed, considering a plane with an X-axis in $[x_{min}, x_{max}]$ and a Y-axis in $[y_{min}, y_{max}]$, the set of space-time positions is the set of points in the cube delimited by $[x_{min}, x_{max}], [y_{min}, y_{max}]$ and $[e_0, l_0]$ (e_0 and l_0 are the minimal and maximal values for the time windows). Consider a vehicle in the depot (x_0, y_0) at t_0 . The set of points (x, y, t) that are accessible by this vehicle are described by the following inequality:

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} \leq (t - t_0)$$

The (x, y, t) satisfying this inequality are those that are positioned inside the cone \mathcal{C} of vertex (x_0, y_0, t_0) and with the equation $\sqrt{(x - x_0)^2 + (y - y_0)^2} = (t - t_0)$ (c.f Fig. 5). This cone represents the Action Zone of a *Vehicle* agent in

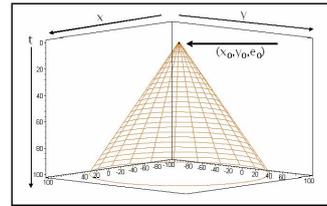


Fig. 5. Initial Action Zone

the Euclidean case. It represents all the possible space-time positions that this *Vehicle* agent is able to have in the future.

We use the Action Zone of the *Vehicle* agents when a *Customer* agent has to choose between several *Vehicle* agents for his insertion. We have to be able to compare the Action Zones of different *Vehicle* agents. To do so, we propose to quantify it, by computing the volume of the cone \mathcal{C} representing the future possible positions of the vehicle:

$$Volume(\mathcal{C}) = \frac{1}{3} \times \pi \times (l_0 - e_0)^3$$

This is the quantification of the initial Action Zone of any new *Vehicle* agent joining the MAS. When a new *Customer*

agent appears, a *Vehicle* agent computes his new Action Zone, the cost that he proposes to the *Dispatcher* agent is the difference between his old Action Zone and his new one. The new Action zone computation is detailed in the following paragraph.

D. Dynamics of the Action Zones

Consider a customer c_2 (of coordinates (x_2, y_2) and with a time window $[e_2, l_2]$) that joins the system, and suppose that v is temporarily the only available *Vehicle* agent of the system and has an empty route. The agent v has to deduce his new space-time action zone, i.e. the space-time nodes that he can still reach without violating the time constraints of c_2 . The new action zone answer the following questions: "if v had to be in (x_2, y_2) at l_2 , where would he have been before? And if he had to be there at e_2 where would he be after $e_2 + s_2$?". The triples (x, y, t) where the *Vehicle* agent can be before visiting c_2 are described by the inequality [a], and the triples (x, y, t) where he can be after visiting c_2 are describe by the inequality [b].

$$\sqrt{(x - x_2)^2 + (y - y_2)^2} \leq (l_2 - (t + s)) \quad [a]$$

$$\sqrt{(x - x_2)^2 + (y - y_2)^2} \leq (t - (e_2 + s_2)) \quad [b]$$

The new Action Zone is illustrated by the Fig. 6: the new measure consists in the intersection of the initial cone \mathcal{C} with the union of the two new cones described by the inequalities [a] and [b] (denoted respectively by \mathcal{C}_1 and \mathcal{C}_2). The new measure of the Action Zone is equal to the volume of the intersection of \mathcal{C} with the union of \mathcal{C}_1 and \mathcal{C}_2 . The complete computation of the volume of the intersection of these two cones is reported in [15].

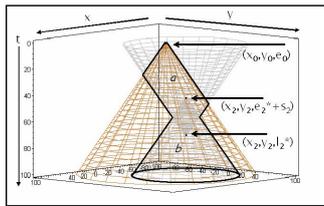


Fig. 6. Space-Time Action Zone after the insertion of c_2

The cost of the insertion of a customer in the route of a vehicle is equal to the measure associated with the old Action Zone of the vehicle minus the measure of the new Action Zone, after the insertion of the customer. The quantity measured represents the space-time positions that the vehicle cannot have anymore, if he had to insert this customer in his route. The retained *Vehicle* agent to visit a given customer is the one for which the insertion of the customer causes less loss in his space-time Action Zone. This corresponds to choosing the vehicle that loses the minimal possibilities to be candidate for future customers.

The physical environment in the non-Euclidian case is not a space-time cube, but a space-time network. In [14], we propose

Δ Distance		
	Number of vehicles	Distance
25	6.4	637.1
50	10.7	1203.7
100	19.1	1968.4
Δ Action Zones		
	Number of vehicles	Distance
25	6.3	679.3
50	10.6	1286.7
100	18.8	2149.7

TABLE I
EXPERIMENTAL RESULTS WITH ON CLASS R1 WITH 25, 50 AND 100 CUSTOMERS

a method for the self-organization of the MAS in the general case, where each *Vehicle* agent associates with each node n of the network an interval defining the moments where he can be in n , which represents his Action Zone.

V. RESULTS

Marius M. Solomon [10] has created a set of different problems for the dynamic VRPTW. In the Solomon *benchmarks*, six different sets of problems have been defined : C1, C2, R1, R2, RC1 and RC2. The customers are uniformly distributed in the problems of type R, clustered in the problems of type C, and a mix of the two is used in the problems of type RC. We have used instances from both the classes R and C.

When the size of the fleet of vehicles is fixed in advance, the central concern with a dynamic VRPTW system is the amount of rejected requests, which should be limited. However, since we create vehicles dynamically when no vehicles can serve a new customer, our system does not reject any request; hence, our central concern becomes the size of this fleet. We have implemented a system which behavior is similar to ours. The only difference is the cost computed by a *Vehicle* agent, which is equal to the increase of the traveled distance, and not the loss of Action zone as we do. Table I reports the results with files of the class R1 where we consider successively 25, 50 and 100 customers, while Table II reports the results with files of the class C1.

The results show, with both classes of the problem, that the use of our measure mobilizes less vehicles than the traditional measure, and this is the case whatever the number of considered customers. Note that we had a different result - the traditional measure behaving better than our measure w.r.t the number of vehicles used - with only one file, the file 5 with 100 customers of the C1 class. This result validates the intuition of the measure which consists on the maximization of future insertion possibilities for a *Vehicle* agent. However, since our measure focuses exclusively on insertion feasibilities, the total distance traveled by all the vehicles with our measure is superior to the distance traveled with the traditional measure. We think that a compromise between the two measures, e.g. a weighted sum of the increase of the distance and the loss of Action Zones is able to give better results w.r.t the two criteria.

Δ Distance		
Number of vehicles		Distance
25	3.4	316.6
50	6	671.2
100	12.1	1601.3
200	21.6	6315.5
Δ Action Zones		
Number of vehicles		Distance
25	3.3	347.9
50	5.9	731.5
100	11.9	1774.4
200	21.4	6979.8

TABLE II
EXPERIMENTAL RESULTS WITH ON CLASS C1 WITH 25, 50 AND 100
CUSTOMERS

VI. CONCLUSION ET PERSPECTIVES

In this paper, we have proposed an agent-oriented self-organization model for the dynamic VRPTW based on the agents' action zones. The action zones of the *Vehicle* agents reflect their space-time coverage of the environment. We use the variation of these action zones as a new metric between *Vehicle* agent to reduce the myopic behavior of traditional metrics. By optimizing the space-time coverage of the environment by the *Vehicle* agents, our model allows the MAS to self-adapt by exhibiting an equilibrated space-time distribution of the *Vehicle* agents, and to lessen this way the number of vehicles mobilized to serve the customers. Our current works are oriented towards taking into account historic data of customers requests on the network nodes. We use these data as a weighting of the action zones of the *Vehicle* agents that concern the nodes frequently requested, and this to make them converge towards high density zones.

REFERENCES

- [1] Z. J. Czech and P. Czarnas, "A parallel simulated annealing for the vehicle routing problem with time windows," in *Proceedings of the 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing*, Canary Islands (Spain), 2002, pp. 376–383.
- [2] M. Desrochers, J. Lenstra, M. Savelsbergh, and F. Soumis, "Vehicle routing with time windows: Optimization and approximation," in *Vehicle Routing: Methods and Studies*. Amsterdam (Netherlands): B.L. Golden, A.A. Assad (Eds.), 1988, pp. 65–84.
- [3] R. B. Dial, "Autonomous dial-a-ride transit introductory overview," *Transportation Research Part C: Emerging Technologies*, vol. 3, pp. 261–275(15), October 1995.
- [4] L. M. Gambardella, E. D. Taillard, and G. Agazzi, "MACS-VRPTW: A multiple ant colony system for vehicle routing problems with time windows," *New Ideas in Optimization*, pp. 63–76, 1999.
- [5] J. Homberger and H. Gehring, "Two evolutionary metaheuristics for the vehicle routing problem with time windows," *INFOR*, vol. 37, no. 6, pp. 297–318, 1999.
- [6] R. Kohout and K. Erol, "In-Time agent-based vehicle routing with a stochastic improvement heuristic," in *Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence (AAAI'99/IAAI'99)*. Menlo Park, CA (USA): AAAI Press, 1999, pp. 864–869.
- [7] F. Liu and S. Shen, "A route-neighborhood-based metaheuristic for vehicle routing problem with time windows," *European Journal of Operational Research*, vol. 118, pp. 485–504, 1999.
- [8] Y. Rochat and E. Taillard, "Probabilistic diversification and intensification in local search for vehicle routing," *Journal of Heuristics*, vol. 1, pp. 147–167, 1995.
- [9] R. G. Smith, "The contract net protocol: High-level communication and control in a distributed problem solver," *IEEE Trans. on Comp.*, vol. C-29, no. 12, pp. 1104–1113, December 1980.
- [10] M. Solomon, "Algorithms for the vehicle routing and scheduling with time window constraints," *Operations Research*, vol. 15, pp. 254–265, 1987.
- [11] E. D. Taillard, P. Badeau, M. Gendreau, F. Geurtin, and J.-Y. Potvin, "A tabu search heuristic for the vehicle routing problem with time windows," *Transportation Science*, vol. 31, pp. 170–186, 1997.
- [12] K. Tan, L. Lee, and K. Ou, "Artificial intelligence heuristics in solving vehicle routing problems with time window constraints," *Engineering Applications of Artificial Intelligence*, vol. 14, pp. 825–837, 2001.
- [13] S. R. Thangiah, O. Shmygelska, and W. Mennell, "An agent architecture for vehicle routing problems," in *Proceedings of the 2001 ACM symposium on Applied computing (SAC '01)*. New York, NY (USA): ACM Press, 2001, pp. 517–521.
- [14] M. Zargayouna, "Une représentation spatio-temporelle de l'environnement pour le transport à la demande," in *Atelier: Représentation et raisonnement sur le temps et l'espace, Plate-forme AFIA 2005*, Nice (France), 2005.
- [15] —, "Modèle et langage de coordination pour les systèmes multi-agents ouverts. application au problème du transport à la demande," PhD Dissertation, University of Paris-Dauphine, Paris (France), 2007.
- [16] B. Zeddini, A. Yassine, M. Temani, and K. Ghédira, "Collective intelligence for demand-responsive transportation systems: a self organization model," in *Proceedings of the 8th international conference on New technologies in distributed systems (NOTERE'08)*. New York, NY (USA): ACM Press, 2008, pp. 1–8.

Besma Zeddini is Ph.D Student in le Havre University (France) and in the ENSI School of Computer Science (Tunisia). She is also a Computer Science Lecturer in Evry University (France). Her research interests are Multi-agent Systems and Vehicle Routing Problems.

Mahdi Zargayouna is Research Assistant in the French National Institute for Transport and Safety Research (INRETS) and currently Visiting Researcher in TU-Delft University (The Netherlands). His research interests are the formal specification of Multi-agent Systems and transportation problems such as Traveler Information and Dial A Ride Problems.

Adnan Yassine is Professor in Le Havre University and is Head of the LMAH Laboratory. His research interests are centered on the numeric analysis, convex and nonconvex analysis, convex and nonconvex optimization, global optimization, etc. with varied application domains such as image processing, finance and transportation problems.

Moncef Temani is Head of the Computer Science Institute (ISI) in Tunis (Tunisia) and member of the LI3 Laboratory. His research interests concern mainly Artificial Intelligence Techniques and Multi-agent Systems for routing and scheduling. He's also interested in Web Services and Security.

French Ports Reform: reasons and perspectives

Sidi Mohamed. OULD MOHAMED MOCTAR

PhD Student

Laboratory: Centre for Study and Research in Economics and Logistics Management

University of Le Havre

25 rue Philippe Lebon – 76600 France

E – mail: sidimohamedammah@yahoo.fr

I. Introduction

Globalization returns the world small a village and creates interdependences between the nations. This interdependence the ones with the others poses the problem of investment which includes a question of management of transport and logistics.

The economic globalization, consequence of this globalization, exerts strong pressures on the harbor authorities to answer with more flexibility, and in real-time at the worldwide markets which change quickly, and to acquire a modern system of transport and telecommunications which facilitates the international business.

The port is a link of a chain of transport. It is thus a node of a total supply chain. Its interaction with this chain requires efforts as regards investment. The port deals with the reception and the operation of the ships, the transshipment, storage and close and post routing of the traffic. The ships and the cargo tend to increase their size, the port must have the characteristics nautical and terrestrial suitable as well as rather vast storage spaces. All its activities require an installation of territory whose cost is often very high, and make that the harbor authority collaborates with other private partners in the research of the economic optimum.

The Course of the Accounts in its 2006 report and the stressed the overall degradation of market shares of the French ports and lack of competitiveness compared to competing ports in the North Sea and the Mediterranean and particularly in terms of container. Then a French port reform is necessary to ensure their competitiveness.

In this context the Supervisory Board at the Port Harbor has adopted the strategic project of Greater seaport that will be renamed "Havre Port. The strategic project comes from the port reform passed by parliament last summer. The project wants to be very ambitious; it sets to reach the cap of 6 million containers in 2015. To achieve this objective, the project foresees the realization of investment. "By 2013, underlines Laurent Casting, the chairman of the

Board, the investment represents a total of around 700 million euro" (LE HAVRE PRESSE, Friday April 10 2009).

In this context why is this reform necessary? What are its main features? And how to explain theoretically the lack of competitiveness of French ports?

II. Harbor Reform

In his report/ratio of January 15th, 2008 on the harbor reform of the ports, the Minister for Ecology, the Development and Durable Installation (MEDAD) underlined the total degradation of market shares of the French ports and the insufficiency of their competitiveness compared to the large competing ports of the Mediterranean and North Sea and more particularly in terms of container.

Also the Course of the Accounts in his ratio of 2006 and the General advice of the Highways Departments addressed statistics alarming:

- Between 1989 and 2006, the market share of the French ports dropped by 3,9% passing from 17,8% to 13,9%;
- The market share of the French ports in terms of containers went from 11,7% to 6,2%; whereas in Europe the market of containers knew a growth of more than 5%;
- Particularly the market share of wearing of Marseilles, which according to the report/ratio is the largest French port in tonnage, dropped by 8% and more particularly by 13,3% in the containers;
- Compared to the European competitor ports, the French ports record the lowest productivity except for the new terminals of “Port 200”. For a gantry, when the wearing of Marseilles records 46000 annual movements, Valence record 76 000, whereas Antwerp records between 100 and 150 000 movements.

A reform of the ports is essential. For Jean Chapon (2007) four reasons justify a reform of the ports:

- Effects of globalization: fierce competition to which are exposed the French ports obliges them to exploit the couple costs/quality.
A charger when it chooses a port, it seeks before all that which offers “the totality of the cost/quality of the routing from beginning to end”. The French ports should not be satisfied with their geographical proximity with national industries and trade, because the pre one and terrestrial post-routing constitutes a factor among others of the cost total of the carriage of the goods;

- In terms of operation of the ports, the industrial activities and commercial must be entrusted to the private ones, because they are better to manage by professionals. The harbor authority when with it, deals then “with the mission of organizing the public service which constitutes the harbor passage for the international economy”. this redistribution of mission makes it possible to organize in an effective way the harbor passage or the harbor authority will ensure not only one role of supervisor but also of actor in the investment in infrastructure; and the private ones carry out the industrial and commercial acts;
- To act vis-a-vis the constraints of space for the ports: the French ports must obligatorily adapt vis-a-vis the change and the evolution of technology. The ships are increasing harls and increasingly specialized. That requires on behalf of the harbor authorities to acquire new tools and to adapt its infrastructure, but also to have the land fields allowing infrastructures creation or the extension of those existing. The harbors authorities have also needs for space to open new sites or to extend the sites exist. But they are obliged to face constraints environmental;
- The rarefaction of the public resources obliges the States to call upon private for the realization of expensive equipment in highly powerful terms of tools but also in terms of infrastructure like the terminals and the quays. As a Mr. Chapon (2007), the harbor authority must have the possibility “of utilizing private operators who will deal with the financing of the equipment, or, at the very least, can quickly begin in Partnership Public Private, which will be often the only way that the private companies launch out in financially heavy operations of creation of terminals”. for Li the recourse to private is all the more necessary as the State does not fill its engagements envisaged by the law of 1965 which stipulates that the State must take part to a total value of 80% in the financing of the quays, and of 60% in machines and repair and 100% in the maintenance of the maritime accesses.

Mr. Chapon (2007) concludes that:

- The harbor authority must concentrate its financial means and the participation of the State on the operations of common interests (works of access, basins...);
- To continue to call upon the territorial collectivities which are concerned with their activities;

- To encourage the private ones to invest in the ports.

The reform is articulated with the turn of four main axes which is appropriate to specify the methods of their operation:

1. The heart of the reform is the transfer of the activities of tools: it is a question of transferring to private as well as the material of tools as the employees working on this material. Such a transfer will make the port more effective because it will relate to the whole of its financial means on the infrastructure. The private one on its side will control the superstructure.
2. Centring of the ports on specific missions namely the kingly missions as the police force of the port. Infrastructures and works of access and their maintenance, the installation of the harbor field. The port will assume the responsibility for service road terrestrial (MEDAD, Reform of the ports, 2008).
3. Modernization of the governorship of the ports: it is about the creation of a council of sustainable development associating the whole of the recipients of the ports (economic actors, communities, employee representatives, ONG). This council aims at the implementation “of the integrated policies fascinating of account the economic aspects, social and environmental of the development of the ports” (MEDAD, Reform of the ports, 2008).
4. To establish a capital spending program:
 - Maintenance of the maritime accesses: the State promises to reinforce:
 - ✓ Its participation in the financing of the maintenance of the maritime accesses;
 - ✓ Effort of productivity for the dredging of the maritime accesses through a multiannual program.
 - The investment in the ports: The objective is to align the French ports with their competitors of Northern Europe which invest massively in the infrastructure. Thus the reform envisages a policy of investment by each harbor place within the framework of what it calls “strategic project” aiming at developing “new

infrastructure of international scale by 2020". the reform envisages the completion of port 2000, the project Fos 3XL and the future extensions of the container terminals of Marseilles-Fox.

It should be recalled that in the contract of project 2007-2013, the State already envisaged to take part in the harbor investment in height of 245 million euro.

- The terrestrial investment of service road: it is a question of improving the river service roads and particularly that of the port 2000. When to the railway service road, after the ports took its control, the State provides the required investments to the implementation of an optimized exploitation of the harbor railways. But to carry out all that the State needs the funds. The port, taking into account the promises of the State, must work out a strategy of financing.

III. Microeconomic base of the specificity of infrastructure assets

The question is to know if the harbor infrastructures are specific assets or not?

According to J. M. Josselin (1997) the infrastructures are assets little redeployable and thus specific. It justifies its point of view by:

- the weak development of the markets secondary relative or associated with its infrastructures. That is with the difficulty of resale of the assets which is due to the non adaptability of the assets to external uses at the firm;
- the specific features to its infrastructures return them specific assets;
- the design of its infrastructures is related to very specific projects;
- often its projects of infrastructure are not numerous what compromised a reassignment of the infrastructures is necessary alternative processes.

In terms of investment, the irreversibility is the result of the combination of three factors (McDonald and Siegel, 1986):

- the company must pay an irremediable cost for the implementation of the project;
- the flow of income associated with the project is random;
- the investment can undergo a delay of with the need for extra information.

The decision of investment is then irreversible and the credit is not redeployable and thus the costs of the infrastructure are irremediable.

Moreover “like the main part of the heavy investments, the projects of infrastructure generate particularly high initial costs, which increases the weight of the irreversibility at the time of decision making. However these irremediable costs are often dubious, sometimes even more than the future net incomes of the project” (J.M.Josselin, 1997).

IV. Costs of the harbor investments

According to Michel Loir (1980) these costs can be subdivided into:

- **Capital costs in infrastructure:** it is about digging of channels, construction of mole, construction of the quays, coating of surfaces, creation of the roadway systems;
- **Capital costs in superstructure:** its costs relate to the machines of handling or transport, the hangars, the systems of indication;
- **Costs of exploitations:** “the additional employment whose recruitment proves to be essential, additional supplies, new operations such as dredging of maintenance in the channel lately created” (Michel Loir, 1980).

V. Advantages of the harbor investment

According to Michel Loir (1980) the benefit of the harbor investment can be appreciated starting from the following elements:

5.1. Maritime economy

The harbor investment makes it possible to increase the productivity of the ships by accelerating its rotations or by allowing its recourse to new technologies, or by giving the access to ships of greater dimension.

5.2. Harbor economy

The effects depend on the fact that the investment is an investment of capacity or productivity. The increase in the capacities could reduce the costs of handling by limiting the distances from extrapotage, by reducing the constraints of intensives storage on a rarefied space, or by decreasing the proportion of overtime, the hours of night or the production bonuses. The investment of productivity like the opening of a container terminal. “Of the productivity gains will be to measure on the level of the costs of infrastructure per ton and with that of the costs of equipment and material of handling”.

VI. Theoretical corpus

Several theories can be mobilized to explain the lack of competitiveness of the French ports.

6.1. The theory of efficiency-X

For Leibenstein (1966) the external absence of pressure constitutes the first factor of inefficiency in the public organizations. For the author these companies are often in situation of monopoly, which would support a “quiet life” and would not incite those with a permanent effort of search of competitiveness, efficiency and effectiveness, contrary to the private sector.

The other argument evoked by Leibenstein (1966) is to explain the bad performance of the state enterprises which is due to their immortality. Indeed “the state enterprises are immortal, at least when the monetary policy and financial is sufficiently large to limit the probabilities of their failure” (Patrick Plane, 1994).

According to the theory of efficiency-X, the sources of inefficiency in the public organizations are justified by the unsuited behaviors of the State and its agents, on the one hand, and on the other hand by the strongly bureaucratized organizational structure of those. “Leibenstein (1966) evokes in the end the incidence of the multiplicity of the economic and social objectives that the political directors are carried to entrust to the state enterprises” (Patrick Plane, 1994).

“This is why, the theorists of efficiency-X support that the partnerships public-private (PPP) could contribute to reduce in a substantial way the sources of inefficiency-x in the public organizations, thus making it possible to make up again with the performance and competitiveness” (Hachimi Sanni Yaya, 2005).

6.2. The theory of Public New Management

It is a management style aiming at the introduction of the values and operating processes of the firm deprived in the public administration. The principal idea of the NMP is that “the public sector is considered to be ineffective, excessively bureaucratic, rigid, expensive, is centered on its own development, not innovating and having a too centralized hierarchy” (Anne Amar and al.).

For much, and as a reform, New Public Management precipitated the birth of the PPP. Thus New Public Management is a will “To modernize the State, to reinvent it, modernize the

public services, to improve management of the public organizations until the reform of the State, to found contracts of performance which are in fact the links of the new public managerial ideology, such is in a few words, the objective of New Public Management” (Hachimi Sanni Yaya, 2005).

VII. Conclusion

Various reports have highlighted the deteriorating global market share of French ports and lack of competitiveness compared to competing ports in the North Sea and the Mediterranean, particularly in terms of container. The port reform is necessary to ensure their competitiveness. This reform will certainly transfer to the private sector both hardware tools that employees working on this equipment by refocusing ports on specific missions to missions such as the sovereign of the port police, for the modernization of governance ports but also through the investment of land and river services as it has been because of "the battle of the Sea to gain ground."

VI. References

- 1- Anne Amar, Ludovic Berthier, Le Nouveau Management Public : Avantages et Limites
- 2- Chapon. J / Réflexion sur la nécessaire réforme des ports maritimes autonomes français / Transport, n° 446. novembre - décembre 2007 .
- 3- Hachimi Sanni Yaya, Les partenariats privé-public comme nouvelle forme de gouvernance et alternative au dirigisme étatique: ancrages théoriques et influences conceptuelles, 2005.
- 4- Josselin. M .J, Irreversibilité et financement des infrastructures par les fonds propres, 1997.
- 5- LE HAVRE PRESSE du vendredi 10 avril 2009
- 6- Leibenstein, H. Economic theory and organizational analysis. New York, Harper, 1966.
- 7- Michel Loir, Manual of Harbor Management, UNCTAD, 1980.
- 8- Ministère de l'Ecologie, du Développement et de l'Aménagement Durable (MEDAD), Dossier de presse Réforme des ports, 15 janvier 2008.
- 9- Patrick Plane, La privatisation dans les pays en développement : qu'avons-nous appris ?, Revue française d'économie, Année 1994, Volume 9, Numéro 2

Industrial Dynamics as applied to raw material supply process modeling: the case of a small company from the building industry

Ch.H. Fredouet

Abstract— Whatever the company, inventory disruption is a costly situation. To help reduce disruption risks, simulation stands out as a very efficient tool, as it gives the possibility to analyze the sensitivity of inventory levels to the variation of different internal and external factors. This paper describes the successive stages of building and then operating a simulation-based decision-support system dedicated to raw material supply management in the case of a semi-continuous production process. This decision-support system uses the Industrial Dynamics modeling framework.

Index Terms— decision-support systems, industrial dynamics, supply chain management, simulation.

I. INTRODUCTION

For many years now, companies have led, and are still leading, offshoring strategies, sourcing their raw materials and components from one part of the world, manufacturing their products in some other and selling these products on markets again located elsewhere. Furthermore, and although the trend is in some contexts slowing or even reversing, companies have also had for some time a growing tendency to outsource what they consider as non-core activities, within the supply, production and/or distribution processes.

Consequently, logistics, defined as the design, control and operation of demand-driven networks of actors, has gained an unprecedented share of contribution to the optimization of the performance of these companies.

To help improve company's global performance through the improvement of its logistics-based performance, research should be conducted addressing such issues as the design and implementation of logistics-aware decision-support systems and performance measurement systems [1], [2], [3].

This paper follows this line of thought in dealing with raw material process optimization in a building industry, SME environment:

Manuscript received May 16, 2009.
Ch. H. Fredouet is with the School of Engineers in Logistics (ISEL) of Le Havre University, France (phone: +33 2 32 74 49 00; fax: +33 2 323 74 49 11; e-mail: chf@univ-lehavre.fr).

Whatever the company, would it be from the production or the retailing industry, inventory disruption is a costly, and therefore unacceptable, situation of which this company must try to minimize the probability of occurrence. This necessity is all the stronger as companies are leading continuous or semi-continuous production activities: there, interrupting the process may have extremely disturbing consequences. Raw material and/or component inventories must therefore be sufficiently high to ensure continuous feeding of the activities in progress.

So as to reduce, hopefully eliminate, disruption risks, a given company will try to identify as precisely as possible the variables likely to influence the increase or the decrease of their inventories. To this end, simulation stands out as a very efficient tool, as it gives the possibility to analyze the sensitivity of the modeled system to the variation of different internal and external factors, therefore helping in pointing out the most influential.

In such a context, this paper describes the successive stages of building and then operating a simulation-based decision-support system dedicated to raw material supply management in the case of a semi-continuous production process. This decision-support system uses the Industrial Dynamics modeling framework, which has re-emerged in the mid-90's as a powerful, well-suited way of modeling logistics systems [4], [5], [6]. Beyond a certain level of complexity, the feedback loops encountered within logistics systems cannot be handled through mathematics, and their dynamic behavior can only be described through non-linear models [7], [8].

II. MODEL BUILDING

When designing a decision-support system (DSS), first step is to define the purpose of this tool, through the specification of the questions it is supposed to help answer.

21 – DSS purpose

The company retained as a case-study for DSS development belongs to the building industry; more specifically, it produces ready-to-use concrete, following a process which may be labeled as semi-continuous: as soon as a “production at the plant – delivery to the customer” set of activities begins, this set must be performed right to its end with no interruption. During this time-period, it is therefore necessary that raw materials are available in quantities sufficiently large to satisfy the needs of the on-going production process.

Solution presently implemented consists in maintaining high inventories of all raw materials needed, through frequent and systematic deliveries to the plant: while production is on its way, raw material usage is such that applying a classical order-point supply method would be too risky. Therefore, raw material suppliers automatically replenish inventories according to contractually pre-defined quantities and frequencies. However, this solution is obviously all the costlier as average inventories maintained are higher. Company’s top executives have consequently shown some interest in knowing whether inventory levels could be lowered while still avoiding disruption.

The DSS described in the following lines has been designed and implemented to help answer this question. As inventory levels depend upon multiple fixed (e.g.: maximum output capacity) or variable (e.g.: sales) factors, the global purpose of the tool can actually be extended to the analysis of inventory levels’ sensitivity to the variations of these factors.

22 – DSS content

To fulfill its purpose, the DSS must describe

- the various raw material inventories in the company,
- the variables likely to influence the evolution of their respective levels,
- the relationships between these variables.

Once identified, these data are then structured following the principles of Industrial Dynamics:

221 – raw material inventories:

Inventories may easily assimilated to those tanks the levels of which are depending upon incoming (supply) and outgoing (usage) flows, and which are described by J.W. Forrester in his best-selling book on Industrial Dynamics [9].

Graphically, each raw material inventory has therefore been modeled in the hereunder format (fig.1):

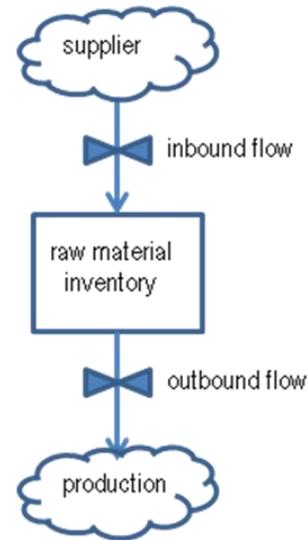


Fig.1: basic Industrial Dynamics modeling

Mathematically, the level of a given inventory S at a given time K is calculated from the level of S at immediately preceding time J , the inbound flow during the JK period, and the outbound flow during this same JK period:

$$\text{LEVEL.S}(K) = \text{LEVEL.S}(J) + \text{DT}(\text{INBOUND FLOW}(JK) - \text{OUTBOUND FLOW}(JK)),$$

where DT is the value of the selected time interval between two consecutive evaluation of the state of the system.

222 – inventory level variation factors:

The level of a raw material inventory is changing given the deliveries received from the suppliers and the quantities needed for the production of concrete.

a) inbound flows from suppliers:

For each raw material, determinant variables identified in this case-study are the supplier’s delivery unit and the frequency of delivery. Considering the systematicity of deliveries, neither the order point nor the safety stock nor the delivery lead-time are formally described as such; however, they are implicitly taken into account when specifying the rate at which deliveries are made by the supplier.

- Delivery unit: it corresponds to the capacity of the truck used by the supplier of a given raw material. In fact, as for obvious reasons of cost-effectiveness the truck is always running FTL, delivery unit cannot be considered as an actual decision variable; but when performing simulation scenarios, it may be used to test the inventory level’s sensitivity to variations of the

inbound flow in a more precise way (ton by ton) than only through the rate of delivery (truck by truck, that is 25 tons at a time).

- Frequency of delivery: contractually pre-defined by the company and its supplier there again for a given raw material, and notwithstanding contextual adjustments due to variations in the daily demand from the production system, the frequency of delivery depends upon 1) the usual rate of usage of the raw material, 2) the combination of delivery unit and geographical proximity of the supplier, and 3) the inventory level which the company wishes to maintain to ensure the continuity of its production activity.

b) outbound flows to production plant:

For the rate of a raw material outbound flow, all determinant variables are formally described in the DSS:

- Sales: estimated on a daily basis, sales feature three components, which are 1) the number of orders placed by the customers and to be delivered within a given day, 2) the share held by each concrete mix in the company's sales structure, and 3) the quantity to be produced per order and per mix.
- Customer order handling lead-time: in-between order reception at the plant and actual delivery to the customer, it is one-day long. Thus, the level of production of a given day, and therefore the rate of usage of raw materials, are determined by the nature and volume of orders received the previous day.
- Respective raw material usage of the different concrete mixes: as they are rather standardized, concrete mix compositions should be considered as parameters of the DSS rather than actual simulation variables. However, it happens that (very) large orders for specific building projects require unusual concrete mixes; so it should be possible to modify the compositions accordingly to estimate the impact of such or such future sales contract on raw materials' inventory levels.

the one calculated from available raw material inventories; if the latter are too low for all orders to be satisfied, a backlog of orders will start to build up, and will keep growing as long as suppliers' deliveries and/or demand reduction won't have brought inventories back to such levels that demand may be met.

The hereunder diagram, roughly sketching the model described in the present paper, shows this type of "loop" structure (fig.2). Built from the main variables and inter-relationships of the raw material supply system, this diagram serves as a basis for the actual development of the computerized simulation decision-support system.

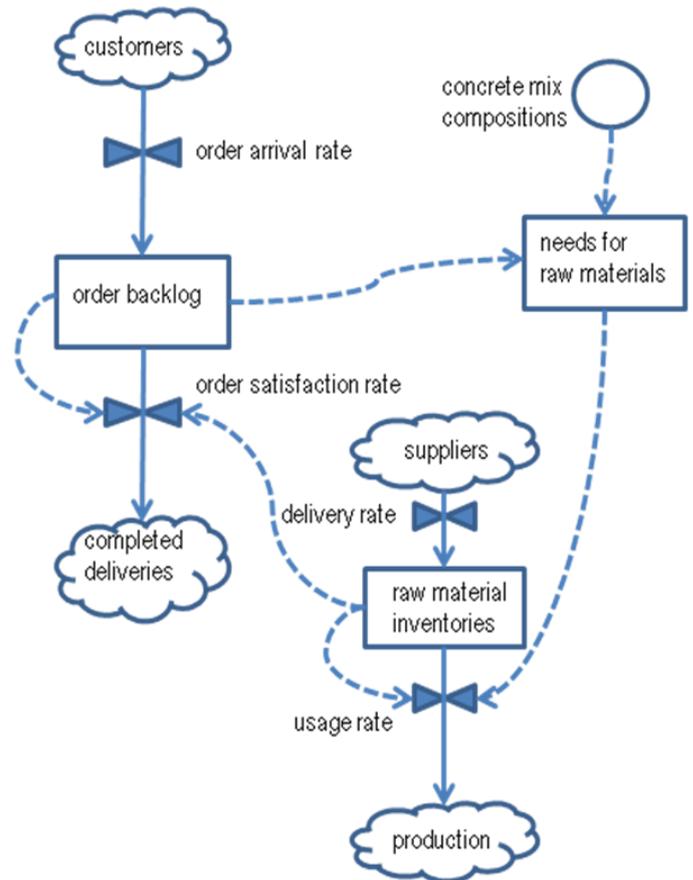


Fig.2: rough sketch of the model

223 – relationships between variables:

The main ambition of Industrial Dynamics-based models is to describe systems the operation of which is characterized by feedback loops between (some of) their components.

Here, raw material inventory levels reached at time J determine the nature and volume of customer orders likely to be satisfied during JK period: the rate of satisfaction retained will be the smaller of the one calculated from orders received and

III. MODEL IMPLEMENTATION

The first step in the implementation of the Industrial Dynamics model has been the development of a relevant

computer-based decision-support system, using a dedicated programming language.

Then, once the program was properly running, a number of simulation scenarios have been performed, to analyze the sensitivity of the raw material inventory levels to various evolution hypotheses in the supply strategy and/or the commercial activity of the company.

31 – DSS programming

Choosing the programming language, and activity-level generation, have been the two most important issues to be addressed during this first step of model implementation.

311 – choosing of the programming language:

An Industrial Dynamics-based model features three types of equations:

a) level equations:

They describe the evolution of variables (inventories, order backlogs, ...) the value of which, as it would be for a tank, changes from one period to another due to the action of other variables.

b) flow equations:

They describe the evolution of those variables (sales, supplies, ...) which increase or conversely decrease the various levels identified in the system.

c) auxiliary variable equations:

They calculate the value of other system variables or the value of some inbound / outbound flows.

Each iteration of the model must bring out the state of the system at time K, depending upon its preceding state at time J and the events having occurred during the JK time interval; to this end, the equations are processed in the following order (fig.3):

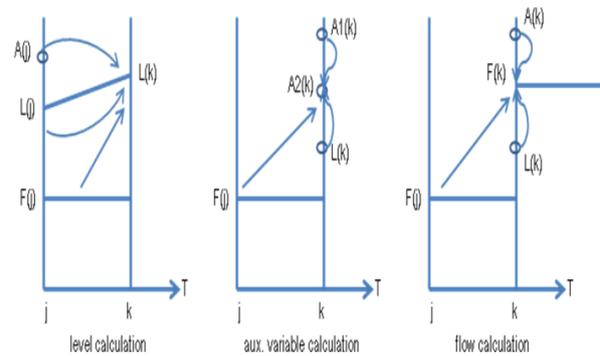


Fig.3: equation evaluation process

For optimal programming of such a specific DSS, with its multiple iterations and feedback loops, a dedicated language has been sought for. Among other available solutions, the language which has been opted for is the Fortran-based DYNAMO, as it fitted so well within the programming constraints and kept alive the link with Industrial Dynamics and the research work of J.W.Forrester. However, a Vensim-based version should be implemented in a near future, among other reasons for improved end-user friendliness.

312 – activity-level generation:

The company's level of activity is evaluated through its volume of sales, which is itself measured in cubic meters of finished product. This volume of sales must be disaggregated to identify the quantities sold per each composition of concrete mix, as they have led to different usages of raw materials.

Daily sales have therefore been analyzed into three components:

- the total number of orders (TNBO(K))
- the relative share of each mix X in total sales (SXTS(K))
- the amount of cubic meters per each order of each mix X (QSOX(K))

The total volume sold of a given mix X (TQSX(K)) is then calculated as follows:

$$TQSX(K) = TNBO(K) * SXTS(K) * QSOX(K)$$

where $1 \leq X \leq N$ and N is the number of different mixes sold

Thanks to a function randomly generating, on each iteration of the model, a number comprised between 0 and 1 (included), as well as the availability of a long time-series of the company's daily sales, probabilities have been associated to the determinant TNBO(K), SXTS(K) and QSOX(K) variables.

The tables hereafter give some instances of values actually implemented (tables 1, 2 & 3):

Table 1: values for TNBO (K)

Day of the week	Values for TNBO (K)	Probability of occurrence
Mon., Tues., Wed.	24	0.33
	14.8	0.67
Thurs., Fri.	18.33	0.30
	28.29	0.70

Table 2: values for SXTS (K) for concrete mix 1

Values for SITS (K)	Probability of occurrence
19%	0.03
32%	0.15
61%	0.40
86%	0.42

Table 3: values for QSOX (K) for concrete mix 1

Values for QSO1(K)	Probability of occurrence
2 m ³	0.05
3 m ³	0.13
4.5 m ³	0.15
6 m ³	0.67

If the random number generated at time K is for instance 0.8, then
 TNBO(K) will be 14.8 or 28.29 orders depending upon the day of the week,
 SITS(K) will be 61%,
 QSO1(K) will be 4.5 m³
 and TQS1(K) will be $14.80 * 0.61 * 4.5 = 40.62$
 or $28.29 * 0.61 * 4.5 = 77.65$
 depending upon the day of the week.

Sales per mix thus obtained lead to the calculation of raw material needs, and therefore inventory outputs for production. Obviously, full order satisfaction cannot be attained if available inventory levels are not high enough. Besides, supplier deliveries are performed according to actual volumes and frequencies. Finally, these outbound and inbound flows make for the inventory levels, targeted by the simulation.

Once the program is reliably operational, simulation sessions may start.

32 – DSS operation

The time frame presently in use within the DSS is set to 90 days (18 weeks). Considering the system is evaluated every two hours, (a quarter of a day), it implies 360 model iterations, which has been validated as satisfactory by the company's chief operating executives.

Simulation may then now lead to a better knowledge of raw material inventories' behavior, and identify the conditions leading to an improved balance between high inventory costs and risks of disruption.

To this end, two sets of simulation operations have been performed: one pertains to the level of sales; the other, to the level of supplies.

321 – simulating the sales level:

Question to be answered here was "What if ... sales increase?", more specifically "How much time do the inventory levels give the company to adjust its suppliers' delivery rates?".

A simulation has therefore been run where monthly sales were increased by 10% from the "normal" level of activity, standing at 2400 cubic meters of concrete mix.

Keeping to the corresponding rate of delivery from the suppliers, it takes 10 days for all inventory levels but one to reach a steady state, where the level is equal to the delivery unit, that is the full truck load capacity. Disruption occurs only, every two days, for the cement inventory, due to its lower delivery rate: 3 trucks every two days instead of one every half-day, or even one every hour, for the other raw materials.

Of course, disruption is avoided at the cost of a sharp increase in order backlog. Average inventory levels are therefore not that high that the company can easily absorb the simulated change in its sales volume. However, they give a time buffer of nearly two weeks to help adjust to the new level of activity, and, considering the (close) geographic proximity and (high) delivery flexibility of the suppliers, this buffer does not need to be that long.

The next simulation to be run was then dedicated to answering another question: "to what extent the time buffer, and therefore global volume and cost of inventories, may be reduced without going disrupt?".

322 – simulating the supply levels:

While delivery rates remained unchanged, raw material delivery units have been progressively reduced from one simulation to the next, leading to the design of a new delivery

rate / unit supply policy, compared to the present one in the following table (table 4):

Table 4: supply policies comparison

Type of raw material	Delivery rate	Present delivery unit (P)	Suggested delivery unit (S)	(P) – (S) difference
Coarse gravel	Every 3h	25	25	0
Fine gravel	Every 4h	25	25	0
Coarse sand	Every 1h	7	5.75	1.25
Fine sand	Every 4h	25	17	8
Cement	Every 2 days	80	66	14

Such a decrease in volumes leads to purchasing cost reductions. For a 20-working days, 8h per day, reference month, the suggested supply policy would result in total savings of around 15,000 euros, detailed in the hereunder table (table 5):

Table 5: purchasing cost savings

Type of raw material	Unit cost (in euros)	Monthly volume reduction (in tons)	Monthly cost savings (in euros)
Coarse gravel	8	0	0
Fine gravel	8	0	0
Coarse sand	9	200	1800
Fine sand	11	320	3520
Cement	75	140	10500

The enhanced cost-effectiveness of the raw material supply policy has no impact on the reliability of the production process: at no time during the whole simulation period have the inventories suffered from disruption.

To measure the available time-safety margin, and incidentally the sensitivity of inventory levels to variations in the supply policy, delivery units have been further reduced by one ton / one half-ton at a time from one simulation to another, starting from the newly defined values. As a consequence, all inventories have gone disrupt within one month.

More specifically, would the company decide to adopt the suggested new raw material supply policy, longest intervals between two deliveries would be

- 1 truck every 6 days for coarse gravel,
- 1 truck every 16 days for fine gravel,
- 1 truck every 3 days for coarse sand,
- 1 truck every 12 days for fine sand,
- 1 truck every 53 days for cement

IV. CONCLUSION

Those latter results bring an end to the simulation process dedicated to answering the initial question asked by the company's top executives.

From an operational standpoint, this decision-support system may however be used more extensively by companies from the building industry for any set of simulations of the impact of a given determinant variable on such or such inventory level. For instance a table could thus be fed which would give, for each monthly sales level, the optimal delivery unit / rate supply policy to be conducted accordingly.

From an academic standpoint, this paper should be viewed as another contribution to the revival of Industrial Dynamics as a basis for the design and implementation of logistics-dedicated, simulation-based decision-support systems, would it be on operational or on strategic issues.

REFERENCES:

- [1] M. Busi, and U.S. Bititci, "Collaborative performance management: present gaps and future research," *International Journal of Productivity and Performance Management*, vol.55, no. 1, pp. 7-25, 2006.
- [2] K.H. Lai, C.W.Y. Wong, and T.C.E. Cheng, "A coordination-theoretic investigation of the impact of electronic integration on logistics performance," *Information & Management*, vol. 45, no. 1, pp. 10-20, 2008.
- [3] M.J. Meixell, and M. Norbis, "A review of the transportation mode choice and carrier selection literature," *The International Journal of Logistics Management*, vol. 19, no. 2, pp. 183-211, 2008.
- [4] D.R. Towill, "Industrial Dynamics modeling of supply chains," *International Journal of Physical Distribution and Logistics Management*, vol. 26, no. 2, pp. 23-42, 1996.
- [5] D.R. Towill, M.M. Naim, and J. Wikner, "Industrial Dynamics simulation models in the design of supply chains,"

International Journal of Physical Distribution and Logistics Management, vol. 22, no. 5, pp. 3-13, 1992.

[6] E.F. Wolstenholme, and R.G. Coyle, "The development of System Dynamics as a methodology for system description and qualitative analysis," *Journal of the Operations Research Society*, vol. 34, no. 7, pp. 569-581, 1983.

[7] D. Thiel, "Towards a better knowledge of production system dynamics. Conception of generic models and continuous simulation," Ph.D. dissertation, Aix-Marseille III University, Aix-en-Provence, France, 1993.

[8] G.P. Richardson, "Problems with causal-loop diagrams," *Systems Dynamics Review*, vol. 2, no. 2, pp. 158-170, 1986.

[9] J.W. Forrester, *Industrial Dynamics*. Cambridge, MA: MIT Press, 1961.

Nonlinear Analysis of Time Series and Applications

Organizing Committee

- Ralph Gregor ANDRZEJAK
Universitat Pompeu Fabra, Barcelona, Spain
- Dimitris KUGIUMTZIS
Aristotle University of Thessaloniki, Greece
dkugiu@gen.auth.gr

Description

Nonlinear time series analysis is the practical spin-off from complex dynamical system and chaos theory. It allows one to characterize, or even make predictions of, dynamical systems in which nonlinearities give rise to a complex temporal evolution by analyzing signals measured from these dynamics. Importantly, this concept allows extracting information which cannot be resolved using classical linear techniques such as for example the power spectrum or spectral coherence. In recent years, the framework nonlinear time series analysis has been extended to also comprise methods derived from statistical physics, information theory, statistics and computer science. Application of nonlinear time series analysis can be found in many fields, ranging from biology, neuroscience, engineering, to geophysics and economics. This session will cover some recent advances in the methodology of nonlinear time series analysis and provide examples for applications to a variety of real-world experimental dynamics.

Contents

Localizing epileptic foci using surrogate-baseline corrected nonlinear synchronization measures	
Ralph Gregor ANDRZEJAK , Daniel CHICHARRO , Florian MORMANN , Klaus LEHNERTZ	331
Using a nonlinearity detection as a prior step for global modeling	
Ubiratan S. FREITAS , Christophe LETELLIER	333
Improvement of symbolic transfer entropy	
Dimitris KUGIUMTZIS	338
Assessing the degree of synchronization in time series using symbolic representations	
Roberto MONETTI , Wolfram BUNK , Thomas ASCHENBRENNER , Stephan SPRINGER	343
Characterization of pre-seizure states by a multi-signal analysis of scalp EEG and ECG signals during slow wave sleep	
Stavros NIKOLOPOULOS , Mario VALDERRAMA , P. MILANI , V. NAVARRO , Michel LE VAN QUYEN	347
Complex networks as a tool for nonlinear time series analysis	
Michael SMALL , Xiaoke XU	347
Non-Gaussian signal processing of biological signals	
Max LITTLE	353

Localizing epileptic foci using surrogate-baseline corrected nonlinear synchronization measures

Ralph G. Andrzejak, Daniel Chicharro, Florian Mormann, and Klaus Lehnertz.

Abstract—We applied bivariate nonlinear time series analysis techniques to electroencephalographic (EEG) recordings from epilepsy patients. In particular, we combined a novel bivariate synchronization measure with bivariate surrogates. We tested the discriminative power of this surrogate-baseline corrected nonlinear synchronization measure to detect the seizure-generating hemisphere in medically intractable medial temporal lobe epilepsy. For this purpose, we analyzed intracranial EEG recordings from the seizure-free interval of 29 patients. Our results demonstrate that the surrogate-baseline correction is essential for a successful characterization of the spatial distribution of the epileptic process.

Index Terms—Nonlinear time series analysis, Synchronization measures, Surrogate time series, Electroencephalography, Epilepsy, Focus localization.

THE disease epilepsy is characterized by sudden and recurrent malfunctions of the brain that manifest themselves as epileptic seizures. During epileptic seizures large groups of neurons discharge hyper-synchronously. In consequence, the electroencephalogram (EEG) recorded during epileptic seizures is characterized by rhythmic oscillations of high amplitude. During the seizure-free interval only intermittent short bursts of hyper-synchronous activity of local neuron groups occur, resulting in so-called interictal epileptiform activity in the EEG. Except for these -mostly brief- episodes of interictal epileptiform activity, the EEG recorded during the seizure-free interval often appears

Manuscript received May 22, 2009. Ralph G Andrzejak acknowledges the grant BFU2007-61710 of the Spanish Ministry of Education and Science. Daniel Chicharro was supported by the grant 2008FI-B 00460 of the 'Generalitat de Catalunya' and European Social Funds. Florian Mormann was supported by the 6th Framework Program of the European Commission (Marie Curie OIF 040445).

R. G. Andrzejak is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain. (corresponding author e-mail: ralphandrzejak@yahoo.de).

D. Chicharro is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain..

F. Mormann is with the Division of Biology, California Institute of Technology, Pasadena, USA and with the Department of Epileptology, University of Bonn, Bonn, Germany

K. Lehnertz is with the Department of Epileptology, the Helmholtz-Institute for Radiation and Nuclear Physics, and with the Interdisciplinary Center for Complex Systems, University of Bonn, Bonn, Germany

unspecific with regard to the epileptic process. It can be conjectured, however, that even in the absence of evident interictal epileptiform activity, the epileptic process causes an elevated level of neuronal synchronization.

Indeed there is growing evidence that this elevated level of synchronization can be detected from EEG recordings of the seizure-free interval using different nonlinear time series analysis techniques. In particular, nonlinear synchronization measures based on instantaneous phases (e.g. [1]), reconstructed state spaces (e.g. [2]) or techniques derived from information theory (e.g. [3]) have been used to analyze intracranial EEG recordings from patients with medically intractable unilateral medial temporal lobe epilepsy. The particular recordings studied in [1-3] offer well-defined conditions since the EEG was measured using electrodes symmetrically implanted in the hippocampal formations in the left and right brain hemispheres. Furthermore, based on these diagnostics, one hippocampal formation was identified as the seizure-generating structure (epileptic focus), while no seizures originated from the opposite brain hemisphere. These studies [1-3] congruently showed that the mean level of synchronization between recording sites within the focal hemisphere is higher than the one between recording sites within the opposite hemisphere. In none of these studies, however, were the synchronization measures combined with the concept of surrogates. In the context of univariate nonlinear time series analysis, surrogates were shown to substantially improve the localization of the epileptic focus [4, 5]. The combination of nonlinear bivariate synchronization measures and bivariate surrogates in application to EEG recordings from epilepsy patients is therefore tested in the present study.

From two simultaneously recorded time series a pair of bivariate surrogates is constructed such that the surrogate time series share, for example, the linear cross- and autocorrelation with the original time series, but are otherwise random (e.g. [6, 7] and references therein). Accordingly, bivariate surrogates can be used to estimate the value of the nonlinear synchronization measures expected for a bivariate linear stochastic process. A surrogate-baseline corrected nonlinear synchronization measure can then be defined as the difference between the synchronization measure's value obtained for the

original time series and the one obtained for the surrogate time series.

For the present study we used a novel nonlinear state space based synchronization measure (L), which was introduced recently in [8] and was shown to detect couplings between dynamical systems with higher sensitivity and specificity as compared to previously published nonlinear state space based techniques [8]. We here combined this measure with bivariate surrogates that were constructed to preserve the cross-correlation, autocorrelation, and also the amplitude distribution of the original time series [6]. Except for these constraints, the surrogates were random, and were used to define a surrogate-baseline corrected measure ΔL , as described above.

We analyzed intracranial EEG recordings from the seizure-free interval of 29 patients with medically intractable medial temporal lobe epilepsy. A total of 84 EEG recordings with an average length of 130 min per patient were analyzed using a moving window technique. (These datasets were previously analyzed using univariate time series analysis measures in [5] and overlap with the recordings studied using bivariate synchronization measures in [1-4]).

Increased values of the measure L allowed us to correctly determine the side of the focal hemisphere in 22 of 29 cases. This performance is comparable to the one obtained by other synchronization measures in previous studies [1-4]. The surrogate-baseline corrected version ΔL allowed us to correctly determine the focal hemisphere in 26 of 29 cases. Extending results obtained by univariate time series analysis techniques [5], our results further demonstrate the importance of the concept of surrogates for a successful characterization of the spatial distribution of the epileptic process.

REFERENCES

- [1] Mormann F, Lehnertz K, David P, Elger CE. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D*. 2000; 144: 358-369.
- [2] Arnhold J, Grassberger P, Lehnertz K, Elger CE. A robust method for detecting interdependencies: application to intracranially recorded EEG. *Physica D*. 1999; 134: 419-430.
- [3] Osterhage H, Mormann F, Staniak M, Lehnertz K. Measuring synchronization in the epileptic brain: A comparison of different approaches. *Int J Bifurcat Chaos*. 2007; 17: 3539-3544.
- [4] Andrzejak RG, Widman G, Lehnertz K, Rieke C, David P, Elger CE. The epileptic process as nonlinear deterministic dynamics in a stochastic environment—an evaluation of mesial temporal lobe epilepsy. *Epilepsy Res*. 2001; 44: 129-140.
- [5] Andrzejak RG, Mormann F, Widman G, Kreuz T, Elger CE, Lehnertz K. Improved spatial characterization of the epileptic brain by focusing on nonlinearity. *Epilepsy Res*. 2006; 69: 30-44.
- [6] Andrzejak R.G., Kraskov, A., Stögbauer, H., Mormann, F., Kreuz, T. Bivariate surrogate techniques: necessity, strengths, and caveats. *Phys. Rev. E*. 2003; 68: 066202.
- [7] Schreiber, T., Schmitz, A., 2000. Surrogate time series. *Phys. D*. 2000; 142: 346-382.
- [8] Chicharro D, Andrzejak RG. Reliable detection of directional couplings using rank statistics. Submitted.

Using a nonlinearity detection as a prior step for global modeling

U. S. Freitas & C. Letellier

Abstract—Identifying chaos from experimental data remains a very challenging problem for which conclusive arguments are still very difficult to provide. One possible answer is to find and validate a global model, although this could prove a rather challenging task. In an attempt to address this question, Poon and Barahona introduced a numerical titration procedure based on a nonlinearity detection method. We show that such numerical titration procedure fails to distinguish non-deterministic signals from low-dimensional deterministic chaos and, therefore, is unable to reliably detect chaos from time series.

Nevertheless, we show that the nonlinearity detection method can be used to browse the identification parameter space before applying a global modeling technique. Used in this way, the nonlinearity detection method can serve as a pre-test to ease the more complex global modeling techniques.

Index Terms—Chaos, time series, global modeling, nonlinearity.

I. INTRODUCTION

When one investigates time series from the real world, one necessarily faces an underlying dynamics which most often results from a complex interplay between deterministic and stochastic components. A special attention is thus devoted to detect or to identify the deterministic component. Sometimes this aim consists in a detection technique that is supposed to prove the existence of an underlying chaotic component. Many techniques were proposed to detect chaos but none of them is fully reliable. All of them rely on certain topological or information measures of attractors reconstructed from the data and present some problems of specificity and reliability [1], [2]. It is known that the Largest Lyapunov exponent fails to distinguish chaotic behaviors from noise [3].

Let us be precise to what a chaotic behavior corresponds. First of all, a chaotic behavior is deterministic, that is, governed by a process which can be described by a set of ordinary differential or difference equations. It might also be described by a delay-differential equation since a delay is often required before the answer to a given event is provided by the system, as usually observed in biology, for instance. Typically, determinism is the paradigm in which future events are a consequence of past and present events combined with the law of nature. Such determinism dates back from the Laplace's thought experiment [4]. What blurred the image provided by Laplace is that when there is chaos, it is no longer possible to foresee the future for an infinite time. Predictions of chaotic dynamics can only be made for short time. Due to this latter property, identifying a determinism underlying experimental

data is quite a complicated problem, necessarily because deterministic chaotic behaviors cannot be distinguished from randomness using statistical analysis [5]. As clearly written by Glass [6], prior to asserting that some dynamics is chaotic, there should be clear evidence that deterministic equations govern the dynamics.

Implicitly, this underlying determinism is considered in terms of low-dimensional determinism, just because when the system dimension is too large, such a determinism can no longer be distinguished from a stochastic process. Usually, low dimensional dynamics means that the behavior can be described in a phase space whose dimension is roughly less than ten. Proving that the dynamics underlying a noise contaminated short time series corresponds to a "low-dimensional" chaotic behavior is one of the most difficult problem to address. This is mainly why surrogate data analysis have been so often used [13]. Unfortunately, this technique only test whether the investigated dynamics can be distinguished from a linear stochastic process, or not. This is therefore not a direct — and definite — answer to the original question, that is, identifying an underlying determinism.

Once the deterministic character of the dynamics is evidenced (or assumed), in order to assert its chaotic nature, one has to show that that the measured time series results from a dynamical process which is *sensitive to initial conditions*, *bounded* and *recurrent*. The fact this measured data set was produced by a bounded process is certainly the less risky assumption which can be made about the underlying dynamics. Then remains the sensitivity to initial condition and the recurrence property of the dynamics. To show that the dynamics is sensitive to initial conditions can be done by computing the largest Lyapunov exponent, although such a computation is still a great challenge when the available time series is short and noise contaminated [14]. Now the recurrence property can only be tested, by definition, from long enough time series. Such a property is related to the population of unstable periodic orbits around which chaotic behaviors are organized [15]. The relative organization of these periodic orbits leads to the architecture of chaotic attractors [16]. Getting periodic orbits from short time series necessarily requires the estimation of a global model which can then be integrated over a long time (see [19] among others). Moreover, possessing a validated global model corresponding to the investigated data set is one of the best proofs for the underlying determinism. When the measured data is long enough, global models are ideally validated by a topological analysis [19] but this is no longer possible when the time series is short (typically around 10 or 20 cycles).

Although finding a valid global model is the ultimate test

CORIA UMR 6614 — University of Rouen, Saint-Etienne du Rouvray, France, e-mail: freitas@coria.fr
Manuscript received May 22, 2009.

for determinism, it is also a complex procedure. It requires some experience of its user and can fail for several reasons. Therefore, a good screening procedure that is able to detect cases where the global modeling would fail could be a invaluable tool.

In 2001, Poon and Barahona [17] proposed a method for detection of chaos in time series. This method uses a non-linearity detection procedure introduced by the same authors on a previous paper [11]. The chaos detection method proved to be unreliable. It will erroneously classify as chaotic time series produced by pure stochastic systems. This shows that the nonlinearity detection procedure cannot be used as a chaos detector, at least not directly. Instead, we propose to use the nonlinearity detection method as a pre-test procedure for the identification of global models.

II. NUMERICAL TITRATION AND NONLINEARITY DETECTION

Poon and Barahona's method of chaos detection [17] is composed by two principles: a nonlinear detection method [11] and gradual addition of noise.

The nonlinear detection method is based on the comparison between one-step ahead prediction performances of linear and nonlinear models. The linear models are autoregressive discrete-time structures like

$$y_k = \sum_i \theta_i y_{k-i}, \quad (1)$$

where the signal at an instant k , y_k , is considered as a linear combination of the same signal at previous instants. Nonlinear models are also autoregressive, but include polynomial combinations of the of the time-delayed terms. A structure like eq. (1) is a one-step ahead predictor if the y_{k-i} on the right-hand side contain only values taken from the time series. The coefficients, θ_i , are computed from the time series using a least squared method. Several linear and nonlinear models are tried, varying the number of terms in each model. The nonlinearity is detected if a nonlinear model has smaller one-step prediction error than any linear model in a statistical sense.

The chaos detection procedure is as follows. The nonlinear detection method is applied to the time series under test. Then, white or linear correlated noise is gradually added to the data until the nonlinear detection method fails to see any nonlinearity. The standard deviation of the added noise, divided by the standard deviation of the time series, at this point, is called the noise limit (NL) and is claimed to be a measure of chaos in the data. The condition $NL > 0$, that is, nonlinearity is detected and noise of a significant variance must be added for the detection to fail, is claimed to be sufficient to deduce the presence of chaos in the data.

Unfortunately, this technique does not provide a definite answer in the sense that, in some cases, it provides incorrect positive answers. Let us give an example with a nonlinear colored noise. In order to do that, a nonlinear moving average filter is applied to random noise according to

$$x_{n+1} = av_n + bv_{n-1}(1 - \nu_n) \quad (2)$$

where ν_n is a uniform i.i.d. random variable with values between 0 and 1. This is a purely random signal (Fig. 1) which can be considered as a sort of nonlinear colored noise. Its stochastic character is well evidenced when a first-return map is computed (Fig. 2). No deterministic structure (like a parabolic shape or other) can be evidenced from this first-return map.

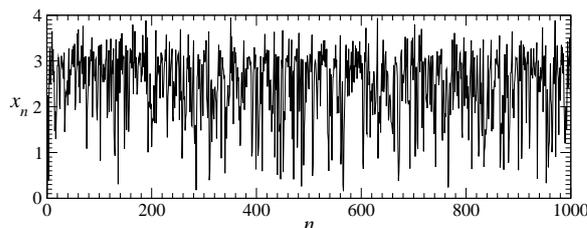


Fig. 1. 1000 points of a stochastic solution to map (2). Parameter values: $a = 3$ and $b = 4$.

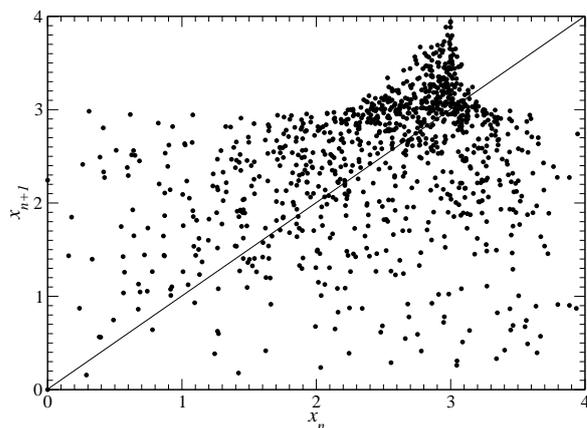


Fig. 2. First-return map computed from a trajectory produced by (2).

Applying noise titration leads to a noise limit $NL=35\%$. Thus this technique would erroneously conclude in favor of a chaotic deterministic behavior although the underlying dynamics is clearly not deterministic. The reason is that this nonlinear colored noise is predicted more accurately (one-step ahead) with a nonlinear model than with a linear one. Moreover, we search for a purely chaotic dynamics for which the noise limit was also about 35%. In order to do so, the Logistic map

$$x_{n+1} = \mu x_n(1 - x_n) \quad (3)$$

was investigated with increasing values for parameter μ . It was finally found that with $\mu = 3.62$, the noise limit was about 35%. For this μ -value, the first-return map looks like a 2-banded parabola (Fig. 3). This means that in a blind test, the noise titration does not make difference between a nonlinear colored noise (Fig. 2) and a purely chaotic behavior (Fig. 3)[12].

III. A PERIODIC GLOBAL MODEL DOES NOT MEAN UNDERLYING DETERMINISM

Since the pioneering paper by Crutchfield and McNamara [7], it is known that it is also possible to get a set of differential

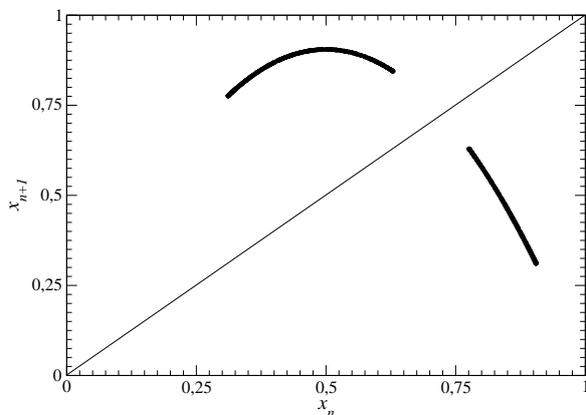


Fig. 3. First return map computed from a trajectory produced by the Logistic map (3). Parameter value: $\lambda = 3.62$.

or difference equations which produces a chaotic attractor equivalent to the experimental one. It is important to note that a single model is thus used to describe the whole underlying dynamics. Global models are thus opposed to “patchwork-models” which result from collection of local models. Global modeling thus refers to getting a single model for reproducing chaotic dynamics. In this paper, the global models obtained can be written as

$$y_k = f(y_{k-1}, y_{k-2}, y_{k-3}, \dots)$$

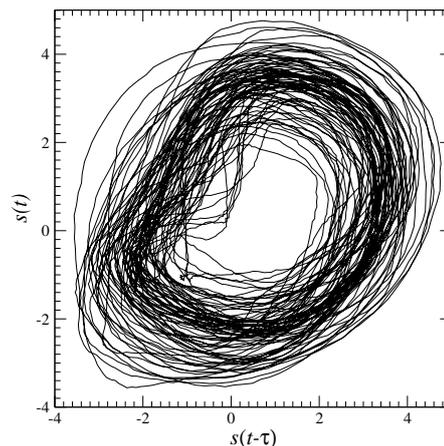
where $\{y_k\}$ is the measured time series [8], [9]. This is the so-called Nonlinear Auto-Regressive Moving Average (NARMA) model. ARMA, therefore, refers to the “ingredients” used to explain (predict) the data y_k and N indicates that such ingredients are combined in a nonlinear way. One possible nonlinear way of combining such variables is using a polynomial to approximate f , although many other alternatives exist. Usually finding a global model is a signature of an underlying determinism. Nevertheless, when the obtained model produces a limit cycle, this is not always the case.

Let us start from data which have no underlying nonlinear dynamics, that is, surrogate data [10]. Surrogate data were produced from the smoothed x -time series of the noise-contaminated Rössler system [18]

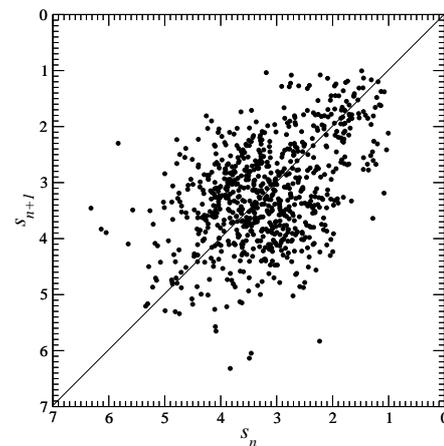
$$\begin{cases} \dot{x} = -y - z \\ \dot{y} = x + ay \\ \dot{z} = b + z(x - c) \end{cases} \quad (4)$$

with $a = 0.3$, $b = 2$ and $c = 4$. The corresponding phase portrait (Fig. 4a) does not provide any regularity that could suggest us an underlying determinism. In particular, there is no folding on this portrait, a relevant ingredient for producing chaos [16]. When a first-return map to a Poincaré section of the surrogate phase portrait is computed a simple cloud of points without any structure is identified (Fig. 4b). From the surrogate data, it was not possible to obtain a single stable model that would settle (whenever stable) to anything but a limit cycle. This results from the type of surrogate data used in the present case. Only phases were shuffled to remove the non

linear contribution within the dynamics. As a consequence, only the “linear” cycling around the inner fixed point was left in the surrogate data. The “linearized” deterministic dynamics cannot provide anything more complicated than a period-1 limit cycle. Consequently, getting a model that only produces a limit cycle is *not* an evidence for determinism.



(a) Phase portrait



(b) First-return map

Fig. 4. Surrogate data computed from the smoothed x -time series of the noise-contaminated Rössler system. Phases are shuffled in such a way that the power spectrum is preserved. Parameter values: $a = 0.3$, $b = 2$ and $c = 4$.

IV. NONLINEARITY DETECTION AS A SCREENING PROCEDURE FOR GLOBAL MODELING

Testing for nonlinearity is a simpler task than identifying and validating global models. The former is easily automated while the later needs user interaction, specially in the validation part. We therefore propose to use the technique to detect nonlinearity as a first step before applying a global modeling technique. We apply this procedure on the Rössler system with parameter values as $a = 0.398$, $b = 2$ and $c = 4$. The ability to obtain a global model from a time series produced by a system can be evaluated using observability indices as

introduced in [20], [21]. In the case of the Rössler system, it has been shown that variable y provides the best observable since there exists a diffeomorphism between the original phase space and the y -induced differential embedding spanned by y and its successive derivatives [21]. Conversely, variable z is the worst and roughly, it is nearly impossible to get a 3D global model without a strong structure selection [22]. Since we do not want to test the observability of the Rössler dynamics via a given variable but rather how it is possible to discriminate times series that are good candidates for applying a global modeling technique, the time series measured will correspond to the time evolution of the y variable of the Rössler system.

To simulate a realistic case, i.i.d. gaussian noise is added to the time series. Depending on the noise level η (standard deviation of noise normalized by standard deviation of time series), the original dynamics is more or less perturbed (Fig. 5). When the noise level is too high, it becomes impossible to extract the deterministic component from the measured time series and, consequently, the global modeling technique fails.

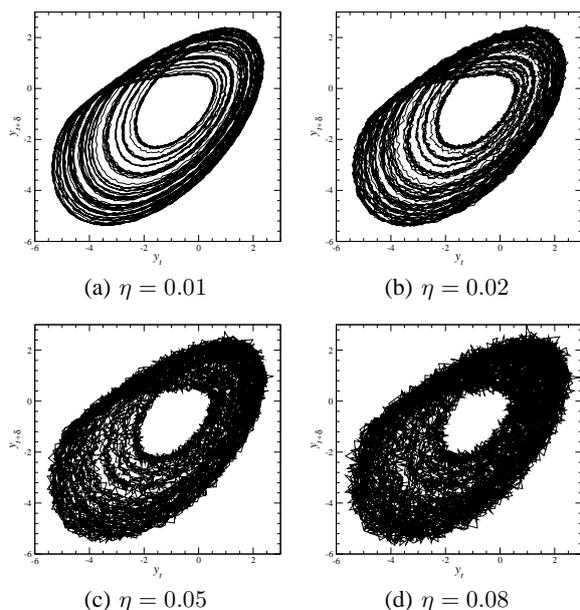


Fig. 5. Delay embedding induced by variable y of the Rössler system. Parameter values: $a = 0.398$, $b = 2$, $c = 4$ and embedding delay 1.0s.

As we saw in the previous section, getting a global model producing a limit cycle does not mean there is an underlying determinism. It is therefore required to obtain a global model producing a chaotic attractor. In our attempt to obtain global model from noise-contaminated data, we will only select those producing chaotic behaviors. Two parameters were varied during the global modeling procedure: i) the noise level (0.01, 0.02, 0.05 and 0.08) added to the time series and ii) the sampling number τ (1,4,8,10 and 20) used to create the time series. For $\tau = 1$, every point from the original time series, sampled at 0.05s, is used. For $\tau = 4$, only a point every 4 is taken; for $\tau = 8$, only a point every 8 and so on. Depending on values of these two parameters, it was possible or not to identify a global chaotic model. For each set of

parameters, the result of the nonlinear detection algorithm, that is, the probability for having a better prediction with nonlinear models than with linear models was computed. Our results are reported in Tab. I. As expected, the nonlinearity becomes difficult to detect as the noise level is increased. As a consequence, it becomes nearly impossible to get a chaotic global model from too noisy data.

TABLE I
PROBABILITY p FOR HAVING BETTER ONE-STEP AHEAD PREDICTION BY A NONLINEAR MODEL THAN BY A LINEAR ONE. CASES FOR WHICH A CHAOTIC GLOBAL MODEL WAS OBTAINED ARE MARKED WITH AN ASTERISK.

Noise level	0.01	0.02	0.05	0.08
$\tau = 1$	0.0148	0.0018	0.0008	0.0007
$\tau = 4$	1.0000	0.9997 *	0.7083	0.2809
$\tau = 8$	1.0000 *	1.0000 *	0.9999	0.9615
$\tau = 10$	1.0000 *	1.0000 *	1.0000	0.9997
$\tau = 20$	1.0000 *	1.0000 *	1.0000 *	1.0000

An interesting feature is that, for a given noise level, our ability to obtain such a chaotic global model increased as τ is increased. As an example, the last chaotic model we obtained (Fig. 6) was for the largest delay ($\eta = 0.05$).

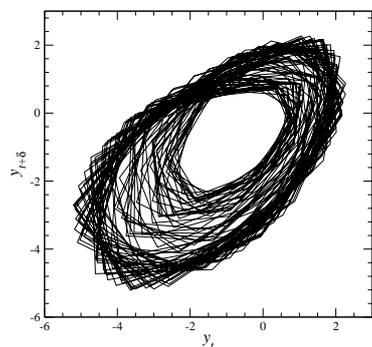
An important thing to note is that all chaotic global models obtained correspond to a probability p greater than 0.99 to have better predictions by nonlinear models than by linear ones. This threshold, 0.99, is the one proposed for the nonlinearity detection [11]. This suggests that a negative nonlinear detection is a sign that the probability to find a chaotic model is low. Since the procedure to detect nonlinearity in time series can be automated, the parameter space — in the present case, the delay versus the noise level — could intensively investigated to select potential values at which the global modeling procedure could be attempted.

V. CONCLUSION

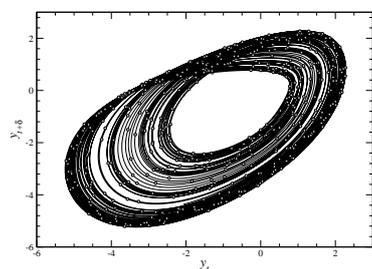
Identifying determinism underlying data from the real world that are resulting, most often from a combination of a deterministic component with a stochastic process remains quite challenging. We showed that noise titration failed to discriminate some nonlinear stochastic processes from chaotic behaviors. Nevertheless, comparison between one-step ahead predictions by nonlinear and linear models can be used to detect nonlinearity in noisy time series. A positive detection can be considered thus encouraging us to attempt a global model since a negative detection was always associated with a failure in applying a global modeling technique.

ACKNOWLEDGMENTS

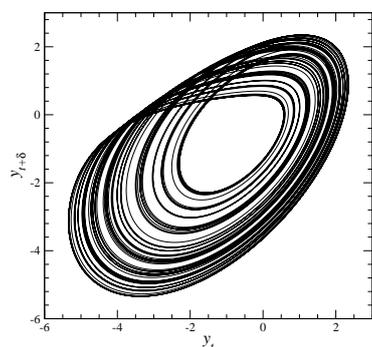
U. S. Freitas' post-doctoral position was supported by ADIR ASSOCIATION and *Région de Haute-Normandie*. We wish to thank Luis A. Aguirre (UMFG, Belo Horizonte, Brazil) for stimulating discussions and for providing the surrogate data used in this paper.



(a) Trajectory produced by the model



(b) Interpolated trajectory



(c) Original system

Fig. 6. (a) Chaotic attractor solution to the global model obtained from a y -time series contaminated with a noise level equal to 0.05 and using $\tau = 20$ (equivalent to a sampling time equal to 1.0s). (b) The trajectory (continuous line) is reconstructed through frequency-domain interpolation of the output of model simulation (circles). (c) Original noiseless attractor. These three figures are time delay plots.

REFERENCES

- [1] J. P. ECKMANN, S. OLIFSSON, D. RUELLE, S. CILIBERTO, Lyapunov exponents from time series, *Physical Review A*, **34** (6), 4971-4979, 1986.
- [2] A. PROVENZALE, L. A. SMITH, R. VIO & G. MURANTE, Distinguishing between low-dimensional dynamics and randomness in measured time series, *Physica D*, **58** (1-4), 31-49, 1992.
- [3] M. DÄMMIG & F. MITSCHKE, Estimation of Lyapunov exponents from time series: the stochastic case, *Physics Letters A*, **178**, (5-6), 385-394, 1993.
- [4] P. S. LAPLACE, *Théorie analytique des Probabilités*, Paris 1820.
- [5] J. LEYSEN & H. PASTIJN, Chaos, determinism and stochasticity, *Lectures Notes in Control and Information Sciences*, **111**, 646-653, 1988.
- [6] L. GLASS, Chaos and heart rate variability, *Journal of Cardiovascular Electrophysiology*, **10**, 1358-1360, 1999.

- [7] J. P. CRUTCHFIELD & B. S. MCNAMARA, Equations of motion from a data series, *Complex Systems*, **1**, 417-452, 1987.
- [8] L. A. AGUIRRE & S. A. BILLINGS, Dynamical Effects of over-parametrization in nonlinear models, *Physica D*, **80** (1,2), 26-40, 1995.
- [9] L. A. AGUIRRE & E. C. FURTADO, Building dynamical models from data and prior knowledge: the case of the first period-doubling bifurcation, *Physical Review E*, **76**, 046219, 2007.
- [10] J. THEILER, A. LOGTIN, S. EUBANK, B. GALDRIKIAN & J. D. FARMER, Testing for nonlinearity in time series: the method of surrogate data, *Physica D*, **58**, 77-94, 1992.
- [11] M. BARAHONA & C.-S. POON, Detection of nonlinear dynamics in short noisy time series, *Nature*, **381**, 215-217, 1996.
- [12] U. S. FREITAS, C. LETELLIER & L. A. AGUIRRE, Failure for distinguishing colored noise from chaos by the "Noise titration" technique, *Physical Review E*, **79**, 035201(R), 2009.
- [13] H. KANTZ & T. SCHREIBER, *Nonlinear time series analysis*, Cambridge University Press, 1997.
- [14] M. CENCINI, M. FALCIONI, E. OLBRICH, H. KANTZ & A. VULPIANI, Chaos or noise: difficulties of a distinction, *Physical Review E*, **62** (1), 427-437, 2000.
- [15] P. CVITANOVIĆ, Periodic orbits as the skeleton of classical and quantum chaos, *Physica D*, **51**, 138-151, 1991.
- [16] R. GILMORE & M. LEFRANC, *The topology of chaos*, Wiley, 2002.
- [17] C.-S. POON & M. BARAHONA, Titration of chaos with added noise, *Proceedings of the National Academy of Sciences (USA)*, **98**, 7107-7112, 2001.
- [18] O. E. RÖSSLER, An equation for continuous chaos, *Physics Letters A*, **57** (5), 397-398, 1976.
- [19] C. LETELLIER, L. LE SCHELLER, P. DUTERTRE, G. GOUESBET, Z. FEI & J. L. HUDSON, Topological Characterization and Global Vector Field Reconstruction from an experimental electrochemical system, *Journal of Physical Chemistry*, **99**, 7016-7027, 1995.
- [20] C. LETELLIER & L. A. AGUIRRE, Investigating nonlinear dynamics from time series: the influence of symmetries and the choice of observables, *Chaos*, **12**, 549-558, 2002.
- [21] C. LETELLIER, L. A. AGUIRRE & J. MAQUET, Relation between observability and differential embeddings for nonlinear dynamics, *Physical Review E*, **71**, 066213, 2005.
- [22] C. LAINSCSEK, C. LETELLIER & I. GORODNITSKY, Global modeling of the Rössler system from the z -variable, *Physics Letters A*, **314** (5-6), 409, 2003.



Ubiratan S. Freitas received his Electrical Engineer degree in 2000 at the Federal University of Minas Gerais, Brazil, and his Master degree in Electrical Engineering in 2001 at the same university. In 2006, he received his PhD in Computer Science at Brazil's National Institute for Space Research. Since 2007 he has a Post-doc position at the University of Rouen, France.



Christophe Letellier Christophe Letellier received B.S. degrees from the Physics Department (1991) at Paris VII and his Ph.D. from the Physics Department at Paris VII in 1994. He has been professor of physics at the University of Rouen since 1996. He is the head of a national research group from CNRS on Dynamics and Control of Complex Systems (GdR 2984 DYCOEC) since 2004. He is responsible for a course in Biomedical Engineering since 2007.

Improvement of Symbolic Transfer Entropy

Dimitris Kugiumtzis

Abstract—A number of measures have been proposed for the direction of the coupling between two time series, and transfer entropy (TE) has been found in recent studies to perform consistently well in different settings. Symbolic transfer entropy (STE) has been very recently proposed as a variation of the transfer entropy operating on the ranks of the components of the reconstructed vectors rather than the reconstructed vectors themselves. Here, an improvement of STE is proposed. Specifically, the ranks of the samples of the response system for given time steps ahead are computed with regard to the current reconstructed vector. The grounds of this modification are given and the new measure, called Transfer Entropy on Rank Vectors (TERV), is compared to STE and TE on different settings of state space reconstruction, time series length and observational noise. The results on two simulated systems have shown that the detection of the direction and strength of coupling is improved with TERV over both STE and TE.

Index Terms—bivariate time series, coupling, information measures, transfer entropy, rank vectors

I. INTRODUCTION

The fundamental concept for the dependence of one variable Y measured over time on another variable X measured synchronously is the Granger causality [1]. While Granger defined the direction of interaction in terms of the contribution of X in predicting Y , many variations of this concept have been developed, starting with linear approaches in the time and frequency domain (e.g. see [2], [3]) and extending to nonlinear approaches focusing on phase or event synchronization [4], [5], [6], comparing neighborhoods of the reconstructed points from the two time series [7], [8], [9], [10], [11], [13], [14], and measuring the information flow between the time series [15], [16], [17], [18], [19].

Among the different proposed measures we concentrate here on the last class of measures, and particularly on the transfer entropy (TE) [15] and the most recent variant of TE operating on rank vectors, called symbolic transfer entropy (STE) [18] (see also [20] for a similar measure). Other information measures, such as mean conditional mutual information [19] and coarse-grained transinformation rate [16] are rather similar to transfer entropy and are therefore not included in this study. There have been a number of comparative studies of information flow measures and other coupling measures giving varying results. In all the studies where TE was considered, it performed at least as good as the other measures [21], [22], [28]. The STE is proposed as an improvement of TE in real world applications, where noise may mask details of the fine structure, that can be better treated by coarse discretization using ranks instead of samples.

D. Kugiumtzis is with the Department of Mathematical, Physical and Computational Sciences, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece, e-mail: (see <http://users.auth.gr/dkugiu>).

We have studied STE and propose here a modification of it in order to improve the correct detection of the direction as well the strength of coupling when it is present. In the following, the TE and STE are presented briefly in Section II, and the proposed measure is described in Section III. Then the results of a simulation study comparing the proposed measure to TE and STE are presented in Section IV, and finally conclusions are given in Section V.

II. INFORMATION MEASURES

Transfer entropy (TE) is a measure of the flow of information from the driving system, denoted X , to the response system, denoted Y [15]. Supposing a representative quantity of system X is measured in terms of a scalar time series $\{x_t\}_{t=1}^N$ and respectively $\{y_t\}_{t=1}^N$ for Y , TE for the direction from X to Y can be defined in terms of the Shannon entropy $H(x) = \sum p(x) \log p(x)$ as

$$\text{TE}_{X \rightarrow Y} = -H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) + H(\mathbf{x}_t, \mathbf{y}_t) + H(y_{t+1}, \mathbf{y}_t) - H(\mathbf{y}_t), \quad (1)$$

or directly in terms of distribution functions as

$$\text{TE}_{X \rightarrow Y} = \sum p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) \log \frac{p(y_{t+1} | \mathbf{x}_t, \mathbf{y}_t)}{p(y_{t+1} | \mathbf{y}_t)}, \quad (2)$$

where $p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t)$, $p(y_{t+1} | \mathbf{x}_t, \mathbf{y}_t)$, and $p(y_{t+1} | \mathbf{y}_t)$ are the joint and conditional probability mass functions (pmf). The summation is over all the cells of a suitable partition of the joint variable vectors appearing as arguments in the pmfs or H . The points \mathbf{x}_t and \mathbf{y}_t appearing as arguments in eq.(1) and eq.(2) are reconstructed with the method of delays, so that we have $\mathbf{x}_t = [x_t, x_{t-\tau_x}, \dots, x_{t-(m_x-1)\tau_x}]'$ and $\mathbf{y}_t = [y_t, y_{t-\tau_y}, \dots, y_{t-(m_y-1)\tau_y}]'$, allowing different delay parameters τ_x, τ_y and embedding dimensions m_x, m_y for the systems X and Y , respectively.

The estimation of TE requires the estimation of the pmfs in eq.(2), or the probability density functions assuming the integral form and no binning. The pmfs are estimated directly by the relative frequency of occurrence of points in each cell, so the only complication is to choose a suitable binning [24], [23]. However, for high-dimensional reconstructions, the binning estimators are data demanding, and therefore estimators of the probability density functions are more appropriate for TE estimation, such as kernels [25], nearest neighbors [26], and correlation sums [27]. Using the latter approach, the TE estimator is given as

$$\text{TE}_{X \rightarrow Y} = \log \frac{C(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) C(\mathbf{y}_t)}{C(\mathbf{x}_t, \mathbf{y}_t) C(y_{t+1}, \mathbf{y}_t)}, \quad (3)$$

where $C(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t)$, $C(\mathbf{y}_t)$, $C(\mathbf{x}_t, \mathbf{y}_t)$ and $C(y_{t+1}, \mathbf{y}_t)$ are the correlation sums, which estimate the probability of

inter-points distances less than some given radius for the points of the form $[y_{t+1}, \mathbf{x}_t, \mathbf{y}_t]$, \mathbf{y}_t , $[\mathbf{x}_t, \mathbf{y}_t]$ and $[y_{t+1}, \mathbf{y}_t]$, respectively. The corresponding vector dimensions are $1 + m_x + m_y$, m_y , $m_x + m_y$ and $1 + m_y$. We use the Euclidean norm for the distances and define the radius as the product of 0.1, multiplied with the standard deviation of the data, and the square root of the vector dimension at each case (the latter is used to standardize the Euclidean norm). The use of 0.1 is a trade-off of having enough points within a radius to assure stable estimation of the point distribution and preserving neighborhoods to retain details of the point distribution. Still, for high-dimensional points, even this radius may be insufficient to provide stable estimation.

A. Symbolic transfer entropy

The authors in [18] derived the so-called *symbolic transfer entropy* (STE) as the transfer entropy defined on rank vectors formed by the reconstructed points. For each point \mathbf{y}_t , the ranks of its components in ascending order assign a rank vector $\hat{\mathbf{y}}_t = [r_1, r_2, \dots, r_{m_y}]$, where $r_j \in \{1, 2, \dots, m_y\}$ for $j = 1, \dots, m_y$, is the rank order of the component $y_{t-(j-1)\tau_y}$ (for two equal components of \mathbf{y}_t the smallest rank is assigned to the component appearing first in \mathbf{y}_t). Substituting y_{t+1} in eq.(1) with the rank vector at time $t + 1$, $\hat{\mathbf{y}}_{t+1}$, STE is defined as

$$\text{STE}_{X \rightarrow Y} = -H(\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) + H(\hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) + H(\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_t) - H(\hat{\mathbf{y}}_t), \quad (4)$$

or equivalently and with regard to eq.(2)

$$\text{STE}_{X \rightarrow Y} = \sum p(\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t) \log \frac{p(\hat{\mathbf{y}}_{t+1} | \hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t)}{p(\hat{\mathbf{y}}_{t+1} | \hat{\mathbf{y}}_t)}. \quad (5)$$

The estimation of STE from eq.(4) or eq.(5) is straightforward as the pmfs are naturally defined on the rank vectors and no binning or advanced estimator of probability density function is involved. There is a great advantage of using a rank vector $\hat{\mathbf{y}}_t$ over a binning of \mathbf{y}_t , say using b bins for each component: the possible vectors from binning are b^{m_y} while the possible combinations of the rank vectors are $m_y!$. For example, for $b = m_y = 4$, there are 256 cells from binning and only 24 combinations of rank vectors. Still, the estimation of the probability of occurrence of a rank vector becomes unstable as the dimension increases. Especially, for the joint vector of ranks $[\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{x}}_t, \hat{\mathbf{y}}_t]$ the dimension is $2m_y + m_x$, for which the equivalent of TE is $[y_{t+1}, \mathbf{x}_t, \mathbf{y}_t]$ and has dimension $1 + m_x + m_y$.

III. MODIFICATION OF SYMBOLIC TRANSFER ENTROPY

The conversion of the scalar y_{t+1} to the rank vector $\hat{\mathbf{y}}_{t+1}$ was chosen rather arbitrarily by the authors in [18] in order to express y_{t+1} in terms of ranks. Under this conversion, STE is not the direct analogue to TE using ranks instead of samples. The problem is not the use of y_{t+1} instead of \mathbf{y}_{t+1} in the definition of TE in eq.(1) or eq.(2) because $p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) = p(\mathbf{y}_{t+1}, \mathbf{x}_t, \mathbf{y}_t)$, as all components but y_{t+1} of the vector \mathbf{y}_{t+1} are also components of \mathbf{y}_t . The same holds for the conditional pmfs (and the same holds also for the two correlation sums in which y_{t+1} appears in eq.(3)).

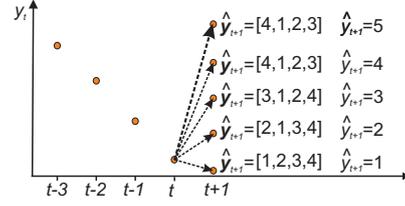


Fig. 1. Sketch of a position of samples $y_{t-3}, y_{t-2}, y_{t-1}, y_t$ and the possible rank position of y_{t+1} together with the corresponding rank vector $\hat{\mathbf{y}}_{t+1}$ defined for STE and the actual rank of y_{t+1} considering all 5 samples.

Let us first assume that $\tau_y = 1$. A first problem lies in the fact that when deriving the rank vector $\hat{\mathbf{y}}_{t+1}$ associated with \mathbf{y}_{t+1} , the rank of the last component of \mathbf{y}_t , y_{t-m_y+1} , is not considered. As an example, consider the vector $\mathbf{y}_t = [y_t, y_{t-1}, y_{t-2}, y_{t-3}]'$ with a corresponding rank vector $\hat{\mathbf{y}}_t = [1, 2, 3, 4]$, i.e. the samples decrease with time. If the decrease continues at the next time step, then $\hat{\mathbf{y}}_{t+1} = [1, 2, 3, 4]$, if y_{t+1} is between y_t and y_{t-1} then $\hat{\mathbf{y}}_{t+1} = [2, 1, 3, 4]$, if it is between y_{t-1} and y_{t-2} then $\hat{\mathbf{y}}_{t+1} = [3, 1, 2, 4]$, and finally if y_{t+1} is larger than y_{t-2} (the largest of all components in \mathbf{y}_{t+1}) then $\hat{\mathbf{y}}_{t+1} = [4, 1, 2, 3]$. The 4 possible scenarios are shown in Fig. 1.

The definition of rank vector $\hat{\mathbf{y}}_{t+1}$ accounts only for the possible rank positions of y_{t+1} with respect to the last $m_y - 1$ samples, ignoring the sample y_{t-m_y+1} , here y_{t-3} . With regard to the same example, $\hat{\mathbf{y}}_{t+1} = [4, 1, 2, 3]$ assigns to both cases $y_{t-2} < y_{t+1} < y_{t-3}$ and $y_{t-3} < y_{t+1}$ (see Fig. 1). Indeed there are 5 possible rank positions of y_{t+1} in the augmented vector $[y_{t+1}, y_t, y_{t-1}, y_{t-2}, y_{t-3}]$, as shown in Fig. 1. Thus for $m_y = 4$ there are $5! = 120$ different rank orders for the joint vector $[y_{t+1}, \mathbf{y}_t]$, but when forming the joint rank vector $[\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_t]$ (as in the computation of STE) there are only $4! \cdot (4!/3!) = 96$ possible rank orders. In general, there are $(m_y + 1)!$ possible rank orders for the joint vector $[y_{t+1}, \mathbf{y}_t]$, but STE estimation represents them in $m_y! \cdot \frac{m_y!}{(m_y-1)!}$ rank orders of $[\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_t]$.

The pmf of the rank vector derived from $[y_{t+1}, \mathbf{y}_t]$ and the rank vector $[\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_t]$ are shown in Fig. 2 for uniform white noise data and $m_y = 3$. There are $(m_y + 1)! = 24$ equiprobable rank orders for $[y_{t+1}, \mathbf{y}_t]$ (see Fig. 2a) but only $m_y! \cdot \frac{m_y!}{(m_y-1)!} = 18$ different vectors $[\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_t]$ are found, where $m_y! = 6$ of them have about double probability, each corresponding to two distinct rank orders that could not be distinguished. As a result, the Shannon entropy is underestimated here. Using $n = 10^{16}$ samples and the ranks of $[y_{t+1}, \mathbf{y}_t]$ we found $H = 4.5846$ bits and using $[\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_t]$ we found $H = 4.0865$ bits, while the true Shannon entropy is $H = -\log_2(1/24) = 4.5850$.

The situation changes if a further time step ahead is used. In general, allowing for y_{t+T} , where $T \geq 1$, the possible rank orders of $[y_{t+T}, \mathbf{y}_t]$ are again $(m_y + 1)!$, but for $[\hat{\mathbf{y}}_{t+T}, \hat{\mathbf{y}}_t]$ are $m_y! \cdot \frac{m_y!}{(m_y-T)!}$. For example, $m_y = 3$ and $T = 2$ gives 36 possible rank orders $[\hat{\mathbf{y}}_{t+T}, \hat{\mathbf{y}}_t]$, while the possible rank orders for 4 samples are 24. This increase holds in general for $T > 1$. For uniform white noise data, this results in overestimation of the true $H = 4.5850$ using the rank orders of $[\hat{\mathbf{y}}_{t+T}, \hat{\mathbf{y}}_t]$, $H =$

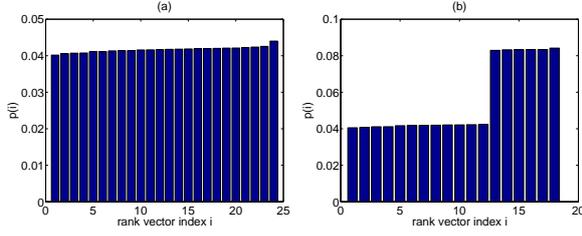


Fig. 2. (a) Estimated pmf for the ranks of $[y_{t+1}, y_t]$ with $m_y = 3$ (probabilities are in ascending order), where the samples y_t are from a uniform white noise time series of length $n = 10^{16}$. (b) Same as in (a) but for the rank vector $[\hat{y}_{t+1}, \hat{y}_t]$.

4.9736 ($n = 10^{16}$) while using the rank orders of $[y_{t+T}, y_t]$ we estimated $H = 4.5849$.

Thus we propose the following modifications to STE:

- 1) If $T = 1$, then in the definition of STE replace \hat{y}_{t+1} by \hat{y}_{t+1} , i.e. the rank of y_{t+1} in the augmented vector $[y_{t+1}, y_t]$.
- 2) If $T > 1$, then replace \hat{y}_{t+T} by $\hat{y}_t^T = [\hat{y}_{t+1}, \dots, \hat{y}_{t+T}]$, the ranks of y_{t+1}, \dots, y_{t+T} in the augmented vector $[y_{t+T}, y_{t+T-1}, \dots, y_{t+1}, y_t]$.

For $T > 1$, the proposal is to use all the ranks for times $t+1, \dots, t+T$ in order to keep track of the effect of X on the evolution of the time series of Y up to T time steps ahead. Similar reasoning for $T > 1$ was used for the measure of the coarse-grained transinformation rate [22] and we have used $T > 1$ also for TE [28]. Thus the proposed measure of transfer entropy with rank vectors (TERV) for T steps ahead is

$$\text{TERV}_{X \rightarrow Y}^T = -H(\hat{y}_t^T, \hat{x}_t, \hat{y}_t) + H(\hat{x}_t, \hat{y}_t) + H(\hat{y}_t^T, \hat{y}_t) - H(\hat{y}_t), \quad (6)$$

The TERV measure is the direct analogue to TE using ranks and extends the measure of information flow from X at time t to Y for a range of T time steps ahead t .

We note that when a lag $\tau_y > 1$ is used for the state space reconstruction of y_t , there are up to $m_y! \cdot m_y!$ different rank vectors $[\hat{y}_{t+T}, \hat{y}_t]$ in the computation of STE. On the other hand, for TERV there are $(T + m_y)!$ different rank vectors $[\hat{y}_t^T, \hat{y}_t]$. Thus for $\tau_y > 1$, the distortion of the domain of the rank vectors by STE may be large, e.g. for $\tau_y = 2$ and $T = 1$, the pmfs and entropies are computed on $(m_y + 1)!$ different rank orders for TERV and $m_y! \cdot m_y!$ for STE.

IV. ESTIMATION OF INFORMATION MEASURES FROM SIMULATED SYSTEMS

As it was shown for the example of uniform white noise the distortion of the domain of the rank vectors $[\hat{y}_t^T, \hat{y}_t]$ using the rank vectors $[\hat{y}_{t+T}, \hat{y}_t]$ instead has a direct effect on the estimation of entropy. While for uncoupled systems X and Y the entropy terms involving $[\hat{y}_{t+T}, \hat{y}_t]$ cancel out in the expression of TERV (and respectively for STE), in the presence of coupling some bias is introduced in the estimation of the coupling measure by STE. Using TERV instead this bias is removed.

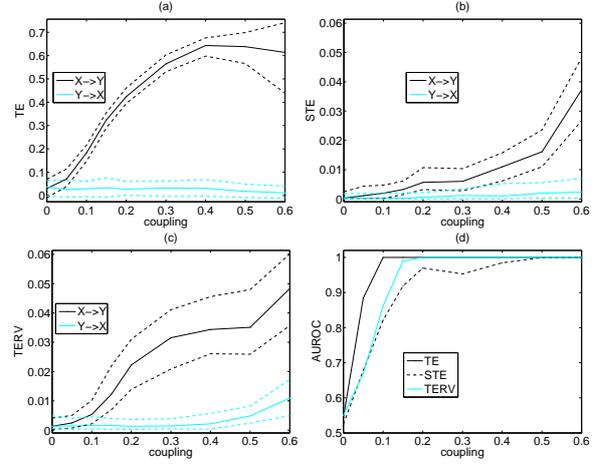


Fig. 3. (a) Median (solid line), 10% and 90% percentiles (dashed lines) of TE computed on 100 noise-free realizations of length $n = 1024$ from the system of two unidirectionally coupled Henon maps for varying coupling strengths. The other parameters are $T = 1$, $\tau_x = \tau_y = 1$ and $m_x = m_y = 2$. The direction $X \rightarrow Y$ is shown with black lines and $Y \rightarrow X$ with grey (online cyan) lines, as shown in the legend. (b) Same as (a) but for STE. (c) Same as (a) but for TERV. (d) AUROC computed on the 100 realizations for each of the two directions and for the measures TE, STE and TERV, as given in the legend.

We compare the estimation of coupling (strength and direction) in a system of two unidirectionally coupled Henon maps

$$\begin{aligned} x_{t+1} &= 1.4 - x_t^2 + 0.3x_{t-1} \\ y_{t+1} &= 1.4 - cx_t y_t + (1-c)y_t^2 + 0.3y_{t-1} \end{aligned}$$

with coupling strengths

$$c = 0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6.$$

The results on the coupling measures TE, STE and TERV for $T = 1$, $\tau_x = \tau_y = 1$ and $m_x = m_y = 2$ are shown for 100 noise-free time series of length $n = 1024$ in Fig. 3. For TE the correlation sums are used to estimate the entropies (see eq.(3)). TE seems to give the best detection of the correct direction of coupling even for very weak coupling, whereas STE performs worst. To quantify the level of discrimination of the correct direction of information flow, $X \rightarrow Y$, people often use the net information flow, defined as the difference of the coupling measure in the two directions. Here, we assess the level of discrimination in a statistical setting by computing the area under the receiver operating characteristic (ROC) curve on the samples of 100 realizations for each direction (e.g. see [29]). We denote the measure AUROC. For uncoupled systems, we expect that AUROC derived from a coupling measure is close to 0.5. For a measure to detect coupling with great confidence AUROC has to be close to 1. In Fig. 3d, the AUROC shows that TE detects coupling with great confidence for as low coupling strength as $c = 0.1$, followed by TERV reaching the same level of confidence at $c = 0.15$, while STE fails to provide confident discrimination of the two information flow directions unless the coupling gets strong (at $c = 0.5$).

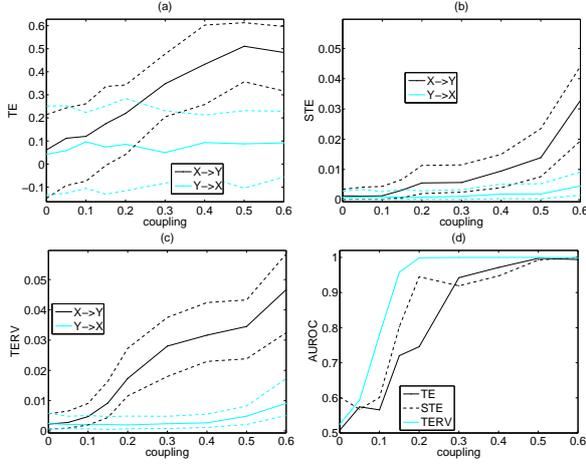


Fig. 4. As Fig. 3, but with 20% Gaussian white noise added to the data.

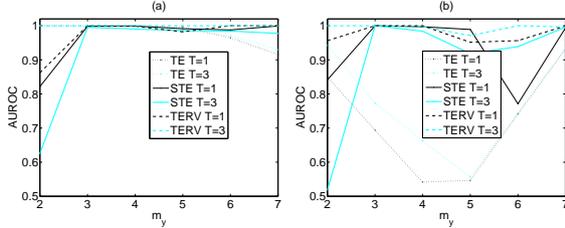


Fig. 5. (a) AUROC computed for different m_y ($m_x = m_y$) on 100 realizations of the weakly coupled Henon system ($c = 0.1$) for each of the two directions, for the measures TE, STE and TERV and for time steps ahead T as given in the legend. The time series are noise-free and $n = 1024$. (b) As in (a) but for $n = 4096$ and 20% additive Gaussian white noise.

The performance of the coupling measures changes in the presence of noise. For the same setup as that in Fig. 3, but adding to the bivariate time series 20% Gaussian white noise, we observe that TERV performs best, followed by STE, while TE has larger variance and fails to detect the correct direction of coupling when it is weak (see Fig. 4). It is notable that the discriminating power of TERV has not been affected much by noise. The AUROC for TERV increases faster with c than for the other two measures and reaches the highest level for $c = 0.2$, while both TE and STE reaches this level when coupling gets strong ($c = 0.5$).

We have estimated TE, STE and TERV on the coupled Henon system for different settings of embedding dimensions (keeping $m_x = m_y$), time steps ahead T , time series length n and noise level. For n small and m_y large and mostly for noisy time series, the computation of TE was not possible due to the lack of points within the given radius. Therefore the performance of TE was worse than for STE and TERV. This can be seen in Fig. 5, where the AUROC is shown for the three measures as a function of m_y for two settings of noise-free short time series and noisy but longer time series. It seems that estimating the information flow for $T = 3$ time steps ahead increases the detection of correct direction of weak coupling (e.g. $c = 0.1$ in the results of Fig. 5) for all but

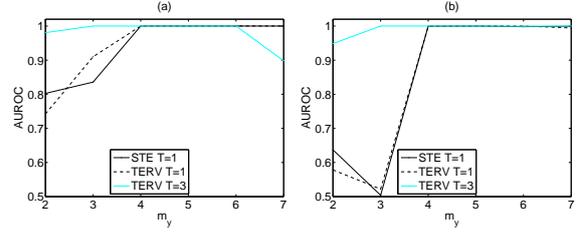


Fig. 6. (a) AUROC computed for different m_y ($m_x = m_y$) on 100 realizations of the weakly coupled Rössler – Lorenz system ($c = 0.5$) for each of the two directions and for the measures STE ($T = 1$) and TERV ($T = 1$ and $T = 3$), as given in the legend. The time series are noise-free and $n = 1024$. (b) As in (a) but when 20% Gaussian white noise is added to the data.

the STE measures. For the noise-free data, the differences in AUROC among the three measures are small, and TERV for $T = 3$ performs best giving AUROC=1 for all m_y (see Fig. 5a). TERV for $T = 3$ performs best also for the noisy data (see Fig. 5b) and generally the AUROC for TERV was almost always higher than for STE, which in turn was higher than for TE.

Similar simulations have been run on a Rössler system driving a Lorenz system given as (subscript 1 for Rössler, 2 for Lorenz)

$$\begin{aligned} \dot{x}_1 &= -6(y_1 + z_1) & \dot{x}_2 &= 10(x_2 + y_2) \\ \dot{y}_1 &= -6(x_1 + 0.2y_1) & \dot{y}_2 &= 28x_2 - y_2 - x_2z_2 + cy_1^2 \\ \dot{z}_1 &= -6(0.2 + z_1)(y_1 - 5.7) & \dot{z}_2 &= x_2y_2 - \frac{8}{3}z_2 \end{aligned}$$

for different coupling strengths c . In all parameter settings, TERV was improving the results of STE, especially when $T > 1$ was used (we tested for $T = 2$ and $T = 3$). An example is shown in Fig. 6 for weak coupling, $c = 0.5$. While STE and TERV score about the same in AUROC when $T = 1$, the use of a longer time horizon for the information flow ($T = 3$) in the estimation of TERV gives perfect detection of coupling direction for all but insufficient embeddings, $m_x = m_y = 2$ (a small decrease is observed for $m_y = 7$). This high performance of TERV and $T = 3$ holds also when noise is added to the data, while in this case STE and TERV with $T = 1$ fail to detect the coupling direction also for $m_y = 3$.

V. CONCLUSION

The use of ranks of consecutive samples instead of samples themselves in the estimation of the transfer entropy (TE) seems to gain robustness in conditions often met in real world applications, i.e. the presence of noise and the use of large embedding dimensions. This was confirmed by our results in this simulation study. Given that TE based on ranks can be a useful measure of information flow and direction of coupling, we have studied the recently proposed rank-based transfer entropy, termed symbolic transfer entropy (STE), and suggested a modified version of STE, which we termed TERV. The first modification is to use the rank of y_{t+1} (one time step ahead for the response time series) in the augmented reconstructed state vector y_t including also y_{t+1} , instead of

considering a whole rank vector for \mathbf{y}_{t+1} as done in STE. We showed that indeed this correction gives accurate estimation of the true entropy of the rank vector derived from the joint vector of \mathbf{y}_t and \mathbf{y}_{t+1} . Further, we suggested to allow the time step ahead to be $T > 1$ and use the ranks of all samples at the T future times ($\mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+T}$) derived from the augmented vector containing the current vector \mathbf{y}_t and these future samples. The proposed TERV measure was compared to TE and STE on two synthetic systems, a coupled map and a coupled flow, and the level of detection of the coupling direction was assessed by the area under the receiver operating characteristic curve (AUROC). TERV gave consistently higher AUROC than STE, and when the data were noisy also higher than TE. In particular, the use of $T > 1$ improved the performance of TERV.

There are other issues that have not been addressed in this study, such as the use of other settings of state space reconstruction (e.g. delays larger than one and different embedding dimensions for the driver and the response system). Also, there are problems in the estimation of the measures, including TERV, that have not been discussed here, such as the statistical significance of a measure when the systems are uncoupled, and the increase of a measure also for the opposite (wrong) coupling direction when the coupling strength increases. This study is by no means extensive or complete and the measures should also be compared in many different systems (identical and non-identical).

REFERENCES

- [1] J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Acta Physica Polonica B*, vol. 37, pp. 424 – 438, 1969.
- [2] L. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463 – 474, 2001.
- [3] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistriz, D. Klan, R. Bauer, J. Timmer, and H. Witte, "Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems," *Signal Processing*, vol. 85, no. 11, pp. 2137 – 2160, 2005.
- [4] M. Rosenblum and A. Pikovski, "Detecting direction of coupling in interacting oscillators," *Physical Review E*, vol. 64, no. 4, p. 045202, 2001.
- [5] R. Quian Quiroga, T. Kreuz, and P. Grassberger, "Event synchronization: A simple and fast method to measure synchronicity and time delay patterns," *Physical Review E*, vol. 66, no. 4, p. 041904, 2002.
- [6] D. Smirnov and B. Bezruchko, "Estimation of interaction strength and direction from short and noisy time series," *Physical Review E*, vol. 68, no. 4, p. 046209, 2003.
- [7] A. Cenys, G. Lasiene, K. Pyragas, J. Peinke, and J. Parisi, "Analysis of spatial correlations in chaotic systems," *Acta Physica Polonica B*, vol. 23, no. 4, pp. 357 – 365, 1992.
- [8] S. Schiff, P. So, T. Chang, R. Burke, and T. Sauer, "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble," *Physical Review E*, vol. 54, pp. 6708 – 6724, 1996.
- [9] J. Arnhold, P. Grassberger, K. Lehnertz, and C. Elger, "A robust method for detecting interdependences: Application to intracranially recorded EEG," *Physica D*, vol. 134, pp. 419 – 430, 1999.
- [10] R. Quian Quiroga, J. Arnhold, and P. Grassberger, "Learning driver-response relationships from synchronization patterns," *Physical Review E*, vol. 61, no. 5, pp. 5142 – 5148, 2000b.
- [11] R.G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz, "Bivariate surrogate techniques: Necessity, strengths, and caveats," *Physical Review E*, vol. 68, p. 066202, 2003.
- [12] J. Bhattacharya, E. Pereda, and H. Petsche, "Effective detection of coupling in short and noisy bivariate data," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 33, no. 1, pp. 85 – 94, 2003.
- [13] M. Romano, M. Thiel, J. Kurths, and G. C., "Estimation of the direction of the coupling by conditional probabilities of recurrence," *Physical Review E*, vol. 76, no. 3, p. 036211, 2007.
- [14] D. Chicharro, A. Ledberg, and R.G. Andrzejak, "A new measure for the detection of directional couplings based on rank statistics," in *BMC Neuroscience, 7th Annual Computational Neuroscience Meeting*, vol. 9, Portland, USA, 2008, p. 148. [Online]. Available: <http://www.biomedcentral.com/1471-2202/9/S1/P148>
- [15] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461 – 464, 2000.
- [16] M. Paluš, V. Komárek, T. Procházka, Z. Hrnčíř, and K. Šterbová, "Synchronization and information flow in EEGs of epileptic patients," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 5, pp. 65–71, 2001.
- [17] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series," *European Physical Journal B*, vol. 30, pp. 275 – 281, 2002.
- [18] M. Staniek and K. Lehnertz, "Symbolic transfer entropy," *Physical Review Letters*, vol. 100, no. 15, p. 158101, 2008.
- [19] M. Vejmelka and M. Paluš, "Inferring the directionality of coupling with conditional mutual information," *Physical Review E*, vol. 77, no. 2, p. 026214, 2008.
- [20] A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. V. E. McClintock, and H. Kantz, "Direction of coupling from phases of interacting oscillators: A permutation information approach," *Physical Review Letters*, vol. 100, no. 8, p. 084101, 2008.
- [21] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, "Methods for quantifying the causal structure of bivariate time series," *Journal of Bifurcation and Chaos*, vol. 17, no. 3, pp. 903 – 921, 2007.
- [22] M. Paluš and M. Vejmelka, "Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections," *Physical Review E*, vol. 75, no. 5, p. 056211, 2007.
- [23] A. Papana and D. Kugiumtzis, "Evaluation of mutual information estimators on nonlinear dynamic systems," *Complex Phenomena in Nonlinear Systems*, vol. 11, no. 2, pp. 225–232, 2008.
- [24] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1991.
- [25] B. Silverman, *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall, 1986.
- [26] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, p. 066138, 2004.
- [27] C. Diks and S. Manzan, "Tests for serial independence and linearity based on correlation integrals," *Studies in Nonlinear Dynamics & Econometrics*, vol. 6, no. 2, Art. No 2, pp. 1–22, 2002.
- [28] A. Papana and D. Kugiumtzis, "Detection of directionality of information transfer in nonlinear dynamical systems," topics on Chaotic Systems, selected papers from CHAOS 2008 International Conference, Chania, Crete, World Scientific (to be published).
- [29] D. J. Hand and R. J. Till, "A simple generalization of the area under the ROC curve to multiple class classification problems," *Machine Learning*, vol. 45, pp. 171 – 186, 2001.



Dimitris Kugiumtzis Dimitris is Assistant Professor at the Department of Mathematical, Physical and Computational Sciences, Aristotle University of Thessaloniki (AUTH) since 2001. He has a B.Sc. in Mathematics (AUTH, 1987), M.Sc. and Ph.D. in Informatics (University of Oslo, Norway, 1991 and 1997). He was PostDoc at the Max-Planck Institute for Physics of Complex Systems, Dresden (1998 - 1999) and lecturer B at the Department of Statistics, University of Glasgow (2000 - 2001). His main research area is time series analysis in conjunction with dynamical systems and chaos as well as computational statistics and data mining. Applications of interest include the analysis of physiological time series, such as pre-epileptic electroencephalograms, as well as time series in finance, climate and seismology.

Assessing the degree of synchronization in time series using symbolic representations

Roberto Monetti, Wolfram Bunk, Thomas Aschenbrenner and Stephan Springer

Abstract—We extend a recently proposed methodology to characterize synchronization in time series using symbolic representations (R. Monetti et al., *Phys. Rev. E* 79, 046203 (2009)). In this approach, we define symbols which are expected to contain information of the system at different time scales. Here,, a feature vector is created at every position in the time series using a set of local filters. Every feature vector is subsequently mapped into a symbol through the rank-order of its values. A representation of a time series results after mapping all feature vectors into symbols. The dynamics of coupled systems is analyzed using a transcription scheme which allows us to assess the degree of synchronization. A prototype non-linear system is used as a test bed for our method. This approach is also employed in a longitudinal case study of a child with frontal lobe epilepsy (FLE) and tested against an age-matched control group.

Index Terms—Time Series, Synchronization, Symbolic Representations, Transcription Scheme, Information measures.

I. INTRODUCTION

SYNCHRONIZATION of oscillatory systems is a phenomenon broadly discussed in different fields of science and technology. It is also observed in chaotic oscillators where the states of complete, generalized and phase synchronization have been described theoretically [1]. Experimental observations of the various synchronization states can be found in the cardiorespiratory system [2], in the cells of paddlefish [3] and extended ecological systems [4]. In the human brain, synchronization plays an important role in epileptic seizures and its quantitative description is relevant for diagnostic purposes ([5], [6], [7], [8]). It has been argued that chronic epilepsy modifies the brains state even in interictal periods revealing altered anatomical, biochemical and functional properties [9]. Interictal EEG, even without epileptiform abnormalities, has a number of characteristic differences from the EEG of healthy subjects. These discrepancies should vanish under a successful medical therapy. Hence, an important aspect of interictal EEG assessment is the evaluation of spatio-temporal synchronization.

Here, we introduce a method to characterize synchronization in coupled systems where information measures are obtained using symbolic representations of feature vectors. The present approach is a variant of the methodology introduced in [11]. Section II is devoted to the description of symbolic feature vector representations, the transcription scheme, the concept

R. Monetti, W. Bunk, and T. Aschenbrenner are with the Max-Planck-Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany e-mail: monetti@mpe.mpg.de.

S. Springer is with the Heckscher-Klinikum für Kinder- und Jugendpsychiatrie und Psychosomatik, München; Kinderklinik und Poliklinik im Dr. von Haunerschen Kinderspital, Klinikum der Universität München, Germany

of order classes, and the information measures to quantify the degree of synchronization. The application of the method to a coupled chaotic system and to EEG data in the context of epilepsy is presented in section III.

II. METHOD

Let x be a time series and consider a set of operators \mathcal{F}_i suitable to extract local features of the signal at different time scales, i.e. $\mathcal{B} = \{\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}, \dots, \mathcal{F}_{\tau_p}\}$ where τ_i indicates a time scale. Every point in the time series can be represented by a feature vector $\vec{v}_i = (\mathcal{F}_{\tau_1}(x_i), \mathcal{F}_{\tau_2}(x_i), \dots, \mathcal{F}_{\tau_p}(x_i))$ where local coarse-grained information at different time scales is compiled. Now, a symbol V is defined as the rank-ordered indices of the components of \vec{v}_i . For instance, for $\vec{v} = (1.6, 1.3, 1.4, 1.5)$, the symbol associated is $V = (3, 0, 1, 2)$. In contrast to the ordinal time series analysis introduced by Bandt et al. [12] and also applied in [11] where sequences of length p extracted from time series are transformed into symbols, the symbolic representation proposed in this work points to describe changes in local features when evaluated at different time scales.

A relevant issue for the application of this method is to obtain representative feature vectors which provide local information of the system at different time scales. Thus, the choice of the operator \mathcal{F} and the time scales must be guided by a previous knowledge of the system under consideration and its typical time scales. Here, we extract from the time series p local features at different time scales using the fourth order Gaussian wavelet \mathcal{W} , i.e. the normalized 4th derivative of the Gaussian function. Figure 1 indicates how a symbolic representation of a signal is constructed. The signal (lower curve) is analyzed using a set of fourth order Gaussian wavelet (see Fig. 1 middle curves). Then, for every time t the set of five wavelet coefficients defines a feature vector which is subsequently mapped into a symbol as explained above. The resulting symbolic representation is shown in Fig. 1 (upper curve).

The concept of transcription between symbols A_1 and A_2 was already introduced in [11] and refers to the permutation T which satisfies $T[A_1] = A_2$, where the operation must be interpreted as a composition. The action of symbol T is defined as follows. Let $A_1 = (j_0, j_1, \dots, j_{p-1})$ and $T = (k_0, k_1, \dots, k_{p-1})$. Then,

$$T[A_1] = (j_{k_0}, j_{k_1}, \dots, j_{k_{p-1}}). \quad (1)$$

The set of symbols form a finite non-Abelian group of order $p!$ with operation T known as the symmetric group S_p .

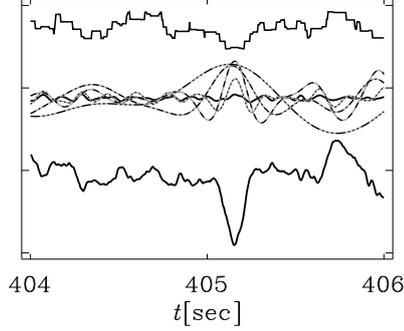


Fig. 1. Example of a piece of an EEG signal (lower curve), the corresponding five wavelet transform (middle curves), and the symbolic representation (upper curve) for $p = 5$ using the fourth order gaussian wavelet at time scales $\tau = 0.02$ (full line), $\tau = 0.06$ (dotted line), $\tau = 0.12$ (dashed line), $\tau = 0.2$ (dashed-dotted line), and $\tau = 0.4$ (dashed-tri-dotted line).

Thus, for every symbol T in S_p an integer number N can always be found such as the composition $T^N = I$ where $I = (0, 1, \dots, p-1)$ is the identity symbol (neutral element of S_p). Since N is not unique, we will always refer to N as the smallest integer which satisfies the above mentioned power relation. A group of symbols which satisfy the same power relation defines an order class \mathcal{C}_N . A detailed discussion about the properties of order classes can be found in [11].

We generate symbolic representations for the time series as explained above and focus on the probability density of transcriptions between symbolic representations (called in the following source and target symbolic representations). The probability density of transcriptions $P_T(p)$ can be written as follows

$$P_{T_k}(p) = \sum_{\Omega=\{(i,j): T_k[X_i]=X_j\}} P^C(X_i, X_j), \quad (2)$$

where X_i (X_j) is a symbol belonging to the source (target) symbolic representation and $P^C(X_i, X_j)$ is the joint probability density. Let $P^{(1)}(X_i)$ and $P^{(2)}(X_j)$ be the marginal probability densities of the symbols X_i and X_j , respectively. The matrix $M_{i,j} = P^{(1)}(X_i)P^{(2)}(X_j)$ is the probability density matrix of transcriptions for two independent processes. In this case, the probability density of transcriptions $P_T^e(p)$ can be evaluated as follows

$$P_{T_k}^{ind}(p) = \sum_{\Omega=\{(i,j): T_k[X_i]=X_j\}} M_{i,j}, \quad (3)$$

The aim is to find an information measure to assess how much P_T deviates from P_T^{ind} . A natural choice to quantify the contrast between probability densities is the Kullback-Leibler (KL) entropy

$$E_{KL}(P, P^{ind}) = \sum_i P_{T_i}(p) \log(P_{T_i}(p)/P_{T_i}^{ind}(p)). \quad (4)$$

Since the E_{KL} is not a symmetric quantity, we use the

following symmetrized form [13]

$$S_{KL}(p) = \frac{E_{KL}(P, P^{ind})E_{KL}(P^{ind}, P)}{E_{KL}(P, P^{ind}) + E_{KL}(P^{ind}, P)}. \quad (5)$$

One can demonstrate that $S_{KL}(p)$ remains always finite and that $S_{KL}(p) \leq \min(E_{KL}(P, P^{ind}), E_{KL}(P^{ind}, P))$. Other properties are discussed in [13]. It can be shown that T and T^{-1} belong to the same order class [11]. Thus, $S_{KL}(p)$ for transcriptions inside a class is a suitable invariant measure under the interchange of source and target time series. In the following, we will refer to S_{KL} obtained using the probability density of transcriptions in order class \mathcal{C}_N as S_{KL}^N .

III. APPLICATIONS

We apply the method to a bi-directionally coupled Roessler-Roessler system [10] defined by the following set of equations

$$\begin{aligned} \dot{x}_{1,2} &= -w_{1,2}y_{1,2} - z_{1,2} + k(x_{2,1} - x_{1,2}), \\ \dot{y}_{1,2} &= w_{1,2}x_{1,2} + 0.165y_{1,2}, \\ \dot{z}_{1,2} &= 0.2 + z_{1,2}(x_{1,2} - 10). \end{aligned} \quad (6)$$

where $w_1 = 0.99$ and $w_2 = 0.95$ are the mismatch parameters. All time series were generated using a fourth-order Runge-Kutta method with an increment $\delta t = 0.001$ and the following initial conditions: $x_1(0) = -0.4$, $y_1(0) = 0.6$, $z_1(0) = 5.8$, $x_2(0) = 0.8$, $y_2(0) = -2$, and $z_2(0) = -4$. Results were saved at intervals $\Delta t = 0.01$. This chaotic system exhibits a rich synchronization behavior which ranges from phase ($k \approx 0.036$) to lag ($k \approx 0.14$) and finally complete synchronization as the coupling parameter k is increased [10]. The results presented here were obtained using the x -components of the Roessler subsystems. We considered time series of length $L = 2^{19}$ (~ 775 orbits). For every time series, we generate p -dimensional feature vectors $\vec{v}_i = (\mathcal{W}_{\tau_1}[x(t_i)], \mathcal{W}_{\tau_2}[x(t_i)], \dots, \mathcal{W}_{\tau_p}[x(t_i)])$ ($i = \{0, 1, \dots, L-1-\tau_p\}$) for $p = 6$ and then the symbolic representation. We have chosen wavelet scales which approximately cover the time span of one orbit of the Roessler system ($676\Delta t$), namely $\tau_1 = 10$, $\tau_2 = 21$, $\tau_3 = 43$, $\tau_4 = 87$, $\tau_5 = 175$, and $\tau_6 = 350$.

Figure 2 shows the Kullback-Leibler entropies for all transcriptions S_{KL} and transcription in different order classes S_{KL}^N for $p = 6$. For small values of the coupling constant k , the time series behave independently since the Roessler subsystems are almost uncoupled. For $k \in [0, 0.036]$, all Kullback-Leibler entropies indicate that the actual dynamics hardly deviates from that of the independent processes. All curves increase at $k \sim 0.036$ due to the transition to phase synchronization and display a peak at $k \approx 0.061$ which corresponds to period three window [10]. For stronger coupling k , curves increase again rather monotonically till $k \sim 0.11$. For $k \in [0.11, 0.145]$, S_{KL} , S_{KL}^3 , S_{KL}^5 , and S_{KL}^6 display strong fluctuations revealing the presence of 'intermittent-lag-synchronization'. This particular synchronization regime is characterized by synchronization periods interrupted by bursts of non-synchronized activity [10], [11]. The strong fluctuations displayed by these Kullback-Leibler entropies sharply vanish at the onset of lag-synchronization ($k \sim 0.145$). Lag-synchronization is defined

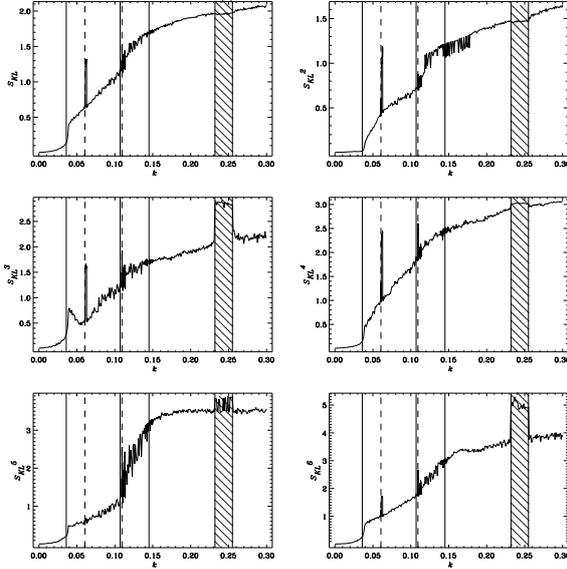


Fig. 2. Kullback-Leibler entropy for all transcriptions S_{KL} and for transcriptions in the order classes S_{KL}^N using a sequence length $p = 6$. Vertical full lines from left to right indicate transitions to phase-synchronization ($k \sim 0.036$), intermittent-lag-synchronization ($k \sim 0.11$), and lag-synchronization ($k \sim 0.145$), respectively. Vertical dashed lines at $k \sim 0.061$ and $k \sim 0.11$, and hatched areas ($k \in [0.232, 0.256]$) indicate periodic windows. The values of the coupling constant for the onset of phase synchronization and the first periodic window were taken from [10].

through the condition $x_1(t + \tau_0) = x_2(t)$, i.e. the coincidence of the time series when shifted in time by a constant time lag τ_0 . Curves increase rather monotonically in the interval $k \in [0.145, 0.30]$ reflecting stronger synchronization. This trend is only interrupted within the coupling range $k \in [0.232, 0.256]$ where a period five window occurs. It should be noted that some order classes are better suited than others to reveal particular features of the system as periodic windows (see Fig. 2). In this sense, different order classes provide complementary information of the coupled system. The results obtained using this new symbolic approach are in complete agreement with those reported in [10], [11].

A. Application to EEG

In this study, the single case results of the analysis of several EEG recordings covering two years of a child (patient A) are compared with the outcome of an age-matched control group. The first four EEG of patient A which entered this study, are taken at an age of 12.1 years and the whole follow-up spans a period of almost two years of successful therapy. The age-matched control group consists of three patients whose EEG recordings cover an age range from 10.6 to 14.4 years. All EEG recordings of the control group were classified as normal. These three patients showed seizure freedom for more than five years of follow-up. The positioning of the electrodes followed that of the standardized 10-20-International System of Electrode Placements. Every EEG recording consists of 21 synchronously obtained time series. Our data base is made up

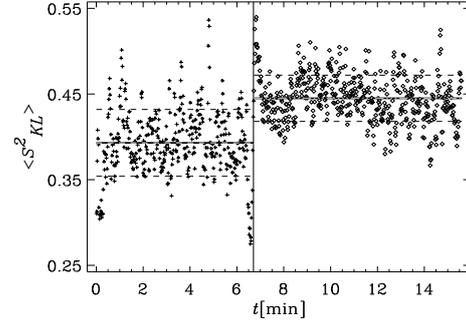


Fig. 3. Time course of the spatial mean of the Kullback-Leibler entropy for transcriptions in order class 2 S_{KL}^2 for an EEG recording of patient A. At about 6.7 min (vertical line) an epileptic seizure starts. The seizure attack is initiated by a loss of synchronization, while during the seizure the correlation is generally increased, thus S_{KL}^2 values are higher. The horizontal full lines mark the mean values of the cloud of points before and after the epileptic attack and the dashed lines indicate the corresponding standard deviations.

by a number of twenty multichannel EEG recordings: Eight EEG are recorded from patient A (age range 12.06-13.96 y). Twelve EEG are derived from the control group (three patients, age range 10.61-14.37 y). Every EEG record measures brain activity for at least 10 minutes at a sampling rate of 250 Hz and a signal depth of 16 bits. Using the same EEG data sample, Bunk et al. [?] have recently presented a detailed description of interictal EEG in pediatric frontal lobe epilepsy, where the performance of a variety of synchronization measures was compared.

To obtain a time dependent characterization of brain activity the data has been analyzed with a sliding window technique using a window size of slightly more than two seconds. The consecutive windows overlap by half the window size, which results in more than 900 separate windows for a typical EEG recording of 15 minutes and a time resolution of ~ 1 second.

In the case study of the patient suffering from frontal lobe epilepsy (FLE) we find that the synchronization level is significantly increased during its clinical manifestation. In fact, Fig. 3 shows the time evolution of the spatial mean of the Kullback-Leibler entropy $\langle S_{KL}^2 \rangle$ for patient A where a seizure episode occurs. The vertical full line in Fig. 3 indicates the onset of the seizure. A comparison of the mean values of $\langle S_{KL}^2 \rangle$ before and after the epileptic seizure reveal a stronger synchronization during the seizure attack (see Fig. 3)

Figure 4 shows the temporal mean of the spatial averaged Kullback-Leibler entropies calculated only during interictal activity of the FLE patient and for the control group versus patient age. We define the contrast, i.e. the discriminative power of an information measure S between the control group and the acute phase of patient A as

$$C = \frac{(\langle S \rangle^{CG} - \langle S \rangle^A)^2}{\sigma^{CG}(S)^2 + \sigma^A(S)^2}. \quad (7)$$

The best contrast ($C = 1.85$) is obtained using S_{KL}^4 . Figure 4 indicates that after successful medical suppression of the acute epileptic state the clinical picture still points to the

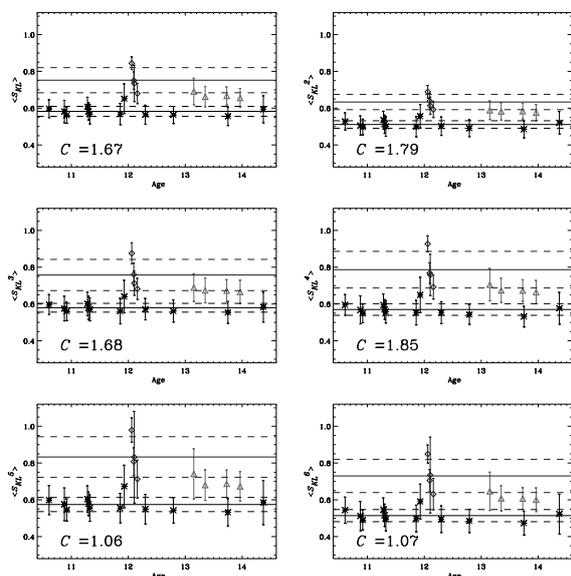


Fig. 4. Mean Kullback-Leibler entropy $\langle S_{KL} \rangle$ evaluated for transcriptions in different order classes for patient A (acute phase: grey diamonds, non-acute phase: light grey triangles) and control group (black squares). The filled circles denote the averaged mean $\langle S_{KL} \rangle$ and the error bars indicate the variability for single EEG recordings. The horizontal lines label the mean $\langle S_{KL} \rangle$ for patient A and control group and the dashed lines label the respective standard deviations. C indicates the values of the contrast between the acute phase of patient A and the control group.

presence of a frontal lobe syndrome (see grey diamonds). As a consequence of a long term effective therapy these symptoms disappeared. In the course of the two year lasting follow up of the patient, the synchronization measures converge to the values obtained for the members of the control group (see Fig 4 light grey triangles).

IV. CONCLUSION

We presented a variant of the methodology proposed in [11] to analyze synchronization in time series using symbolic representations. In this approach, symbols were defined via rank-ordering of a sequence of local features extracted using a suitable operator. The relationship between the time series is studied by means of a transcription scheme applied to the symbolic representations. The synchronization strength is quantified using information measures of the probability density of transcriptions belonging to different order classes. Our results indicate that some order classes may highlight particular features of the coupled system. Thus, they provide complementary information of the synchronization dynamics. We demonstrated the applicability of our method to EEG in the context of epilepsy. The newly introduced information measures proved to be sensitive to changes in the brain activity induced by a effective therapy. Symbolic representations generated using local signal features can easily be generalized to higher dimensional signals and are expected by construction to be robust against noise.

REFERENCES

- [1] S. Boccaletti, J. Kurths, G. Osipov, DL Valladares, and CS Zhou, *Phys. Rep.* **336** 1, 2002.
- [2] C. Schafer, MG Rosenblum, J. Kurths, and HH Abel, *Nature* **392** 239, 1998.
- [3] A. Neiman, X. Pei, D. Russell, W. Wojtenek, L. Wilkens, F. Moss, H.A. Braun, M.T. Huber, and K. Voigt, *Phys. Rev. Lett.* **82** 660, 1999.
- [4] B. Blasius, A. Huppert, and L. Stone, *Nature* **399** 354, 1999.
- [5] P. Tass, MG Rosenblum, MG Weule, J. Kurths, A. Pikovsky, J. Volkman, A. Schnitzler, and HJ Freund, *Phys. Rev. Lett.* **81** 3291, 1998.
- [6] K.K. Jerger, T.I. Netoff, J.T. Francis, T. Sauer, L. Pecora, S.L. Weinstein, et al., *J. of Clinical Neurophysiology*, **18** 259, 2001.
- [7] M. Palus, V. Komárek, T. Procházka, Z. Hrnčíř, K. Sterbová, *IEEE Engineering in Medicine and Biology Magazine*, **20** 259, 2001.
- [8] A. Hedge, D. Erdogmus, D.S. Shiau, J.C. Principe, C.J. Sackellares, *Computational Intelligence and Neuroscience*, doi:10.1155/2007/83416 1, 2007.
- [9] A.A. Fingelkurts, A.A. Fingelkurts, A.Y. Kaplan, *Clinical Neurophysiology*, **117** 208, 2006.
- [10] M.G. Rosenblum, A. Pikovsky, and J. Kurths, *Phys. Rev. Lett.* **78** 4193, 1997.
- [11] R. Monetti, W. Bunk, T. Aschenbrenner, and F. Jamitzky, *Phys. Rev. E* **79** 046203, 2009.
- [12] C. Bandt and B. Pompe, *Phys. Rev. Lett.* **88** 174102, 2002.
- [13] D. Johnson and S. Sinanović, Available from <http://citeseer.ist.psu.edu/johnson01symmetrizing.html>, 2001.
- [14] W. Bunk, T. Aschenbrenner, G. Kluger, and S. Springer, arXiv:0905.3911v1, [physics.med-ph], 2009.

Roberto Monetti obtained the M.S. and Ph.D in Physics from the University of La Plata, Argentina in 1990 and 1995, respectively. He was post-doctoral fellow at the Dept. of Physics of TU-Muenchen, Germany from 1996 to 1998 and research fellow of the CONICET, Argentina from 1998 to 2004. In 2002, he joined the Max-Planck-Institut fuer extraterrestrische Physik where he is currently working in the fields of signal and image processing.

Characterization of pre-seizure states by a multi-signal analysis of scalp EEG and ECG signals during slow wave sleep

Stavros NIKOLOPOULOS , Mario VALDERRAMA , P. MILANI , V. NAVARRO , Michel LE VAN QUYEN

The slow wave sleep (SWS) increases the mean density of electroencephalographic paroxysmal activities whatever the epileptic syndrome. These observations have raised the possibility that SWS, well characterized by slow oscillations (0.5-1 Hz) in scalp EEG, may provide a new window into preictal modifications in the epileptic brain. In particular, the presence of specific changes in heart rate variability (HRV) during sleep, not correlated with seizures, have been reported by several studies in the past, and both ictal and interictal modifications of heart rate regulation have been described. Here, based on the known interactions between brain and cardiac activities, we combined analysis of scalp EEG and ECG signals during SWS for 8 healthy subjects and 15 epileptic patients suffering from partial epilepsy (4 frontal lobe, 8 temporal lobe, 3 cryptogenic epilepsy). Sleep was polygraphically recorded and a series of 5-min epochs were chosen from sleep stages 3 and/or 4. Electrocardiographic signals were analyzed for automatic detection of R-waves and, subsequently, a series of time- and frequency-domain measures were calculated. Furthermore, HRV and slow EEG waves was also characterized by the approximate entropy and a modified root mean square analysis of a random walk, named detrended fluctuations analysis (DFA), which proved efficacy as a diagnostic tool and advantages over existing linear methods. In a first stage, we compared the measures between healthy subjects and the patients and observed that epileptic subjects tended to show an overall lower HRV in both time- and frequency-domain parameters. Among the different bands, this decrease was most evident for the high-frequency band (HF) absolute power, reflecting altered cardiac vagal activity. In a second stage, we have analyzed ECG and EEG during SWS in continuous long-term recordings during at least 7 days duration including SWS periods and a sufficient number of clinically manifest seizures (at least 4). Each seizure was classified according to subject group, sleep/wake state, and time of day of seizure occurrence based on video-EEG-monitoring. Analysis of spectral and dynamical measures of HRV and EEG slow waves showed variable patient-specific fluctuations across the different nights and, for a majority of patients (10/15), changes were more pronounced during the nights before seizures. In conclusion, our findings suggest that easily obtainable noninvasive EEG-ECG can provide during SWS useful information for detecting relative changes in brain dynamics of epileptic patients. We speculate that the SWS represent a cyclic modulation of cortical gross excitability, inducing neuronal and autonomic cardiovascular fluctuations that may provide a new characterization of epileptic seizure susceptibility.

This work is funded by the european FP7 project “Evolving Platform for Improving Living Expectation of Patients Suffering from Ictal Events” (EPILEPSIAE) on advanced ICT for Risk Assessment and Patient Safety.

Complex networks as a tool for nonlinear time series analysis

Michael Small *Member, IEEE*, and Xiaoke Xu

Abstract—Several algorithms have been recently proposed to generate complex networks from time series. In each case the basic idea is to generate a geometric object that accurately represents certain properties of the dynamics. Then, by estimating features of that geometric object it is possible to learn something about the complex nonlinear dynamics that generated the original time series. We describe the relevant features of these methods and show how they can be applied to real experimental time series. We apply these methods to a time series of Canadian lynx numbers and show that these methods can reliably differentiate and classify dynamics as either chaotic, periodic or noisy.

Index Terms—complex networks from time series, nonlinear time series analysis, chaos, Canadian lynx time series

I. INTRODUCTION

THREE methods have recently been proposed to generate complex networks from time series. The first method, proposed by Zhang and Small [13] maps individual cycles in a pseudo-periodic (i.e. roughly cyclic or oscillatory) time series to nodes in a complex network, with links being drawn between nodes which correspond to cycles which are close. The method was based on earlier work by the same authors [12] and was shown to allow one to differentiate between low-dimensional chaotic flow dynamics and a noisy periodic orbit. Based on this method, Lacasa and colleagues [3] described a new method to map individual time series points to nodes of a complex network with nodes being connected based on a scale invariant “visibility” criterion. By construction, this method was therefore effective in identifying fractal or self-similar process. Finally, Xu, Zhang and Small [10] extended their previous work [13] and proposed a more generic method applicable to arbitrary time series. It is this method which will be the focus of the current paper.

In the next section (Sec. II) we describe the basic algorithm and discuss the application of this method to idealised signals. In Sec. III we demonstrate the application of this method to short and noisy experimental population dynamical time series. Finally, in Sec. IV we provide some brief concluding remarks.

II. NETWORKS

The algorithm proposed in [13] is limited to time series exhibiting a clearly oscillatory dynamics. The method is therefore well suited to situations where low-dimensional chaos is a plausible model of the observed dynamics, but not otherwise. One of the greatest strengths of this method is that by comparing cycles of the time series directly, one need

not be concerned with embedding and delay reconstruction. In fact, the most obvious ways to compare cycles of such candidate systems are also very robust to (observational) noise [12]. In contrast, the method proposed in [10] is applicable to any time series data, provided a suitable delay embedding can be found. However, the examples given in [10] required extremely long (around 10^4 embedding data) time series and were demonstrated (in the case of chaos) only for situations with minimal noise. In this communication we demonstrate one method to overcome these issues. First we specify the basic principle of the algorithm [10].

We assume that a successful (or at the very least, robust) embedding can be obtained from the time series data. Let $x_t \in \mathbf{R}^{d_e}$ denote the vector time series of points in the time delay reconstruction. Each point x_t maps to a corresponding node n_t in the network. The adjacency matrix $A = [a_{ij}]$ is constructed such that $a_{ij} = 1$ iff x_i is one of the K closest neighbours of x_j or vice versa. In case x_i is one of the K closest neighbours of x_j and x_j is one of the K closest neighbours of x_i , then we select another node n_k , corresponding to x_k which is the $2K$ -th closest neighbour of either x_i or x_j . In this way, we can ensure that each node contributes K links to the network, and the mean node degree is $2K$ (because the links are undirected). Of course, the distribution of links between nodes need not be uniform.

This transform from time domain, via an embedding, to the network structure introduces a new assortment of measures related to the adjacency matrix A which we can study to unravel information related to the original dynamical systems. In [10] we examined the motif frequency distribution. That is, we looked at the relative frequency of different sets of M connected nodes, for fixed M . In particular for $M = 4$ we list all the different ways in which four nodes may be connected to one another (discounting any arrangements with isolated nodes) and then compute the number of times that each arrangement occurs for a specific network. We found that for networks generated from time series which originated from chaotic flows the sequence of motif frequency is distinct from that for hyper-chaotic systems, or for periodic or noisy dynamics. In [10] we detail the varying frequency with which these patterns occur and the dynamical origin of this variability. Of course, there are many other network-based statistics which one could estimate. However, for now we limit our attention to these motif quantities.

In [10] we consider only $M = 4$ ($M = 3$ is relatively trivial, and $M > 4$ is computationally difficult) and $K = 4$ (the choice of M and K are unrelated, and actually we obtain equivalent results for a wide range of values of K). We then show that this method allows us to consider distinct motif

M Small is with the Department of Electronic and Information Engineering, Hong Kong Polytechnic University, e-mail: small@iee.org.

X. Xu is with the School of Communication and Electronics Engineering, Qingdao Technological University, Qingdao, China.

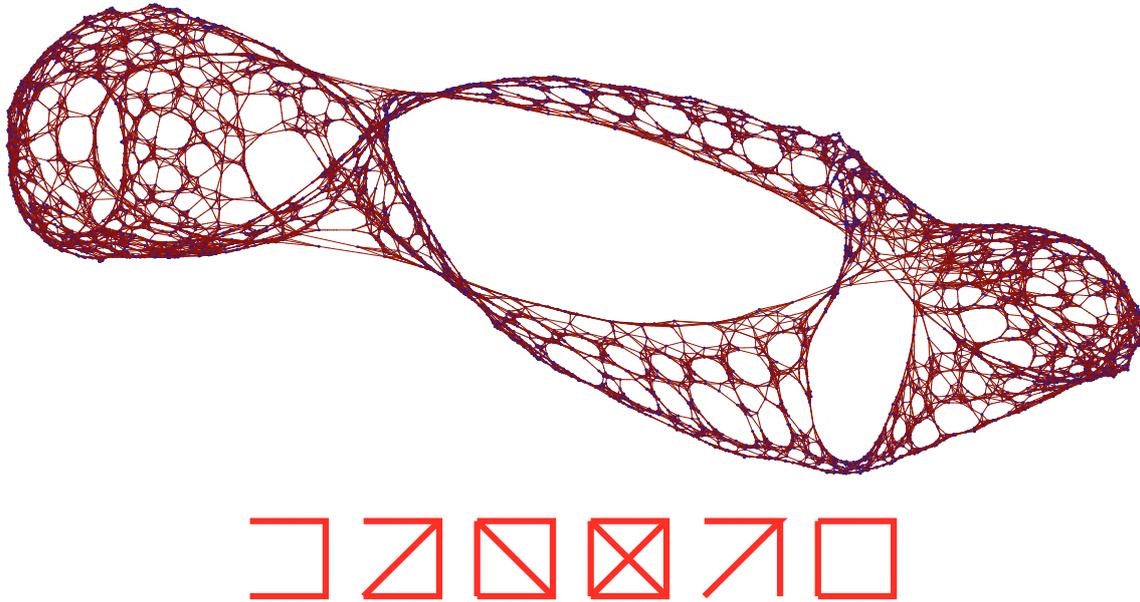


Fig. 1. The network constructed from a trajectory of the chaotic Lorenz system according to the algorithm described in Sec. II. Each node of the network corresponds to a point in the embedded system, and links are drawn between nodes if the corresponding embedded points are near neighbours. Below this we depict the motif frequency distribution (most to least common from left to right) for motif-4. This same Motif superfamily structure is observed in a wide variety of low dimensional chaotic systems: including all the textbook three dimensional flows.

superfamilies. Two networks (and hence the corresponding time series) are members of the same superfamily if different motif of size M (for example, the six different motifs of size 4 are shown in Fig. 1) occur with the same relative frequency. In Fig. 1 we illustrate the results of this method with a depiction of the network constructed from a clean trajectory of the chaotic Lorenz system, along with the motif frequency distribution typical of low-dimensional chaotic dynamics.

A weakness of this method, as presented in [10] is that the algorithm only works for very long ($O(10^4)$) time series. In the next section we illustrate the application of this technique, in conjunction with a nonlinear time series modelling routine to study the asymptotic dynamics of models built from a much shorter time series.

III. LINKS FROM LYNX

Figure 2 depicts the ecological time series data which we analyse in this communication. The data depicts the annual number of Canadian lynx pelts harvested by the Hudson Bay company between 1821 and 1935. Annual reported capture is available for each year excluding the period between 1892 and 1896. For these missing years we complete the data set following an ad hoc procedure detailed in [4]. The raw data therefore consists of 113 observations. The data has then been interpolated, following a procedure described in [4] which

preserves the power spectrum while increasing the sampling rate by a factor of ten. The resultant data was provided to us by the authors of [4] and is shown in Fig. 2.

This data is still insufficient to analyse with the network based methods of [10]. Instead, we build nonlinear models from this data and analyse the dynamics of these models. This provides a form of meta-analysis of the original data. Of course, dynamics observed in the models provide only heresay evidence of the actual dynamics in the data. But, this is to be expected. Any dynamics we infer from the data is based on some *a priori* assumptions (that is, a model) concerning the expected form of those dynamics. By building models we are only making these assumptions more explicit.

The modelling procedure is detailed in [6]. The scalar time series data x_t is subjected to a variable time delay embedding to attempt to capture the various relevant time scales in the system

$$z_t = (x_t, x_{t-5}, x_{t-10}, x_{t-15}, x_{t-20}, x_{t-25}, x_{t-90}) \quad (1)$$

where the pseudo-period of the system is around 96. The choice of embedding strategy is both ad hoc and arbitrary. Various alternative strategies produced similar results provided that the embedding lags span the same range of values. From the vector time series we employ radial basis models to approximate the function f such that $\|z_{t+1} - f(z_t)\|$ is

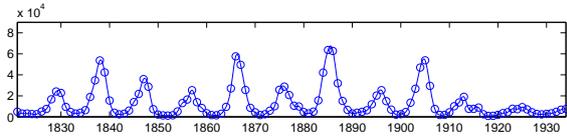


Fig. 2. Annual number of Canadian lynx pelts reported by the Hudson Bay company (1821-1935). Data for the years 1892 to 1896 are missing and filled in following an ad hoc procedure. The data set has been interpolated by a factor of 10 (increasing the amount of available data while preserving the power spectra). The original values are shown as open circles.

minimised subject to a fixed model size d

$$f(x) = \sum_{i=1}^d \lambda_i \phi_i \left(\frac{\|x - c_i\|}{r_i} \right) \quad (2)$$

where the weights λ_i are determined by least-mean-squares fit, the nonlinear parameters c_i and r_i are chosen with a combination of random search and steepest descent, and ϕ_i is one of the following functional forms:

$$\phi_i = \begin{cases} \exp(-\frac{1}{2}x^2) \\ \exp\left(-\frac{1-p}{p} \left|\frac{x}{p}\right|^p\right) \\ (2x^2 - 1) \exp(-x^2) \end{cases} \quad (3)$$

that represent Gaussian, “tophat” (a modified Gaussian) or a Mexican hat wavelet. The remaining parameter d is selected according to the minimum description length principle. The description length is the computational cost (in terms of number of bits) required to describe a particular model and the model prediction errors of that model, rather than just describing the raw data. The “best” model is deemed to be the one which affords the shortest description of the data (that is, it achieves the greatest compression as measured with the smallest value of description length).

For a fixed model size d the model which minimises the mean-squares prediction error is deemed to be best, but by selecting model size based on description length we avoid overfitting without the need for validation data. The idea behind minimum description length is detailed in [5] and the application to radial basis and neural network models which we utilise here is described in [2], [14], [8], [7]. Note that, since the optimisation problem is nonlinear, and the algorithm is stochastic, repeated application of the modelling procedure will yield a different solution.

From an ensemble of 100 models we generate trajectories of each model for the same initial conditions (taken from the original data). Representative trajectories are illustrated in Fig. 3. For each trajectory we apply the network transform and motif super-family techniques described above to identify the frequency of various motifs of size four. We find that the motif pattern provides an accurate and succinct summary of the underlying dynamics in each case.

In total, from 100 models we identified four distinct super-families corresponding to periodic trajectories. In every case the asymptotic dynamics was a dense one dimensional set (i.e. not a periodic map but points on a periodic flow sampled at a frequency incommensurate with the underlying period). We

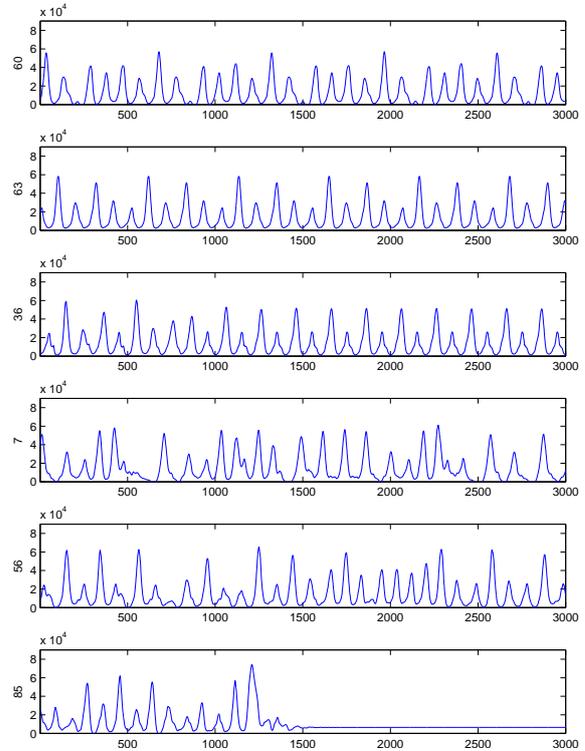


Fig. 3. Typical trajectories corresponding to different motif superfamilies. The vertical scale is the same as Fig. 2, but not the horizontal axis. In each case a transient of 1000 points has been removed. The first two trajectories (models 60 and 63) correspond to models from distinct periodic superfamilies. The third model (36) is typical of the “almost-periodic” trajectories of a third superfamily (in this case, the model exhibited intermittent “periodic” bursts). The fourth and fifth trajectories (models 7 and 56 respectively) correspond to distinct chaotic superfamilies (the most prevalent among all the models). Finally, all systems which exhibited a stable fixed point (focus or node) are captured in yet another distinct superfamily (represented here by a trajectory of model 85). The numerical identifiers on the y -axis correspond to the particular models (see discussion).

find no difference between these trajectories and attribute the distinct groupings to transient dynamics. Trajectories of model 60 and 63 are representative of two of these groups and are shown in Fig. 3. A relatively large proportion of models belong to a separate super-family, typified by apparently chaotic but almost periodic dynamics. A trajectory of model 36 is typical of these as illustrated in Fig. 3. In this case the trajectory exhibited intermittent bursts of almost periodic dynamics (of course, the system state can not be intermittently exactly periodic) and otherwise irregular bounded and aperiodic behaviour.

Trajectories of models 7 and 56 are typical examples of chaotic dynamics and are representative of two separate superfamilies. By far the most commonly occurring superfamily is represented by model 7 and is the same as that identified in Fig. 1: low dimensional chaos. Sample attractors reconstructed from the trajectories in Fig. 3 are illustrated in Fig. 4. In Fig. 5 we contrast the various dynamics of each motif superfamily

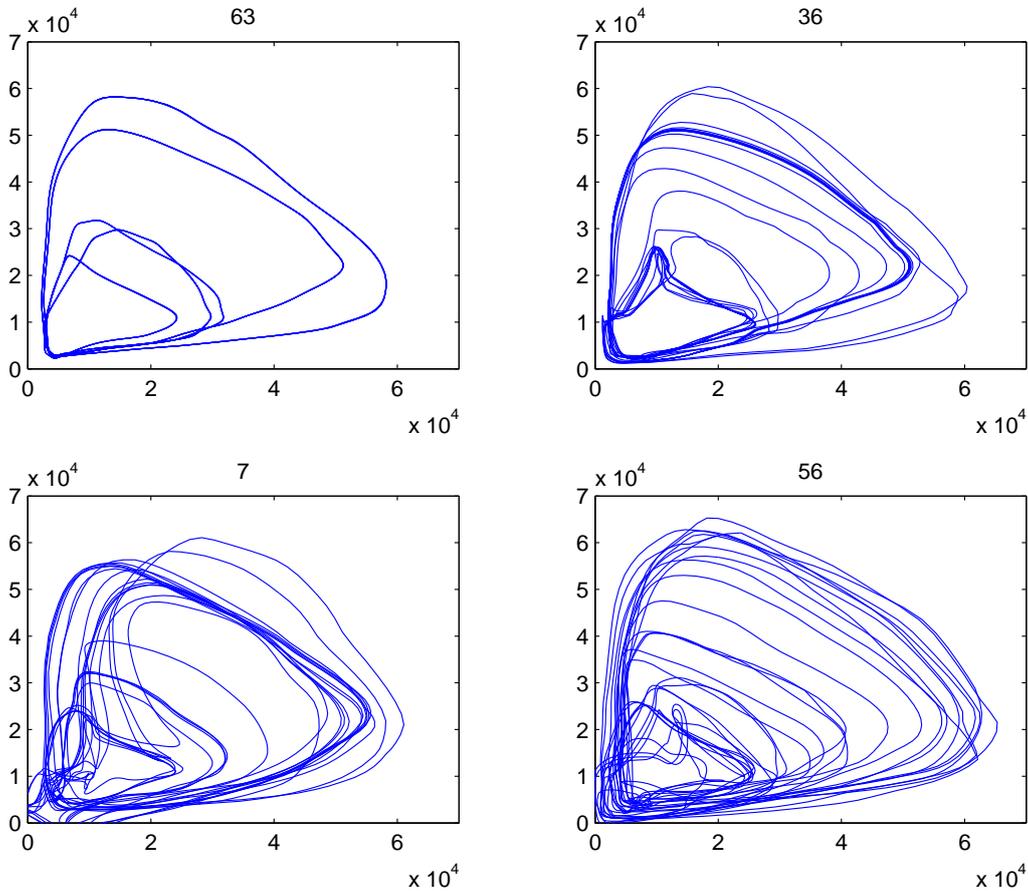


Fig. 4. Time delay embedding ($d_e = 2$, $\tau = 20$) of four of the trajectories shown in Fig. 3. Trajectory 63 (top left) is periodic, the other three (36, 7, and 56) are increasingly disordered and correspond to different motif superfamilies. The third model (36) (of Fig. 3) is typical of the “almost-periodic” trajectories of a third superfamily (in this case, the model exhibited intermittent “periodic” bursts). The fourth and fifth modes (of Fig. 3, numbered 7 and 56 respectively) correspond to distinct chaotic superfamilies (the most prevalent among all the models). In each plot the horizontal and vertical axes are equal, and each trajectory comes from a completely independent model of the data in Fig. III.

by computing the correlation dimension using the Gaussian kernel algorithm described by Diks [1], [11]. The distinction between the periodic asymptotic behaviour of the various superfamilies, and the corresponding distinction between the different chaotic superfamilies is not clear for estimates of correlation dimension. Nonetheless, the distinction between the various broader dynamical categories is clear.

We note that the motif family groupings are extremely good at classifying the dynamics correctly based on the asymptotic behaviour. Even in cases with a very long and atypical transient the dynamics are correctly identified. For example, in model 85 (See Fig. 3) the system exhibits apparently chaotic behaviour for over 10^4 time steps before converging to a fixed point. Although the first 1.3×10^4 points appear chaotic (i.e. they are deterministic, bounded and aperiodic), this does not taint the motif superfamily classification. Of course, for both periodic and chaotic dynamics we do observe more than one superfamily, while it is not clear at this stage which family is most appropriate.

IV. CONCLUSION

The network construction method introduced in [10] provides a new way to look at time series data. The method relies on one being able to reliably reconstruct the system attractor with a time delay embedding, or some other technique. Nonetheless, from the network structure one immediately has an ensemble of network analysis tools which effectively provide a new set of statistics for the analysis of complex networks. In this paper, and in [10] we have only considered the prevalence of various motifs of size 4. The choice of this particular statistic was deliberate as it gives a tool to analyse the complex network (and therefore the original time series) which has no clear analogue in nonlinear time series analysis.

Of course, the adjacency matrix of the complex network is in some ways similar to the recurrence matrix produced for recurrence plots. One can therefore consider the tools of complex network analysis as a new set of statistical measures to be applied to time series, in just the same way as recurrence

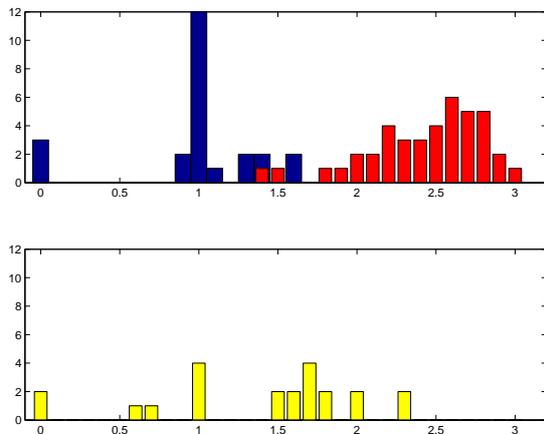


Fig. 5. Estimates of correlation dimension (results for $d_e = 10$ and $\tau = 2$ are shown, but are representative) for periodic superfamily trajectories (blue, upper panel), chaotic superfamilies (red, upper panel) and the “almost periodic” superfamily as typified by the trajectory of model 36 of Fig. 3 (yellow, lower panel). Similar results were obtained for calculations of correlation dimension with other methods, and from calculations of entropy and complexity.

plots allowed for the introduction of recurrence quantification analysis. However, unlike recurrence quantification analysis, the most commonly applied tools from complex network analysis (measures such as path length, network diameter, degree distribution, betweenness, centrality and so on) are dependent on higher order properties of the adjacency matrix, rather than geometric features of the recurrence matrix (such as the mean length of diagonal and vertical components of the matrix). In the case of recurrence matrix, statistics are computed from the geometric structure of that matrix. For complex networks, it is iterates (or other more complicated functions) of the adjacency matrix which are important.

Although we have yet to fully explore the potential of these network analysis tools, this paper has applied complex network transform of [10] to analyse the motif superfamily behaviour in an ensemble of models of an apparently (or at least “plausibly”) chaotic ecosystem. In addition to observing models which exhibit a wide range of interesting dynamic behaviour (which is discussed in more detail in [6]) we show that the models are categorised into one of four basic dynamical behaviours. By comparing the behaviours of these various models, and the corresponding motif superfamilies to data using surrogate type techniques [9] we can now provide an analysis which allows us to determine which dynamical description is most suitable for this data. We can test whether the chaotic dynamics of the models which is apparently plausible is also statistically likely.

ACKNOWLEDGMENT

The authors would like to thank Christophe Letellier for sharing the version of the lynx time series data reported in their paper. We also acknowledge Jie Zhang and Cristian Carmeli for various fruitful and useful discussions. The paper

is supported with funding from the Hong Kong Polytechnic University under grant number G-YG35.

REFERENCES

- [1] C. Diks. Estimating invariants of noisy attractors. *Physical Review E*, 53:R4263–R4266, 1996.
- [2] Kevin Judd and Alistair Mees. On selecting models for nonlinear time series. *Physica D*, 82:426–444, 1995.
- [3] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J.C. Nuño. From time series to complex networks: The visibility graph. *Proc Natl Acad Sci USA*, 105(13):4972–4975, 2008.
- [4] J. Maquet, C. Letellier, and Luis A. Aguirre. Global models from the canadian lynx cycles as a direct evidence for chaos in real ecosystems. *J. Math. Biol.*, 55:21–39, 2007.
- [5] Jorma Rissanen. *Stochastic complexity in statistical inquiry*. World Scientific, Singapore, 1989.
- [6] Michael Small and Cristian Carmeli. Re-examination of evidence for low-dimensional chaos in the canadian lynx data. In *International Symposium on Nonlinear Theory and its Applications*. Research Society of Nonlinear Theory and its Applications, IEICE, 2009.
- [7] Michael Small and Kevin Judd. Comparison of new nonlinear modelling techniques with applications to infant respiration. *Physica D*, 117:283–298, 1998.
- [8] Michael Small and C.K. Tse. Minimum description length neural networks for time series prediction. *Physical Review E*, 66:066701, 2002.
- [9] Michael Small, Dejin Yu, and Robert G. Harrison. Surrogate test for pseudo-periodic time series data. *Physical Review Letters*, 87:188101, 2001.
- [10] Xiaoke Xu, Jie Zhang, and Michael Small. Superfamily phenomena and motifs of networks induced from time series. *Proc Natl Acad Sci USA*, 105:19601–19605, 2008.
- [11] Dejin Yu, Michael Small, Robert G. Harrison, and C. Diks. Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data. *Physical Review E*, 61:3750–3756, 2000.
- [12] Jie Zhang, Xiaodong Luo, and Michael Small. Detecting chaos in pseudoperiodic time series without embedding. *Physical Review E*, 73:016216, 2006.
- [13] Jie Zhang and Michael Small. Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters*, 96:238701, 2006.
- [14] Yi Zhao and Michael Small. Minimum description length criterion for modeling of chaotic attractors with multilayer perceptron networks. *IEEE Transactions on Circuits and Systems I*, 53:722–732, 2006.

Michael Small (M’01) Michael received the BSc(Hons) degree in pure mathematics and the PhD degree in applied mathematics from the University of Western Australia, in 1994 and 1998, respectively. He is presently an Associate Professor with the Department of Electronic and Information Engineering of the Hong Kong Polytechnic University, Hong Kong. Previously, he has worked in the Physics Department of Heriot-Watt University, Edinburgh, United Kingdom and the Mathematics Department of the University of Western Australia. His research interests are nonlinear dynamics and chaos, nonlinear time series analysis, and nonlinear modelling. As well as modelling disease epidemics, his recent work has been on the application of nonlinear mathematical methods to a diverse range of problems, including, neurodynamics, infant respiratory patterns, cardiac arrhythmia, pulse diagnosis and financial time series analysis.

Xiaoke Xu Xiaoke Xu received the BSc(Hons) degree in electronic and information engineering and the PhD degree in communication and information systems from Dalian Maritime University, Dalian, China, in 2002 and 2008, respectively. From 2008 to 2009, he was with the Department of Electronic and Information Engineering, Hong Kong Polytechnic University, where he studied nonlinear time series analysis and complex networks. He is currently with the School of Communication and Electronics Engineering, Qingdao Technological University, Qingdao, China. His general research interests are in the areas of complex networks, nonlinear time series analysis, nonlinear dynamics and radar signal processing.

Non-Gaussian signal processing of biological signals

Max LITTLE

Digital signal processing is dominated by concepts from classical linear systems theory first systematically described in the early 20th century. Despite the enormous success of digital signal processing based on these concepts in practical problems, there are many systems and signals for which the linear-Gaussian theory is inappropriate, in particular signals arising from natural contexts such as in biology and biomedicine. Here, we are faced with signals that show unexpectedly common large fluctuations, multimodality, long memory, and discrete transitions. For such signals, non-Gaussian techniques show improvements such as increased precision, sparseness and selectivity. In this talk, I will describe some of the ways in which non-Gaussian signal processing has advanced in recent years, and demonstrate some example applications which have particularly benefited from these advances.

Dynamic Graphs and Complex Systems

Organizing Committee

- Serge CHAUMETTE
LaBRI, University of Bordeaux 1, France
`serge.chaumette@labri.fr`
- Frédéric GUINAND
University of Le Havre , France
`frederic.guinand@univ-lehavre.fr`

Description

The goal of this special session is to bring together researchers, whatever their original scientific field, working on dynamic graphs for modelling, simulating and/or analysing complex systems. Topics of interests include (but are not limited to):

- dynamic graph models in biology
- dynamic graph models in computer networks
- dynamic graph models in social sciences
- metrics in dynamic graphs
- probabilistic analysis of dynamic graphs
- morphogenetic analysis of dynamic graphs
- time management in dynamic graphs

Contents

A survey of connectivity in mobile ad hoc networks	
Cédric ABOUE NZE , Frédéric GUINAND	357
On the use of social agents for image segmentation	
Richard MOUSSA , Marie BEURTON-AIMAR , Pascal DESBARATS	369
Dynamical systems based on dynamic graphs	
Stéphane DESPRÉAUX , Aude MAIGNAN	376
Dynamic graph as a new type of graphs	
Mounira NEKRI , Abdelkader KHELLADI	379

A Survey of Connectivity in Mobile Ad Hoc Networks

Cédric Aboue Nze, Frédéric Guinand

2009

1 Introduction

Mobile Ad hoc networks are networks without centralized infrastructures in which stations can move permanently in a given environment. One of the main issues in this domain is evaluation of communication algorithms which are used there. An approach to study this problem consists in studying communication graphs induced by mobile ad hoc networks. In these graphs, stations become nodes and when two stations are able to communicate, an edge is automatically created between the two corresponding nodes. Then, the study of the communication of the ad hoc networks is limited to the study of the connectivity of their connection graphs. Most studies on connectivity of ad hoc networks go in this direction. A graph is connected if there exists at least a way connecting any pair of nodes in this graph. Connectivity is the estimate of this connexity. Several studies on the connectivity of mobile ad hoc networks have been undertaken for a few years. These studies are mainly divided into two categories.

The first category concerned the study of the K -connectivity of ad hoc networks. According to the theorem of menger, a graph is k -connected if and only if, for any pair of nodes u, v , there are k internal ways of disjointed nodes connecting u to v . In other words, a graph is K -connected if this graph remains connected after the removal of any subset of $k - 1$ nodes. It is possible to not consider nodes but edges of graph. In this case, a graph is K -connected by edges if this one remains connected after suppression of any subset of $k - 1$ edges of the graph. K -connectivity gives an additional information about the connexity of the graph in particular the mean degree of node. When $k = 1$ the graph is quite simply connected.

The second category of study is the evaluation of the critical transmission range of the signal of stations allowing to guarantee the connexity of the ad hoc network. These studies are used in sensors networks in order to optimize the energy expenditure of the sensors whose the emitted signal power is proportional to the length of the ray of coverage of the signal.

We present in this article a survey of the various studies undertaken on these three fields whose objective is to estimate the connectivity of the ad hoc networks.

2 Preliminaries

To study the connectivity in mobile ad hoc networks, we characterize each ad hoc network according to three parameters : environment, signal transmission ray of stations and stations mobility.

The environment is the surface where stations are located. They can be various types according to the size, the form, the dimension of space considered. we classify the environment according to five criteria : the first criterion is the form (square, circle). The second criterion concern the borders quality which can be closed (any station cannot come in or goes out the considered zone) or opened (the stations can move in or move out the considered zone). The third criterion specifies if the environment is limited (i.e. if the considered zone is a torus) or not. The fourth criterion is the size of the environment. It is limited if it is not infinite and it is not limited in the contrary case. The last criterion is the obstacles. We will specify if the environment contains or not obstacles. An obstacle is a zone of the environment where stations cannot access.

The model of the signal transmission of stations is a very important parameter in the study of the connectivity of connections graphs because results obtained depend directly on the model chosen. Two models were mainly modelled by Hekmat and P. Van Mieghem in [10] in particular the pathloss model and the lognormal model. The Pathloss model simulates the behavior of the signal on large scales. It describes the existing dependence between signal power received by a receiver and distance which separates it of the transmitter. The coverage of the signal of a node is represented by a circle. The lognormal model (more realistic) captures the random variations of the radio signal on average distances to various positions. It presents the fact that the power of the radio signal received at a fixed distance from the transmitter varies considerably according to the positions around the transmitter. The coverage of the signal is deformed. The deformation of the signal can be related to interferences introduced into the model. Connectivity in this last case is evaluated in [10] [11] [15]. The majority of works study the pathloss model because it is easier to model mathematically. We will thus consider only studies using this model and we will specify if the coverage ray of the stations is homogeneous (the same one for all) or heterogeneous and if it signal crosses or not obstacles.

There exist two ways to consider the mobility of stations in order to evaluate the connectivity of the connexions graphs of modile ad hoc networks. In the first, mobility is considered implicitly. The mobility is represented as a change of topology of the graph at every moment what can involve modifications of

connections between nodes. In this case the graph is static and the localization of stations in the environment is obtained by the distribution of density of probability in the environment [3] [2]. The second way to consider the mobility of stations is explicit. When stations move in the environment, the existing edges between each pairs of nodes can disappear and other edges can appear. Then time is important for the evaluation of connectivity in particular by considering periods of connexity and nonconnexity of the graph. Several models of mobility of stations were developed these last years [1] [12]. More studied being the model random waypoint. In this model the stations choose randomly a destination in the environment. They move in straight line of their position towards this destination. Arrived at destination the stations make a pause and the cycle starts again. A large majority of the studies are relate to the first case of figure.

We will specify for each study parameters used to evaluate the connectivity of the considered ad hoc networks

3 Study of Connectivity in Mobile Ad Hoc Networks

The studies on connectivity and more generally of K-connectivity on mobile ad hoc networks are especially related to the networks which mobility is supposed to be implicit. The difficulty for explicitly dynamic ad hoc networks (influence of time in the connectivity computation) resides in the fact that to study them they should be simulated before. Consequently the results obtained can to be skewed by parameters of the simulation. The approach commonly used consists in studying the connectivity of the static graphs with a discrete time which corresponds to a sequence of consecutive snapshots studied separately. This approach is divided into two main categories : on the one hand the study of the asymptotic connectivity of connection graph i.e. when the number of nodes in the environment tends to the infinity and on the other hand the study of the connectivity in finite mobile ad hoc networks (when the number of nodes in the environment is weak).

3.1 Connectivity in Asymptotic Ad Hoc Networks

3.1.1 Connectivity of a paire of nodes

In this study, an environment closed, bounded and limited is considered. The signal coverage is circular (pathloss model) and the network is homogeneous (the length of signal coverage ray is the same for all the nodes). The form of the environment can be a square (or a rectangle) or a circle in order to facilitate computing. As the network is static, a knowledge of the distribution stations in the environment must be known, i.e. the spatial node distribution in the environment. There are many spatial node distributions (uniform, Poisson, Gauss, etc). The most used is the uniform distribution where any node has the

same probability to be at any position of the environment. From the spatial node distribution, we can deduce the probability density function (pdf) for a given environment. This function indicates the probability for a station to be at a precise position in the environment. Thus, at every position of the environment we have thus a probability of presence of stations. Christian Bettstetter et. al. have determined the probability density function of a homogeneous network whose stations move according to the model Random Waypoint [6] [5] and within the framework of a heterogeneous network [4]. Esa Hyytiä et. Al. have widened the field on other forms of environments (circle, polygon and triangle) in [13].

To evaluate connectivity, the probability density function is known in advance [3] [2]. Once the density probability function known, one determines the probability that two stations randomly located are connected. Connectivity between a station a and a station b is only possible if the station b is in the signal coverage area of the station a . Thus if the distance between a and b is lower than the ray of the signal transmission, these two stations are connected [2]. The probability that the station a is connected to any other station in the environment equal to the number of stations located in the coverage area of a (the whole of the positions of its coverage area) on the full number of stations of the environment (the whole of the positions of the environment). That corresponds to an integration of the probability density function of the station coverage a .

Let f the probability density function of stations in the environment, the probability of connectivity of a station with coverage a_0 at the position x in the environment is :

$$p_0(x) = \iint_{a_0(x)} f(x)dx$$

There are two manners to apprehend the computation of the connectivity of a static ad hoc network. Firstly, from a local point of view, the connectivity is evaluated on each position of the environment. That corresponds to a distribution of connectivity in each position of the environment. Then by an integration on the whole of environment, a global value of connectivity is obtained. Secondly, the ad hoc network can be seen as a total entity. The probability to have a isolated node is computed in the ad hoc network. An isolated node is a node which is connected to none of its neighbors. The connectivity is evaluated by computing the probability that the connection graph does not contain an isolated node.

3.1.2 Ad Hoc Networks Local Connectivity

From the local point of view, it is supposed that there are n nodes in the environment and these nodes are independent to each other. As a node can be connected or not to its neighbor, connectivity between two stations can be interpreted like a binary random variable (0 or 1). In this case the probability

that the node degree at every position x of the environment (d_x) is equal to k (k is a constant) follows a binomial probability law [3] :

$$p(d_x = k) = \binom{n-1}{k} (p_0(x))^k (1-p_0(x))^{(n-1-k)}$$

The local mean degree $\mu_0(X)$ is deduced by computing the degree expectation at each position x of the environment

$$\mu_0(x) = E(d_x) = (n-1).p_0(x)$$

On the large environments, when p_0 is very small, the binomial distribution can be approximated by the Poisson distribution. Pasi Lassila et. Al. use Binomial and Poisson laws to estimate connectivity in a circular environment, closed, bounded, limited of a homogeneous network [14]. they are obtained :

$$p(d_x = k) \approx \frac{\mu_0(x)^k}{k!} e^{-\mu_0(x)}$$

The probability for the node to have in more k neighbors is given by :

$$p(d_x \leq k) \approx \sum_{d_x=0}^k \frac{\mu_0(x)^{d_x}}{d_x!} e^{-\mu_0(x)}$$

It is possible to determine the distribution of K-connectivity in the environment. The probability that the node has at least k neighbors at the position x is :

$$p(d_x \geq k) = 1 - p(d_x \leq k-1)$$

Finally the estimate of K-connectivity (when the connection graph has at least k neighbors) on the whole of environment is :

$$p(d \geq k) = \iint p(d_x \geq k) f(x) dx$$

3.1.3 Ad Hoc Networks Total Connectivity

From the total point of view, the ad hoc network is perceived like a single component. The connectivity of the graph depends at the same time on the coverage ray r_0 of stations and the number of stations n in the environment. The study of connectivity (contrary to the local point of view) consists in evaluating the connectivity of the network according to these two parameters. Let G the connection graph resulting from the ad hoc network, the probability that there is no node isolated in the graph is estimated in order to approach the probability of connectivity of the graph G (close to 1). The probability that a node is isolated at the position x by using a Poisson distribution is :

$$p(i_x) = p(d_x = 0) \approx e^{-\mu_0(x)}$$

The probability to have a isolated node on the environment is :

$$p(i) = \iint p(i_x) f(x) dx$$

The probability that there are no isolated nodes in the graph G is not a sufficient condition to have this graph connected :

$$p(G_{connece}) \geq p(i)$$

The probability that a node is isolated among n other nodes is :

$$p(\neg i) = (1 - p(i))^n$$

By approximating with the Poisson law we obtain :

$$p(\neg i) = e^{-n \iint e^{-\mu_0(x)} f(x) dx}$$

The estimate of K-connectivity (when the graph of connection has at least k neighbors) on the whole of environment is :

$$p(d_{min} \geq k) \approx (p(d \geq k))^n \approx e^{-np(d \leq k-1)}$$

3.2 Connectivity in Finite Ad Hoc Networks

Results obtained of the asymptotic study of the connectivity of the ad hoc networks are not reliable when the number of nodes in the environment is finite and weak. An other method is to study the connectivity of the ad hoc networks empirically. In the Monte Carlo method in [19], n stations are randomly distributed with a density D in an square, closed, bounded and limited environment. The network is homogeneous. By using the Dijkstra algorithm, it is determined if the network is entirely connected or not . The process is repeated M time including the number of times m that the network is connected. It is deduced the probability of connectivity of the ad hoc network over the whole of the experiment period :

$$p(G_{connece}) = \frac{m}{M}$$

In the same idea, Bettstetter approximates connectivity by the path probability computation in [3]. Time is discrete. It is computed the probability p_{path} that two stations chosen randomly in the connection graph G_i are connected.

$$p_{path}(G_i) = \frac{\text{number of pair of connected nodes}}{\text{number of pair of possible nodes}}$$

If $p_{path}(G_i) = 1$ then the graph G_i is complete and connected and if $p_{path}(G_i) = 0$ then all nodes of the graph G_i are insolated. The path probability of a graph G over one discretized period ω is :

$$p_{path}(G) = \lim_{\omega \rightarrow \infty} \frac{1}{\omega} \sum_{i=1}^{\omega} p_{path}(G_i)$$

Bettstetter proposes an approximation of the connectivity of the graph $G_{connected}$ compared to the path probability below :

$$p_{path}(G) \geq p(G_{connece})$$

Ao Tang and Al propose an empirical formula in [19] to estimate the probability of connectivity of an ad hoc network whose environment is homogeneous, closed, bounded and limited. They are interested to know if the number of stations is finished (no more 125). Their formula is reliable only if the probability of connexity between two stations is closed to zero or one. The probability for the graph to be connected is given below.

$$p(G_{connece}) \approx \frac{e^{R-Rc/E}}{1 + e^{R-Rc/E}}$$

R is the stations coverage ray, Rc and E are parameters of the model and depend on the length of the environment border.

$$\begin{cases} Rc \approx (1.0362\sqrt{\frac{\ln(n)}{n}} - 0.073)L \\ E \approx (\frac{0.3743n-0.3331}{n \cdot \ln^2(n)})L \end{cases}$$

they propose an upper limit of the probability of connexity of two stations randomly choosen in a square, closed, limited and homogeneous ad hoc network with the stations uniform distribution. If L is the border length of the environment., n is the number of stations, r is the signal coverage ray of stations and $sigma = l/3$ then :

$$p \leq 1 - e^{-n^2 r^2 / 4\sigma^2}$$

In the same way, Madhav Desai and D. Manjunath in [8] propose an approximation of the limit upper of the probability of connectivity of an entire ad hoc network where the environment is in two dimensions closed, limited, limited uniform distribution according to the ray of cover of the signal r , length of with dimensions of the environment and amongst stations in environment the n .

$$p(G_{connece}) \leq \left(\sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{(z-kr)^n}{z^n} u(z-kr) \right)^2$$

When the network is not entirely connected, Ao Tang et. Al. propose an index of connectivity which estimate the connectivity according to the number of connected component in the environment eta and the number of nodes in each component n_i .

$$\eta = \frac{\sum_i n_i (n_i - 1)}{\sum_i n_i (\sum_i n_i - 1)}$$

The stations mobility can be taken into account by considering the index of connectivity at each discretized time interval η_i . Then, the probability of connectivity between two stations becomes :

$$p = 1 - \prod_{i=1}^k (1 - \eta_i)$$

Wang Hanxing, Liu Guilin and Zhao Wei compute analytically the connectivity of an ad hoc networks uniformly distributed in an environment closed, bounded, limited and homogeneous according to a F-node of a connected graph [20]. A f-node is a node of a connected graph whose the deletion returns the graph disconnexe and divides it into several components. Wang Hanxing et. Al. compute the probability of connexity of two stations in this graph according to the number of connected component and show that this graph is composed of more five components.

4 Study of Critical Transmission Range for connectivity

The second way studied is the critical transmission range (critical coverage ray) necessary to the stations to connect the whole network. Two options are possible : the case where the ad hoc network coverage is homogeneous and the case where it is heterogeneous. However the way currently studied concern the homogeneous ad hoc networks.

4.1 Critical Transmission Range of asymptotic connected Networks

P. Gupta and P.R. Kumar [9] have determined a sufficient condition on the coverage ray of stations r to connect the whole ad hoc network when the number of nodes tends to the infinity. The environment is normalized, limited, closed, bounded and without obstacles. The stations are uniformly and independently distributed in the environment. Two stations can communicate only if the distance separating them is lower or equal to the coverage ray r . If $\pi r^2(N) = \frac{\log n + c(N)}{N}$ then the graph $G(n, r(n))$ is asymptotically connected with a probability 1 when n tends to the infinity if and only if $c(n) \rightarrow \infty$. Mobility is not taken into account however it can be perceived as a change of topology of the ad hoc network in which computations are repeated when the topology change.

Paolo Santi and Douglas Mr. Blough have extended the result for square environment in 1, 2 or 3 dimensions [18]. They suppose that if n nodes uniformly and independently distributed are in an area $R = [0, l]^d$ with $d = 2$ or $d = 3$, if $r^d n = k l^d \ln(l)$ for any constant $k > 0$, r very small compared to l , n very

large compared to 1, then if $k > d.k_d$ or $k = d.k_d$ with r very large compared to 1 ($k_d = 2^d d^{D/2}$), then the communication graph is connected when n tends to infinity the

According to Penrose et. al. in [16], the value of common minimal transmission such as the communication graph is connected is equivalent to the longest edge length of an Euclidean minimal spanning tree. Paolo Santi has established that this longest edge of a minimal spanning tree depends only on the minimal value of the probability density function [17]. If M is a model of mobility of nodes in a square environment standardized $R = [0,1]^2$, closed limited and without obstacles. By supposing that the probability density function of nodes in the environment f_m is continuous on borders and the minimum of f_m on R is higher than zero, then when the constant $c \geq 1$ we have :

$$\lim_{n \rightarrow \infty} r_M = c \sqrt{\frac{\ln n}{\pi n}}$$

In the preceding formula, the problems consist in evaluating the constant c when n tends to the infinity. However it is possible to compute probability density function of the nodes moving according to random waypoint mobility model within an environment closed, bounded, limited and without obstacles [17]. The critical range of transmission depends on pause time p of the stations and their travel velocity ($v_{min} = v_{max} = v$). The minimal value of the probability density function is reached on the borders of the environment. It is equal to the pause probability of $P_p = \frac{p}{p + \frac{0.521405}{v}}$ and the critical coverage ray of stations is :

$$\lim_{n \rightarrow \infty} r_p^w = \left(\frac{p + \frac{0.521405}{v}}{p} \right) \sqrt{\frac{\ln n}{\pi n}} \quad si \ p > 0$$

Evaluation of the probability of pause of stations moving according to the Random Waypoint mobility model was studied in [5] :

$$P_{pause} = \frac{E[t_{pause}]}{E[t_{pause}] + E[t_{displacement}]}$$

Guanghui Zhang et Al. compute the probability of pause p_p according to the pause time t_{pause} , maximum speed v_{max} and minimal speed v_{min} in order to measure the effects of the mobility on the critical transmission range for connectivity of ad hoc networks [21].

$$p_{pause} = \frac{t_{pause}(v_{max} - v_{min})}{t_{pause}(v_{max} - v_{min}) + 0.521(\ln v_{max} - \ln v_{min})}$$

4.2 Critical transmission range of Finite Ad Hoc Networks

A fundamental result in the study of critical signal coverage of nodes in finite ad hoc networks was presented by Penrose and Al in [16]. Indeed, the common

minimal transmission value such as the communication graph is connected is equivalent to the length of the longest edge of Euclidean minimal spanning tree. Miguel Sanchez et al. in [7] use this idea in order to compute the critical transmission range of ad hoc networks by considering different types of mobility models (Random Waypoint, Random Gauss-Markov, Random direction model). Their results obtained show that there does not exist strong dependence between the mobility model and the critical coverage ray of the station.

5 Conclusion

We have presented in this article different studies undertaken these last years on connectivity for stations in the ad hoc mobile networks. Two fields were mainly studied. the probability computation of connectivity of a graph and the computation of critical transmission range of the signal of stations. The propagation model of the signal is supposed to be circular, however recent studies propose a model of the signal propagation which integrates the interferences (model lognormal). This model remains difficult to formalize mathematically as well as the interferences of environment. Another field which was not studied in this article concern the heterogeneous ad hoc networks. However, certain articles tackled the subject. Although most these studies concern the static ad hoc networks, the ad hoc networks integrating the stations mobility explicitly start to show interest. Connectivity is very dependant of environment. however, the majority of the studies presented in this article does not integrate the obstacles in the environment (except borders).

Références

- [1] Fan Bai and Ahmed Helmy. Chapter 1 a survey of mobility models in wireless adhoc networks.
- [2] Christian Bettstetter. On the minimum node degree and connectivity of a wireless multihop network. In *MobiHoc*, pages 80–91. ACM, 2002.
- [3] Christian Bettstetter. On the connectivity of ad hoc networks. *Comput. J.*, 47(4) :432–447, 2004.
- [4] Christian Bettstetter, Michael Gyarmati, and Udo Schilcher. An inhomogeneous spatial node distribution and its stochastic properties. In Carla-Fabiana Chiasserini, Nael B. Abu-Ghazaleh, and Sotiris E. Nikolettseas, editors, *MSWiM*, pages 400–404. ACM, 2007.
- [5] Christian Bettstetter, Hannes Hartenstein, and Xavier Pérez Costa. Stochastic properties of the random waypoint mobility model. *Wireless Networks*, 10(5) :555–567, 2004.
- [6] Christian Bettstetter and Christian Wagner. The spatial node distribution of the random waypoint mobility model. In Michael Weber and Frank Kargl, editors, *WMAN*, volume 11 of *LNI*, pages 41–58. GI, 2002.
- [7] Pietro Manzoni Departamento, Miguel Sánchez, Pietro Manzoni, and Zygmunt J. Haas. Determination of critical transmission range in ad-hoc networks miguel snchez. In *in Proceedings of Multiaccess Mobility and Teletraffic for Wireless Communications 1999 Workshop (MMT ?99*, pages 6–8, 1999.
- [8] Madhav Desai and D. Manjunath. On the connectivity in finite ad hoc networks. 2002.
- [9] Piyush Gupta and P.R. Kuma. Critical power for asymptotic connectivity, 1998.
- [10] Ramin Hekmat. *Ad-hoc Networks : Fundamental Properties and Network Topologies*. Springer, Berlin, Germany, 2006.
- [11] Ramin Hekmat and Piet Van Mieghem. Connectivity in wireless ad-hoc networks with a log-normal radio model. *MONET*, 11(3) :351–360, 2006.
- [12] Tracy Camp Jeff, Jeff Boleng, and Vanessa Davies. A survey of mobility models for ad hoc network research. 2 :483–502, 2002.
- [13] Pasi Lassila. Spatial node distribution of the random waypoint mobility model with applications. *IEEE Transactions on Mobile Computing*, 5(6) :680–694, 2006. Member-Hyytiä, Esa and Member-Virtamo, Jorma.
- [14] Pasi Lassila, Esa Hyytiä, and Henri Koskinen. Connectivity properties of random waypoint mobility model for ad hoc networks. In *in Proc. MedHoc-Net, Île de Porquerolles*, pages 159–168, 2005.
- [15] Daniele Miorandi and Eitan Altman. Coverage and connectivity of ad hoc networks presence of channel randomness. In *INFOCOM*, pages 491–502. IEEE, 2005.

- [16] Mathew D. Penrose. The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.*, 7(2) :340–361, 1997.
- [17] Paolo Santi. The critical transmitting range for connectivity in mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 4 :310–317, 2005.
- [18] Paolo Santi and Douglas M. Blough. Critical transmitting range for connectivity in sparse wireless ad hoc networks, 2003.
- [19] Ao Tang, Cédric Florens, and Steven. H. Low. An empirical study on the connectivity of ad hoc networks. page 6, 2003.
- [20] Hanxing Wang, GuiLin Lu, Weijia Jia, and Wei Zhao. Connectivity in finite ad-hoc networks. *Science in China Series F : Information Sciences*, 51(4) :417–424, 2008.
- [21] Guanghui Zhang, Jiandong Li, Yanhui Chen, and Jing Liu. Effect of mobility on the critical transmitting range for connectivity in wireless ad hoc networks. In *AINA*, pages 9–12. IEEE Computer Society, 2005.

On the use of social agents for image segmentation

Richard Moussa, Marie Beurton-Aimar, and Pascal Desbarats

Abstract—In the literature, there are a lot of methods for image segmentation. Unfortunately, they are often limited in their capacity to treat image obtained by an acquisition system (Optical, X-Ray, IRM, . . .). Thus, many of them are dedicated to particular solutions and there is no generic method for solving the image segmentation problem. In this paper, we present a way to implement segmentation methods which use models coming from biology: social spiders and social ants which are implemented by a multi-agent system. After a presentation of the principles of these two methods, we will quickly present two another ones: Region Growing and Otsu thresholding methods, in the aim to compare their results. The simulation of these methods shows results that are promising. Some perspectives have been retained in order to overcome agent-based methods for having a robust segmentation technique.

Index Terms—Image segmentation, Social spiders, Social ants, Multi-agent system, Artificial life.

I. INTRODUCTION

IMAGE SEGMENTATION consists on partitioning an image into a set of regions that covers it. After this process, each pixel is affected to a region and each region corresponds to a part of the image. The discontinuity between the regions constructs the contour of the object. The segmentation approaches can be divided into three major classes [6]. The first one corresponds to pixel-based methods which only use the gray values of the individual pixels. The second one is the edge-based methods which detect edges, for example, this can be done by computing a luminacy function. The last one, the region-based methods which analyze the gray values in larger areas for detecting regions having homogeneous characteristics, criteria or similitude. Finally, The common limitation of all these approaches is that they are based only on local information. Sometimes, only a part of the information is necessary. Pixel-based techniques do not consider the local neighborhood. Edge-based techniques look only for discontinuities, while region-based techniques only analyze homogeneous regions. Robust, automatic image segmentation requires the incorporation and efficient utilization of global contextual knowledge. However, the variability of the background, the versatility of the properties of the regions to be extracted and the presence of noise make it difficult to accomplish this task. Considering this complexity, one often applies different methods during the segmentation process according to the specificities of the images.

A MAS¹ is composed of heterogeneous unembodied agents carrying out explicitly assigned tasks, and communicating via symbols. On the contrary, many extremely competent natural

collective systems of multiple agents (e.g. social spiders and social ants) are not knowledge based, and are predominantly homogeneous and embodied; agents have no explicit task assignment, and do not communicate symbolically. A common method of control used in such collective systems is stigmergy, the production of a certain behavior in agents as a consequence of the effects produced in the local environment by previous behavior [11].

Social insects like ants are one of the most diverse and ecologically important organisms on earth. As superorganisms, they live in intricately governed societies that rival our own in complexity and internal cohesion. For example, they are particularly well suited to post-genome biology age because they can be studied at multiple different levels of biological organization, from gene to ecosystem, and much is known about their natural history [19].

Social spiders belong to spider species whose individuals form relatively long-lasting aggregations. Whereas most spiders are solitary and even aggressive toward conspecifics, hundreds of species show a tendency to live in groups and to develop collaborations between each other, often referred to as colonies. For example, spiders of 5mm in body length are capable to fix silks up to a volume of $100m^3$ [2]. This technique is used to trap preys having big forms.

Ramos et al. have explored the idea of using a digital image as an environment for artificial ant colonies. They observed that artificial ant colonies could react and adapt appropriately their behavior to any type of digital habitat [17]. He also investigated ant colonies based data clustering and developed an ant colony clustering algorithm which he applied to a digital image retrieval problem. By doing so, he was able to perform retrieval and classification successfully on images of marble samples [10]. Liu et al. have conducted similar works and have presented an algorithm for grayscale image segmentation using behavior-based agents that self reproduce in areas of interest [12]. Hamarneh et al. have shown how an intelligent corpus callosum agent, which takes the form of a worm, can deal with noise, incomplete edges, enormous anatomical variation, and occlusion in order to segment and label the corpus callosum in 2D mid-sagittal images slices in the brain [13]. Bourjot et al. have explored the idea of using social spiders as a behavior to detect the regions of the image. The principle is to weave a web over the image by fixing silks between pixels [8].

In this paper, two methods based on an Ant System and a Spider System are described and compared with two classical methods. The first method consists on travelling on the pixels of the image and lays down a pheromone where each pixel validates our criteria: morphologic gradient. The second is a region-based technique which tries to fix silks between homogeneous pixels to construct webs.

This paper is organized as follows. Section II describes

Richard Moussa is a Phd student in the laboratory Labri, 351, cours de la Libération F-33405 Talence cedex e-mail: richard.moussa@labri.fr.

Marie Beurton-Aimar and Pascal Desbarats are associated professor in the laboratory Labri, 351, cours de la Libération F-33405 Talence cedex.

¹Multi-Agent System.

the two types of MAS with an explanation of the usage of such systems in the image segmentation domain. Section III presents the experimentation of our implementation by using comparison criteria with two segmentation techniques: Region Growing and Otsu thresholding. Finally, our implementation is discussed and we conclude with further expected improvements and perspectives for these systems.

II. SEGMENTATION WITH SOCIAL AGENTS

A MAS is a distributed system composed of a group of agents, which interact between themselves through an environment. Agents are classified into two categories: cognitive and reactive. Cognitive agents have a global view of the environment, they know the task for which they work. Conversely, reactive agents only know a restricted part of the environment. They react to environment stimulus and can modify this environment by adding or removing informations. Reactive agents do not know the complex task for which they work: they have a restricted set of simple features and they only apply them. Spiders or ants colonies are an example of reactive agents: each one knows locally what it has to do, but no one knows the more complex task for which they work.

Multi-agent systems are composed of an environment and a set of agents. For segmentation purpose, environment is created from a given grayscale picture: it is a matrix of gray pixels. The system and its agents have a life cycle. A cycle of the system consists in executing the life cycle of each agent. This life cycle is transposed to a step. The number of steps to be executed is given by the user. Two methods will be presented here: social ants and social spiders.

A. First MAS model: social ants

As previously mentioned, ants are social insects. They exhibit very good organization and construction abilities by colony behaviors. One of the important ones is their object searching behavior, in particular, how they can find the path to the object of interest from their nest. While walking from their nest to the object to be detected, ants leave on the way a kind of substance called *pheromone* whose concentration becomes weaker with time due to evaporating, forming in this way a pheromone trail. During their route, ants smell the pheromones deposited and when choosing their way, they tend to choose the most pheromoned direction. And the more the ants choose the same direction, the stronger the pheromone concentration is. Thus, this pheromone concentration helps the ants in choosing their shortest movement to the object of interest. Such algorithm is called ACO² algorithm [7] [18] [5]. In image segmentation domain, lots of proposed multi-agent methods have been inspired from this concept to elaborate a robust edge-based method [4] or region-based method [3].

For segmentation purpose, from the behavior explained above, we have chosen to use the act of depositing pheromones to perform our image segmentation task. This segmentation uses a number of ants that are injected randomly through the environment and guides them with a morphological

²Ant Colony Optimization.

gradient. The kernel used here to compute the gradient is a 3x3 pixels as shown in figure 1. If the pixel passes the test then an ant leaves a pheromone on it and steps to the pixel having the highest gradient in its neighborhood.

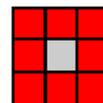


Fig. 1. Gradient computation kernel.

The pixels in the environment are classified into three categories: marked, visited and free. Figure 2 shows an example of the environment having these categories and where an ant is trying to move to another pixel. Firstly, each ant computes the morphological gradient on its own pixel. Then, the pixel is classified as visited or marked depending on the condition established by the user. This ant has the capability to move on its 8-neighborhood. Thus, an ant looks to the free pixels and moves to the one having the highest gradient. If not, the ant in question passes to an impasse status.

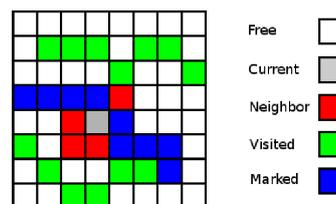


Fig. 2. Ants movement.

Algorithm 1 presents the description of the conditions presented above. The user has to fix three parameters: the percentage of pixels visited, the morphological gradient threshold and the number of agent. After that, the process begin trying to visit the percentage done by the user and marking pixels which passed the gradient condition. The complexity of this algorithm is $O(N_{agent} * NbT)$ where N_{agent} is the number of agents fixed by the user and NbT is the number of times the process passes the condition in line 2.

1) *Optimization*: In order to optimize the number of parameters to be delivered by the user, we have decided to fix the percentage of pixels to 100% to ensure that all the pixels were evaluated. For the morphological gradient threshold, we compute it as the minimum difference between two locals maxima of the histogram of the image having the highest distribution of pixels between them. The number of agent depends linearly from the maxima. Therefore, there is no absolute optimum value for the N_{agent} parameter but this problem can be bypassed by a numerical solution such as injecting one hundred times the number of local maxima.

B. Second MAS model: social spiders

Social spiders have been defined by the biologists to present stigmergic process like social insects. The characteristics of these societies and the importance of the silk in the various

Algorithm 1 Ants method

Require: Pixels: Matrix of pixels $\in \mathbb{N}^2$, PerV: Percentage of pixels visited and Grad: Morphological gradient threshold $\in \mathbb{R}$ and Nbagent: number of agents $\in \mathbb{N}$.

```
1: NbVisited  $\leftarrow$  0.
2: while PerV > Per(NbVisited) do
3:   for Each agent s do
4:      $G_1 \leftarrow$  ComputeGrad(Pixel(s)).
5:     if  $G_1 \geq$  Grad then
6:       Mark(Pixel(s)).
7:     end if
8:      $G_2 \leftarrow$  ComputeGrad(Neighborhood(Pixel(s))).
9:     Move(s, Position(Max( $G_2$ ))).
10:    NbVisited  $\leftarrow$  NbVisited + 1.
11:   end for
12: end while
```

behavior have created a different model from the social insects one. During their cycle, social spiders have the abilities to fix silks, move forward and move back. This model has characteristics which sufficiently distinguishes the levels of the realized spots, the society organization and the communication supports. Indeed, social spiders correspond to an interest model for three reasons [9]:

- 1) social spiders do not present any specialization in morphology and ethology;
- 2) an isolated social spider presents behavioral characteristics very close to lonely species;
- 3) social spiders show spectacular organization and cooperation forms, in particular, the web construction and the prey capture or its transportation phenomenon.

As mentioned before, Bourjot et al. have proposed a method using social spiders as a model of behavior to detect the regions of the image [8]. Its principle is to weave a web over the image by fixing silks between pixels using probabilistic movement. This method has been implemented and evaluated by Bourjot et al. and Moussa et al. [1]. It has given good results on synthetic images but failed on more complex images such as MRI³ images. Thereby, we have decided to build a new method by using some ideas from that described above.

Following the model previously described, we can design spiders as agents. Spiders are reactive agents. They are defined by an internal state composed of a set of parameters values, a current position and the last pixel where a spider has silked. These spiders have also an ability to move in the environment, to fix a silk⁴ and to come back⁵. Spiders which detect the same region can be grouped in a set called a colony. Spiders of a same colony share the same set of parameters values.

Spiders try to move through their 8-neighborhood, they prioritize the non-silked pixels and try to fix silks on them. If they fail, they move back to the last fixed silk. At the end of the process, groups of spiders are formed and are called

³Magnetic Resonance Imaging.

⁴Weave a dragline between two pixels.

⁵Return to the last fixed pixel.

regions. Figure 3 presents an example where a spider try to fix a silk or move back using the intensity variation as a condition.

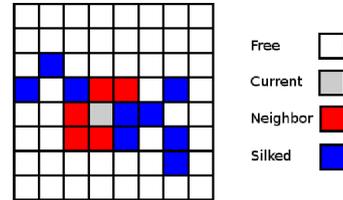


Fig. 3. Spiders movement.

Algorithm 2 performs as follow: for a number of steps delivered by the user, agents try to move through pixels for fixing silks and therefore detecting regions. The number of agents is also fixed by the user and a threshold allowing the spider to fix a silk and therefore to move forward or to move back. Its complexity is about $O(N_{bagent} * N_{bit})$ where N_{bagent} is the number of agents fixed by the user and N_{bit} is the number of steps that the spiders should do.

Algorithm 2 Spiders method

Require: Pixels: Matrix of pixels $\in \mathbb{N}^2$, Nbagent: number of agents $\in \mathbb{N}$, Nbit: Iteration number $\in \mathbb{N}$ and Thres: grayscale Threshold $\in \mathbb{N}$.

```
1: while Ite - > 0 do
2:   for Each agent a do
3:      $T \leftarrow$  computeInt(Pixel(a)).
4:     if  $T \leq$  Thres then
5:       Move(a, Position(Pixel(T))).
6:       Silkfixing(Position(Pixel(a)), Position(Pixel(T))).
7:       LastFixedSilk(a)  $\leftarrow$  Pixel(a).
8:     else
9:       Moveback(a, LastFixedSilk(a)).
10:    end if
11:   end for
12: end while
```

1) *Optimization:* In this case, only the threshold has been optimized. Its computation consists on the minimum variation of two locals maxima. But for the other two parameters, at present, we are not able to compute them automatically due to their dependency between each other.

III. METHODOLOGY USED FOR COMPARISON

In this section, we compare the social spiders method with other segmentation methods while the social ants method is interpreted separately. We do not search for counting the contours but to evaluate the result of the social ants segmentation. These comparisons allow us to determine whether the social spiders and ants methods brings something positive compared to traditional segmentation methods.

We use these comparisons on two other methods:

- a classification method by thresholding: the Otsu method [15];
- a region-based method: the Region Growing method [16].

To compare these methods, we need to establish criteria to be used on all test images. We compare the results on several points:

- 1) the number of regions;
- 2) correspondence between the regions of the model and the segmentation result;
- 3) execution time.

Definition Let Γ be an image and Δ its segmentation result. We call Γ_i the region i of the image and Δ_j the region j of the result.

The number of regions allows us to determine whether the method considered detects a number of regions close to reality. In the case of noisy images, it is possible that some methods detect regions with insignificant size. That is why we add to the total number of regions, the number of regions having insignificant size. For our output segmentation, we consider a region as insignificant if its size is less than 10 pixels.

The computation of the number of regions is done on the segmentation method result on which a labeling is added to the connected components to consider the regions connected.

The correspondence between initial image and its result enables us to determine if the regions identified by the method correspond to the regions defined in the initial image. This is only possible in the case of synthetic images.

To compute the accuracy, it is necessary to determine which region Δ_j matches the most the region Γ_i . This region is determined by:

$$\begin{aligned} n^i &= \frac{\text{the total number of pixels}}{\text{the pixels of } \Gamma_i} \\ n_j &= \frac{\text{the total number of pixels}}{\text{the pixels of } \Delta_j} \\ n_j^i &= \frac{\text{the total number of pixels}}{\text{the common pixels between } \Delta_j \text{ and } \Gamma_i} \end{aligned} \quad (1)$$

Thus, it is possible to compute $\delta_{i,j} = \frac{n_j^i}{n^i}$ and $\gamma_{i,j} = \frac{n_j^i}{n_j}$ representing respectively the proportion of pixels of Γ_i belonging to Δ_j and the proportion of pixels of Δ_j belonging to Γ_i . We have two ways to choose the region that corresponds to Δ_j corresponding the most to Γ_i :

- 1) Δ_k as the value of $\delta_{i,k}$ is maximum: in this case, we prefer the fact that Γ_i and Δ_j have a maximum of pixels in common;
- 2) Δ_k as $\delta_{i,k} + \gamma_{i,k}$ is maximum: same as above, but we add the requirement that Δ_k must have a minimum of pixels in other regions than Γ_i .

In our results, we indicate two points, *accuracy _{δ}* and *accuracy _{$\delta+\gamma$}* , which corresponds respectively to the two choices of Δ_k described above. In both cases, the accuracy will be the average values for all regions of the model.

A. Region Growing

The *Region Growing* method consists on building a region from one chosen pixel and then adding recursively neighbors whose grayscale difference with the original pixel is below a threshold [16].

This method tries to grow an initial region by adding to this region the connected pixels that do not belong to any region. These pixels are the neighborhood pixels already in the region and whose grayscale is sufficiently close to the area. When it is not possible to add pixels, we create a new region with a pixel that has not been selected yet, then we grow the region.

The method ends when all the pixels were chosen by a region.

B. Otsu

Otsu has developed a multi-level thresholding method [20]. Its aim is to determine, for a given number of regions, the optimum values of different thresholds based on the variance of subdivisions created.

The basic method consists on separating the foreground from the background. In this case, we search the optimal threshold to split the pixels in two regions. For a threshold t , it is possible to compute the *between-class variance* $\sigma^2(t)$. This measure is derived from the average intensity μ_1 , μ_2 and μ of classes $[0; t]$, $[t + 1; L]$ and $[0; L]$ where L is the maximum intensity.

The Equation 2 introduce the computation of σ^2 , where w_1 and w_2 represent the proportion of pixels in the class $[0; t]$ and $[t + 1; L]$ compared to the total number of pixels.

$$\sigma^2(t) = w_1(t)(\mu_1(t) - \mu)^2 + w_2(t)(\mu_2(t) - \mu)^2 \quad (2)$$

The Otsu method shows that the optimal threshold t^* is obtained for a between-class variance. The method consists on computing the variance for all possible thresholds ($t \in \{1; \dots; L - 1\}$) and determining its maximal value.

This method could be extended easily to the computation of M classes with $M - 1$ thresholds $\{t_1; t_2; \dots; T_{M-1} - 1\}$ ($t_1 < t_2 < \dots < t_{M-1}$). The between-class variance is defined then as follows:

$$\sigma^2(t_1, \dots, t_{M-1}) = \sum_{M-1}^M w_k(\mu_k - \mu)^2 \quad (3)$$

where w_k represent the proportion of pixels in the class $[t_{k-1}; t_k]$ ⁶, μ_k the intensity average of this same class and μ the intensity average of the class $[0; L]$.

For each $M-1$ -uplet, we compute thresholds of the between-class variance. The optimal thresholds, (t^*, \dots, t_{M-1}^*) , correspond to the maximum value of the between-class variance.

Chen et al. propose an algorithm that minimizes the number of necessary computation to obtain a faster algorithm [15]. This method had been implemented for our evaluations tests.

C. Experimentations

Now, we will present the results of the experimentation on 2D images with and without noise. Firstly, we will see the results of segmentation obtained with a non-noisy image to ensure the functioning of the different methods, then we will see the results on a noisy-image to determine the noise-resistance of the spiders and ants methods. For the spiders

⁶ $t_0 = 0$ and $t_M = L$.

method, the non-detected pixels are colored. Therefore, they will be seen as a noise in the image segmentation results.

The execution time to be given comes from the simulation of the methods on a machine equipped with an Intel Quad Q9550 (4 cores having 2.83GHz) and 4GB of RAM. The operating system of this machine is a Linux kernel 2.6.21 x86_64. The synthetic and the brain images are respectively composed of 10 and 94 regions. The size of the test images is 256x256 pixels.

Figure 4 and 5 presents respectively the results of the different image segmentation techniques applied on a non-noisy synthetic and brain image. Their informations are explored in table I and II.

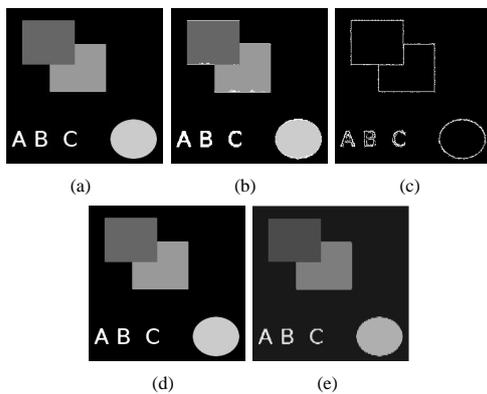


Fig. 4. 2D segmentation of synthetic image: a) Original image b) Social spiders, c) Social ants, d) Region Growing, e) Otsu.

	Social spiders	Social ants	Region Growing	Otsu
Region	11	x	10	10
Region > 10px	11	x	10	10
<i>Accuracy_σ</i>	98.3 %	95.5 %	100 %	100 %
<i>Accuracy_{σ+γ}</i>	98.3 %	93.2 %	100 %	100 %
Time	318 s	0.2 s	0.4 s	15 s

TABLE I
2D RESULTS: SYNTHETIC IMAGE WITHOUT NOISE.

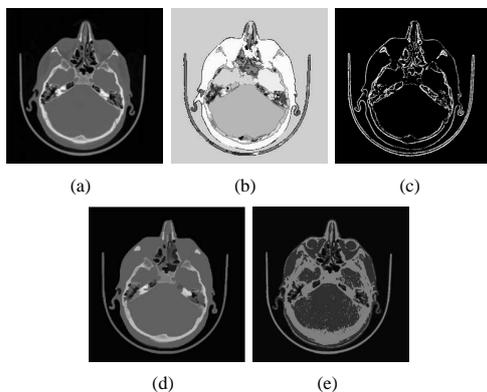


Fig. 5. 2D segmentation of Brain image: a) Original image b) Social spiders, c) Social ants, d) Region Growing, e) Otsu.

	Social spiders	Social ants	Region Growing	Otsu
Region	223	x	8670	1703
Region > 10px	56	x	456	376
<i>Accuracy_σ</i>	71.4 %	95.1 %	65.5 %	91.2 %
<i>Accuracy_{σ+γ}</i>	65.8 %	92.6 %	59.7 %	89.7 %
Time	323 s	0.3 s	0.5 s	14 s

TABLE II
2D RESULTS: BRAIN IMAGE WITHOUT NOISE .

For the synthetic image, the results of Region Growing (threshold = 25) and Otsu (thresholds = 60, 127, 178 and 204) methods have a maximum accuracy with a number of regions that corresponds to the image. The spiders method (iterations = 100000, spiders = 10000 and threshold = 25) has a region that corresponds to the extra pixels that have been detected by any spider. This region is not connected, the pixels that compose it are scattered throughout the image. These pixels are merged with the most likely region. The processing time is bigger than the other methods with less accurate results. The ants method (ants = 10000, gradient = 51) has delivered a good accuracy for the detection of the contours with the same number of contours for the original image. The supplement region discussed above on the segmentation image obtained by social spiders is composed of scattered contours. These contours are found by the social ants segmentation with a good precision and a fast computation time.

In the case of the brain image, the Region Growing method (threshold = 13) has the lowest accuracy and the biggest number of regions. Otsu method (thresholds = 15, 44, 76 and 95) produced the highest number of insignificant regions which leads an oversegmentation of the image. Same as above, The social spiders method (iterations = 100000, spiders = 10000 and threshold = 25) has pixels not selected by any spider (ie. they are colored by black in the image and spreaded like noise). It is therefore possible to perform a post-treatment that would link these pixels to the colony that have a strong presence in their neighborhoods. Its accuracy is better the the other methods. As for the social ants method (ants = 10000, gradient = 26), it has approximatively recovered all the contours with good accuracy.

Despite the fact that the difference between region-based segmentation methods is small, the accuracy of the results of the spiders and the ants methods are worse than the other methods for the synthetic image case and better for the brain case. However, as the spiders and the ants methods are stochastic methods, we do not expect to get maximum accuracy. Let us test that this accuracy will remain stable when adding noise.

For that, we added noise to the original images (20%). The results of the different image segmentation techniques applied on the synthetic and brain noisy images are respectively presented in figure 6 and 7 . Their statistics are mentioned in table III and IV.

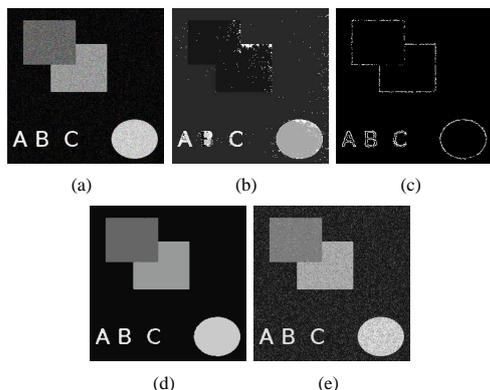


Fig. 6. 2D segmentation of synthetic image with noise: a) Original image b) Social spiders, c) Social ants, d) Region Growing, e) Otsu.

	Social spiders	Social ants	Region Growing	Otsu
Region	425	x	89	3096
Region > 10px	51	x	10	580
$Accuracy_{\sigma}$	84.6 %	75.2 %	99.9 %	50.1 %
$Accuracy_{\sigma+\gamma}$	81.5 %	73.3 %	99.9 %	34.6 %
Time	327 s	0.2 s	0.5 s	18 s

TABLE III
2D RESULTS: SYNTHETIC IMAGE WITH NOISE.

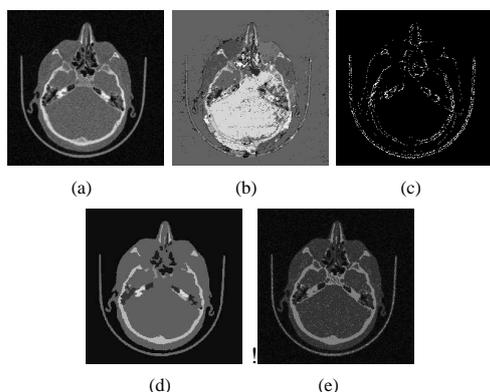


Fig. 7. 2D segmentation of Brain image with noise: a) Original image b) Social spiders, c) Social ants, d) Region Growing, e) Otsu.

	Social spiders	Social ants	Region Growing	Otsu
Region	223	x	1846	3096
Region > 10px	56	x	93	580
$Accuracy_{\sigma}$	75.3 %	60.2 %	65.4 %	72.2 %
$Accuracy_{\sigma+\gamma}$	70.1 %	59.4 %	43.3 %	65.9 %
Time	388 s	0.4 s	0.5 s	15 s

TABLE IV
2D RESULTS: BRAIN IMAGE WITH NOISE .

In the case of synthetic noisy image, adding noise caused a decrease in the accuracy of the results of all methods except Region Growing (threshold = 50) which presents a robustness to noise. The result of the spiders method (iterations = 100000,

spiders = 10000 and threshold = 50) has decreased in term of accuracy. Furthermore, the difference between the accuracy of the non-noisy image and the noisy one is minimal for the spiders method. These two points allow us to say that the spiders method is less sensitive when adding noise to the image. The number of regions has increased for the three methods compared to the non-noisy image segmentation. However, a number of regions rather high can be explained by a number of pixels non-detected more important, leading to disconnection of the regions. We note that the Otsu method (thresholds = 15, 64, 134 and 200) produces the most regions. This method, unlike Region Growing and social spiders method, have produced an important number of insignificant regions which implies oversegmentation of the image. As for the social ants method (gradient = 102), the accuracy has decreased due to noise effect.

For the noisy brain image, the Region Growing (threshold = 25) and the Otsu (thresholds = 17, 52, 90 and 117) have oversegmentated the image despite the fact that the Region Growing method obtained a number of significant region closer to the reality. The accuracy of the social spiders method (iterations = 100000, spiders = 10000 and threshold = 25) has decreased the less and became the best one in term of performance. The accuracy of the social ants method (gradient = 84) has made an important decrease due to noise effect.

The execution time of all methods remained stable. Therefore, it appears that the social spiders segmentation is robust to noise effect. This robustness has however led to a light oversegmentation of the image without influencing the time process.

IV. CONCLUSION AND PERSPECTIVES

In this paper, we have presented two methods of segmentation. The first corresponds to a contour-based technique, the social ants, which has produced a good segmentation where the method has recovered the contours of the non-noisy images. As for the noisy ones, the contours are scattered for an accuracy less important. On the contrary, the social spiders method, a region-based method, has produced a non negligible time processing in the case of non-noised image with a result less important than the others. And when noise is added, the processing time remained stable but with a better accuracy than for the other region-based methods. Note that the results of social spiders method are influenced by the repartition of the agents on the matrix and the number of step to do.

We have made comparisons between the results of the social spiders, social ants, Region Growing and the Otsu methods. These comparisons focused on the accuracy, the number of regions produced and the time processing of the methods. They are not exhaustive comparisons where all aspects of segmentation are not taken into account.

Through these comparisons, we have put forward some drawbacks on the social spiders method. Particularly, we have seen that this method produced a significant number of areas and that the execution time was particularly long as discussed above.

As we can see from the results in table I, the social spiders method produces a new region constructed by pixels

non-silked. These pixels are composed of scattered contours. Therefore, the social spiders and the social ants have complementary roles and merging the two methods will produce a new segmentation having accurate contours on the resulting images.

REFERENCES

- [1] R. Moussa, M. Beurton-Aimar, P. Desbarats and G. Savin. Image segmentation using social agents. Research report, 21 pages, december 2008.
- [2] E. Jackson. Social spiders. *Current Biology*, 17(16):R650–R652, 2007.
- [3] S. Saatchi and C.C Hung: Using Ant Colony Optimization and Self-organizing Map for Image Segmentation. *MICAI 2007*: 570-579.
- [4] Y. Han and P.F. Shi. An improved ant colony algorithm for fuzzy clustering in image segmentation. *Neurocomputing 2007*, volume 70, 665-671
- [5] F. Guinaud and Y. Pign. Problem Solving and Complex Systems. In *Emergent Properties in Natural and Artificial Dynamical Systems*, A. Alaoui and C. Bertelle (Ed.) (2006) 53-86.
- [6] B. Jähne. *Digital Image Processing*. Springer, sixth edition, 2005.
- [7] C. Bertelle, A. Dutot, F. Guinand and D. Olivier: Dynamic Placement Using Ants for Object Based Simulations. *CoopIS/DOA/ODBASE 2003*: 1263-1274.
- [8] C. Bourjot, V. Chevrier and V. Thomas. A new swarm mechanism based on social spiders colonies : from web weaving to region detection. *Web Intelligence and Agent Systems : An International Journal - WIAS*, 1(1) :4764, Mar 2003.
- [9] V. Chevrier. *Contributions au domaine des systemes multi-agents*. Hdr, Universit Henry-Poincarre - Nancy 1, Janvier 2002.
- [10] V. Ramos, F. Muge, and P. Pina. Self-organized data and image retrieval as a consequence of inter-dynamic synergistic relationships in artificial ant colonies. *Frontiers in Artificial Intelligence and Applications, Soft Computing Systems - Design, Management and Applications, 2nd International Conference on Hybrid Intelligent Systems*, IOS Press, in Javier Ruiz-delSolar, Ajith Abraham and Mario Kppen (Eds.), volume 87, ISBN 1 5860 32976:500509, Santiago, Chile, December 2002.
- [11] M. Wooldridge. *An Introduction to MultiAgent Systems*. Wiley and Sons, ISBN 0-471-49691-X, 2002.
- [12] J. Liu and Y. Tang. Adaptive image segmentation with distributed behavior-based agents. *IEEE Transactions Pattern Analysis and Machine Intelligence*, volume 21, Issue 6:544551, 1999.
- [13] G. Hamarneh, T. McInerney, and D. Terzopoulos. Deformable organisms for automatic medical image analysis. *Medical Image Analysis*, pages 6676, 2001.
- [14] H. He and Y. Chen. Artificial life for image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, volume 15, Issue 6:9891003, 2001.
- [15] T.-S. Chen, P.-S. Liao, and P.-C. Chung. A fast algorithm for multilevel thresholding. *Journal of Information Science and Engineering*, 17:713727, 2001.
- [16] L. Shapiro and G. Stockman. *Computer Vision*. Prentice-Hall, 2001.
- [17] V. Ramos and F. Almeida. Artificial ant colonies in digital image habitats - a mass behaviour effect study on pattern recognition. *Proceedings of ANTS2000 - 2nd International Workshop on Ant Algorithms (From Ant Colonies to Artificial Ants)*, in Marco Dorigo, Martin Middendorf and Thomas Stzle (Eds.), pages 113116, Brussels, Belgium, 7-9 September 2000.
- [18] M. Dorigo, V. Maniezzo and A. Colorni, Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 26(1):29-41,1996.
- [19] G. Theraulaz and E. Bonabeau. Coordination in Distributed Building. *Science*, 269: 686-688, 1995.
- [20] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. Systems, Man and Cybernetics*, 9:6266, 1979.

Dynamical Systems Based on Dynamic Graphs

Stéphane Despréaux, Aude Maignan
stephane.despreaux@imag.fr, aude.maignan@imag.fr

Abstract—The aim of this talk is to present a mathematical framework for the modeling of agent networks called *dynamical system based on dynamic graphs*. The agents have local interactions and their behavior obeys a dynamical system. The network evolves, during time. Agents and links between agents can be added or removed thanks to local deterministic transformation rules. This framework have been implemented in the software called *DynSys*.

I. INTRODUCTION

A dynamical system is a fixed rule that describes what future states follow from the current state. Its evolution is described by a fixed phase space and it is supported by a static graph.

But in many applications, including communication networks, embedded systems or biological behaviors (such as collaboration or communication between ants, behavior of a set of cells which share local information, etc.) graphs are subject to discrete changes, such as additions or deletions of agents or links. In the last decade there has been a growing interest in such dynamically changing graphs [1]. The difficulty of such structures is to deal, simultaneously, with the evolution of a number of agents (or nodes), the number of links between agents, the states of agents, and eventually, the states of links.

For biological problems, specific particular models have been developed such as L-systems [6], [7], adaptive dynamics [3], TreeGCS (hierarchical Growing Cell Structures) [2] and DS2 (Dynamical System with a Dynamic Structure) [4]. On the other hand, for robotic problems, graph grammar theory has been used [5].

In this paper, we propose a framework called *Dynamical system based on dynamic graphs* such that agents and links between agents correspond, respectively, to nodes and edges of a graph.

- Each node of the graph obeys a dynamical system.
- Each edge of the graph obeys a dynamical system.
- The evolution of nodes and edges is determined thanks to the state of there neighborhood.
- Under specific conditions, the graph evolves Nodes and edges can be added to or removed from the graph.

The graph evolution is determined thanks to local transformation rules. Our aim is to model agent reactions due to local informations. Consequently, transformation rules are function of a node and its neighborhood, or, of an edge and its neighborhood.

We have implemented the framework of *Dynamical system based on dynamic graphs* in a program called *DynSys*.

First we define the framework of *dynamical systems based on dynamic graph*. Secondly, the graph dynamic is presented. Finally, the *DynSys* program is shortly presented and an example is proposed.

II. MODELING OF A DYNAMICAL SYSTEM BASED ON DYNAMIC GRAPH

Let $G = (V_G, E_G)$ be a graph. V_G is the set of vertices of G and E_G the set of edges. $n(V_G)$ and $n(E_G)$ respectively denote the cardinal of V_G and E_G . Let us suppose that the values of the nodes and the edges of G belong to some sets respectively \mathcal{V} and \mathcal{E} (\mathcal{V} and \mathcal{E} could be a finite set, a finite field, \mathbb{R} , etc).

$S(G)$ is the state space of the system and is equal to $S(G) = \mathcal{V}^{n(V_G)} \times \mathcal{E}^{n(E_G)}$. A dynamic on G will be described by a flow

$$\Phi_G : \mathbb{R} \times \mathbb{R} \times S(G) \rightarrow S(G)$$

$$t, t_0, X_G(t_0), Y_G(t_0) \rightarrow X_G(t), Y_G(t)$$

Now, in order to define *dynamical system based on dynamic graphs*, we introduce some subsets of $S(G)$ where we can applied Φ_G : $\mathcal{D}(G, \Phi_G) \subset S(G)$ is the subset where Φ_G is the legal dynamic. Of course $(X_G(t_0), Y_G(t_0))$ has to belong to $\mathcal{D}(G, \Phi_G)$.

The *dynamical system based on dynamic graphs* mechanism is as follow :

- While $(X_G(t), Y_G(t)) \in \mathcal{D}(G, \Phi_G)$ apply Φ_G
- If $(X_G(t_1), Y_G(t_1)) \notin \mathcal{D}(G, \Phi_G)$ then apply a discrete transition.

A discrete transition arises from a set of rules. It depends on the exiting point of the domain.

We denote \mathcal{F} the set of flows and we denote $\mathcal{P}(\mathcal{E}^\infty)$ the set of parts of \mathcal{E}^∞ . ($\mathcal{E}^\infty = \bigcup_{n \in \mathbb{N}} \mathcal{E}^n$.) A discrete transition is defined by:

$$T : \mathcal{G} \times \mathcal{V}^\infty \times \mathcal{E}^\infty \rightarrow \mathcal{G} \times \mathcal{V}^\infty \times \mathcal{E}^\infty$$

$$G, X_G(t^*), Y_G(t^*) \rightarrow G', X_{G'}(t^*), Y_{G'}(t^*)$$

the discrete transition induces a dynamic on the graph. The new graph G' can have a different size and connections can change.

III. GRAPH DYNAMIC

A. Basic definitions

Basic graph-theory definition are reminded in this section.

Let G be a connected graph with vertices u and v . The distance $d(u, v)$ between u and v is the length of a shortest $u - v$ path in G . The eccentricity $e(v)$ of a vertex v is the distance from v to a vertex furthest from v . The radius of G is $\min\{e(v); v \in V(G)\}$, while the diameter of G is $\max\{e(v); v \in V(G)\}$.

Notation III.1. Let \mathcal{G}_1 be the set of graphs of radius equal to 1. Let \mathcal{G}_2 be the set of graphs of diameter less or equal to 3.

For any nonempty subset S of vertices in G , the induced subgraph $\langle S \rangle$ is the maximal subgraph of G with the vertex set S .

Let σ_i be the set of neighbors of node i and $\bar{\sigma}_i = \sigma_i \cup \{i\}$.

The **sub-graph of G centered on the node i** is the sub-graph Δ_i such that: $V_{\Delta_i} = \bar{\sigma}_i$ and $E_{\Delta_i} = \{(j, k) \in E_G, j \in V_{\Delta_i}, k \in V_{\Delta_i}\}$.

The **sub-graph of G centered on the edge (i, j)** is the sub-graph $\Delta_{(i,j)}$ such that: $V_{\Delta_{(i,j)}} = \bar{\sigma}_i \cup \bar{\sigma}_j$ and $E_{\Delta_{(i,j)}} = \{(j, k) \in E_G, j \in V_{\Delta_{(i,j)}}, k \in V_{\Delta_{(i,j)}}\}$.

Δ_i belongs to \mathcal{G}_1 and $\Delta_{(i,j)}$ belongs to \mathcal{G}_2 .

B. Local transformation rules

The graph dynamic models agent reactions based on local informations. Consequently transformation rules are function of a node and its neighborhood, or, of an edge and its neighborhood.

The neighborhood of a node i not only means the nodes of σ_i but also the edges of the induced sub-graph $\langle \sigma_i \rangle$. Consequently, Δ_i is the neighborhood of i . In the same way, $\Delta_{(i,j)}$ is the neighborhood of the edge (i, j) .

We propose two kinds of local transformation rules, the first one based on a node and the second one based on an edge.

Notation III.2. Let H be an induced subgraph of G , X_H is the dynamic state of the nodes of V_H and Y_H is the dynamic state of the edges of E_H .

Definition III.3. A local transformation rule based on a node is a function:

$$R_n: \quad \mathcal{G}_1 \times \mathcal{V}^\infty \times \mathcal{E}^\infty \longrightarrow \mathcal{G} \times \mathcal{V}^\infty \times \mathcal{E}^\infty \\ \Delta_i, X_{\Delta_i}(t^*), Y_{\Delta_i}(t^*) \longrightarrow N, X_N(t^*), Y_N(t^*)$$

such that $\sigma_i \subset V_N$.

A rule R_n is applicable to a subgraph Δ_i if a constraint of the following type is verified:

$$F_{R_n}(X_{\Delta_i}, Y_{\Delta_i}) = true.$$

This constraint corresponds to one of the boundary of the subset $\mathcal{D}(G, \Phi_G)$.

When R_n is applied to a sub-graph Δ_i of G , G is transformed to a new graph $G' = (V_{G'}, E_{G'})$ defined by:

- $V_{G'} = (V_G \setminus V_{\Delta_i}) \cup V_N$
- $E_{G'} = (E_G \setminus E_{\Delta_i}) \cup E_N$
- The node states and edge states of N are determined by reset function Z_{Δ_i} included in R_n :

$$Z_{\Delta_i}: \quad S(\Delta_i) \longrightarrow S(N) \\ X_{\Delta_i}, Y_{\Delta_i} \longrightarrow X_N, Y_N$$

The node states of $V_G \setminus V_{\Delta_i}$ remains unchanged. The edge states of $E_G \setminus E_{\Delta_i}$ remains unchanged also.

Definition III.4. A local transformation rule based on an edge is a function:

$$R_e: \quad \mathcal{G}_2 \times \mathcal{V}^\infty \times \mathcal{E}^\infty \longrightarrow \mathcal{G} \times \mathcal{V}^\infty \times \mathcal{E}^\infty \\ \Delta_{(i,j)}, X_{\Delta_{(i,j)}}(t^*), Y_{\Delta_{(i,j)}}(t^*) \longrightarrow N, X_N(t^*), Y_N(t^*)$$

such that $\sigma_i \subset V_N$ and $\sigma_j \subset V_N$. This kind of transformation rule occurs when a constraint of the following type is verified:

$$K_{R_e}(X_{\Delta_{(i,j)}}, Y_{\Delta_{(i,j)}}) = true.$$

When R_e is applied to a sub-graph $\Delta_{(i,j)}$ of G , G is transformed to a new graph $G' = (V_{G'}, E_{G'})$ defined by:

- $V_{G'} = (V_G \setminus V_{\Delta_{(i,j)}}) \cup V_N$
- $E_{G'} = (E_G \setminus E_{\Delta_{(i,j)}}) \cup E_N$

The node states and edge states of N are determined by the reset function $Z_{\Delta_{(i,j)}}$ included in the rule:

$$Z_{\Delta_{(i,j)}}: \quad S(\Delta_{(i,j)}) \longrightarrow S(N) \\ X_{\Delta_{(i,j)}}, Y_{\Delta_{(i,j)}} \longrightarrow X_N, Y_N$$

The node states $V_G \setminus V_{\Delta_{(i,j)}}$ remains unchanged and the edge state of $E_G \setminus E_{\Delta_{(i,j)}}$ remains unchanged also.

The discrete transition T is a composition of local transformation rules.

IV. THE DYN SYS PROGRAM AND AN EXAMPLE

DynSys is a program dedicated to the modeling and the simulation of dynamical systems based on dynamic graphs.

Each agent and link of the graph obeys a dynamical system which depend on there neighborhood.

Agents and links between agents can be added to or removed from the graph thanks to transformation rules described below.

Let us consider a simple example. The state of nodes and edges are one-dimensional variables, X and Y respectively. The dynamic function of the nodes is $\dot{X}_i = \frac{6}{\# \sigma(i)}$ where $\# \sigma(i)$ denotes the number of neighbors of the node i . Let Y_{ij} be the age of the edge (i, j) . Its dynamic is $\dot{Y}_{ij} = 1$.

At a given time t^* , a node i and one of its neighbors j create a new node k (see figure 1) when the constraint $X_i(t^*) + X_j(t^*) \geq 12$ is verified.

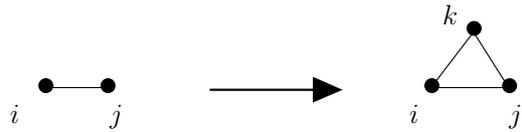


Fig. 1. Local transformation

The value of $X_k(t^*)$ is defined and the values of $X_i(t^*)$ and $X_j(t^*)$ are modified thanks to the reset functions: $X_k(t^*) = \frac{X_i(t^*) + X_j(t^*)}{2}$, $X_i(t^*) = \frac{X_i(t^*)}{2}$ and $X_j(t^*) = \frac{X_j(t^*)}{2}$. The edge (i, j) tenses and the nodes i and j merge when the constraint $Y_{ij}(t^*) \geq 5$ is verified. The reset function of the node i is then $X_i(t^*) = \frac{X_i(t^*) + X_j(t^*)}{2}$.

At each step, DynSys specifies the time variable t , the number of nodes N_s , the number of edges E_s , the maximal degree of the graph d and the number of local transformation rules which have been applied T_r .

Four steps of the example are drawn bellow thanks to DynSys.



Fig. 2. $t=0$, $N_s=2$, $E_s=1$, $d=1$, $tr=0$.

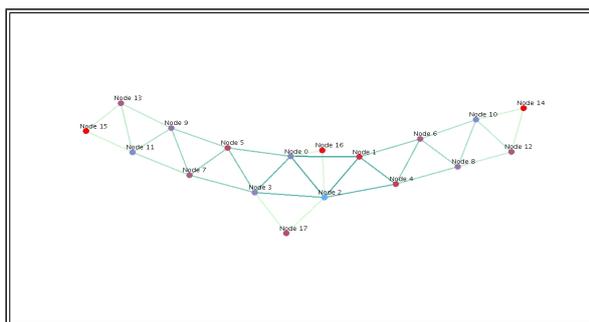


Fig. 3. $t=4.625$, $N_s=18$, $E_s=33$, $d=6$, $tr=13$.

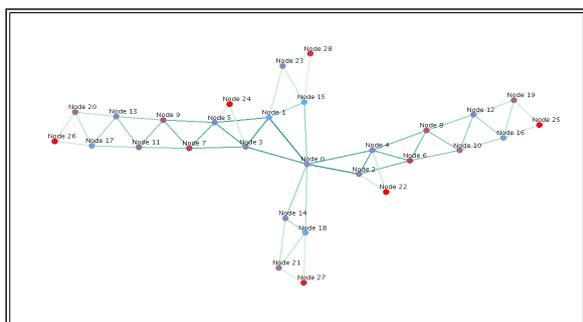


Fig. 4. $t=6.2093$, $N_s=29$, $E_s=53$, $d=7$, $tr=28$.

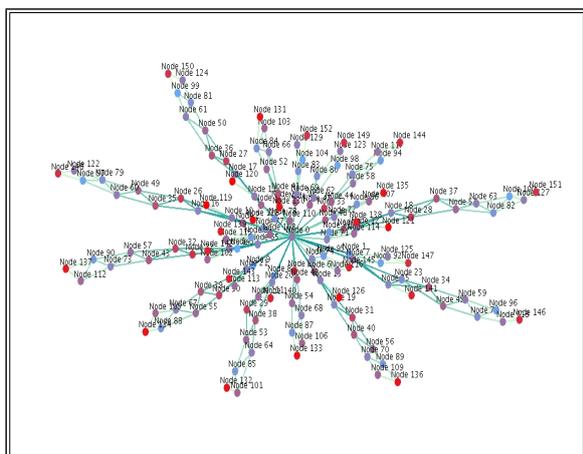


Fig. 5. $t=11.0242$, $N_s=153$, $E_s=289$, $d=38$, $tr=188$.

REFERENCES

- [1] F. Harary, G. Gupta, *Dynamic Graph Models*. Math. Comput. Modelling Vol. 25, No. 7, p 79-87, 1997.
- [2] Hodge, V.J.; Austin, J. Hierarchical, Growing Cell Structures: TreeGCS, Knowledge and Data Engineering, IEEE Transactions on Volume 13, Issue 2, Mar/Apr 2001 Page(s):207 - 218
- [3] Geritz, S. A. H., E. Kisdi, G. Meszna, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 12:35-57.
- [4] J.L. Giavitto, O. Michel, MGS: A rule-based programming language for complex objects and collections. In M. van den Brand and R. Verma, editors, *Electronic Notes in Theoretical Computer science*, volume 59. Elsevier Science Publishers, 2001.
- [5] J.-M. McNew and E. Klavins, Locally interacting hybrid systems with embedded graph grammars, Proc. of the 45th IEEE Conference on Decision and Control, 2006.
- [6] A Lindenmayer, *Mathematical models for cellular interaction in development (Parts I and II)*. Journal of Theoretical Biology, 18:280-315, 1968
- [7] P Prusinkiewicz, A Lindenmayer, *The algorithmic beauty of plants*. Springer-Verlag

Dynamic graph as a new type of graphs

Mounira NEKRI

Centre de Recherche sur l'Information Scientifique et Technique,
Université des Sciences et Technologie Houari Boumediene
mnekri@mail.cerist.dz

(joint work with Abdelkader Khelladi)

The graphs are generally used to model systems. They make it possible to highlight the deep structures of such systems. Hence the complexity of the graph depends directly on the complexity of these systems. The changing nature, of some systems, gives rise to dynamic graphs (which we add or delete vertices and/or edges over time). Thus, a dynamic graph is a graph which is structurally unstable making it difficult to be handled especially when it has an important number of nodes and edges.

Explicitly, A dynamic graph G is a collection of graphs
 $G = \{G_i = (V_i, E_i, t_i) / t_i \geq 0, i \in A \subseteq N\}$ such that :

$G_0 = (V_0, E_0, t_0)$ is called initial graph .

G_{i+1} is obtained from G_i by performing one of the following operations:

- Addition of nodes ($V_i \subset V_{i+1}$)
- Deletion of nodes ($V_{i+1} \subset V_i$)
- Addition of edges or arcs ($E_i \subset E_{i+1}$)
- Deletion of edges or arcs ($E_{i+1} \subset E_i$)

t_i is a positive real number relating to the situation of G_0 after the i th operation.

Between t_i and t_{i+1} the dynamic graph has the structure of the graph $G_i, \forall i \in A$.

F. Harary and G. Gupta suggest two approaches to study a dynamic graph:

1. Modelling a dynamic graph as a sequence of static graphs and studying the properties of this sequence.
2. Modelling a dynamic graph using a logic programming. The properties of dynamic graph can be determinate by executing the logic program that models it.

We can also study the dynamic graph by specifying the classes of graphs that remain invariant with time. For example let G be a C_4 (cycle of length four)

to which we added a vertex connected to all the vertices of C_4 .
By deleting any edge from G , the resulting graph contains a C_5 .

our approach is based on the following observation:

Generally G_i and G_{i+1} have not the same properties. For instance, let $G_{i+1} = G_i - x_0$. If G_i is connex and if x_0 is an articulation point then G_{i+1} is not connex. Hence the idea to transform G to H such that H is more stable than G . The procedure of transformation consists in reducing the sets of the nodes and edges. The graph, thus obtained, is an edge weighted graph. The introduction of the new graph makes it possible to use matric techniques in order to study the properties of a graph G through those of its reduced graph H . In particular the shortest paths and Matching problems.

Index

- N. ABCHA, 36
Haïfa ABDELHAK, 102
Cédric ABOUE NZE, 357
Kazuyuki AIHARA, 197
Modi ALNASR, 218
Gonzalo ALVAREZ, 198
José AMIGÓ, 185, 194, 198
R.E. AMRITKAR, 184
Ralph Gregor ANDRZEJAK, 329, 331
David ARROYO, 198
Thomas ASCHENBRENNER, 343
- Stefan BALEV, 290
Cyril BANDERIER, 97
Arnaud BANOS, 129
A. BARBIER, 29
Luc BARBIER, 29
Ana Cristina BARROSO, 169
E. BARTHEL, 34
A. BATAILLE, 29
Farida BENMAKROUHA, 75, 79
Younès BENNANI, 53
Cyrille BERTELLE, 129, 151
É. BERTIN, 37
Marie BEURTON-AIMAR, 369
Luiz BEVILACQUA, 20
O. BEZENCENET, 29
A. BIRZU, 190
Daniel BONAMY, 25, 29, 30, 46
Emmanuel BONNET, 110, 125
I. BOU MALHAM, 45
E. BOUCHAUD, 44
D. BOUNOIARE, 13
- C. BRUN, 47
Wolfram BUNK, 343
- R. CANDELIER, 35
Xavier CARTON, 193
Ms. S. CHAKRABORTY, 268
Serge CHAUMETTE, 355
Yong CHEN, 213
Daniel CHICHARRO, 331
Sergio CILIBERTO, 31
F. CÉLARIÉ, 46
Eric CLEMENT, 32
Mattia COCCOLO, 200
Gilles COLLETTE, 33
J. COUSTY, 29
Gisella CROCE, 169
Daniel J. CROSS, 3
Olivier CRUMEYROLLE, 49
A. CUVELIER, 13
- Davy DALMAS, 34, 46
Syamal Kumar DANA, 189
T. DARNIGE, 32
Subir DAS, 264
Tien DAT CHU, 48
J. DAUCHET, 32
Olivier DAUCHOT, 35
Éric DAUDÉ, 102, 110, 116, 297
François DAVIAUD, 25
A. DEBOEUF, 41
Daniel DELAHAYE, 121
Pascal DESBARATS, 369
Stéphane DESPRÉAUX, 376
Hassan DIARAMOUNA SIDIBE, 175

Rui DILÃO, 185, 201
 Johnny DOUVINET, 121
 Moez DRAIEF, 64
 M. DROZ, 37
 E. DUBOIS, 39
 Edwige DUBOS-PAILLARD , 140
 B. DUBRULLE, 33
 Gérard H.E. DUCHAMP, 51, 60
 V. DUPUIS, 39
 Antoine DUTOT, 151
 Arsen DZHANOEV, 202

Ahmad Rami EL-NABULSI, 223
 Emmanuel ELIOT, 99, 110
 Amine ELTAIEF, 175
 G. Hussian ERJAE, 218
 Alexander EZERSKI, 36, 48

Philippe FAURE, 24
 Thierry FAURE, 42
 Jérémy FIEGEL, 129
 R. FISCHER, 43
 Mikhail V. FOURSOV, 84
 Charles Henri FREDOUET, 321
 Ubiratan S. FREITAS, 333
 Ubiratan. S. FREITAS, 13

David GAILLARD , 125
 Laurent GAUBERT, 64
 G. GAUTHIER, 41
 R. GHADERI, 247
 Robert GILMORE, 3
 P. GONDERT, 43
 Didier GONZE, 8
 Silvia GOODENOUGH, 60
 M.A. GOROKHOVSKI, 50
 Guillaume GRÉGOIRE, 37
 Ioan GROSU, 190
 C. GUERRA, 46
 Raphaël GUILLERM, 38
 Frédéric GUINAND, 290, 355, 357

M. HASLER, 190
 Christiane HESPEL, 75, 79, 84
 S.H. HOSSEINIA, 237, 242, 247
 M. HOYOS, 32

Damien JAMET, 64
 A. JARNO-DRUAUX, 48
 N. JARRIGE, 45
 Yiliang JIN, 213

Sylvain JOUBAUD, 31

K. KATSUYOSHI, 39
 Fahem KEBAIR, 134
 Fayçal KELAI, 49
 Judy KENNEDY, 191
 A. KERLAIN, 39
 Abdelkader KHELLADI, 379
 Takashi KOHNO, 197
 M. KONCZYKOWSKI, 39
 Daniel KROB, 97
 Dimitris KUGIUMTZIS, 329, 338

Denis L'HÔTE, 39, 47
 Rabah LABBAS, 157
 F. LADIEU, 47
 Abdelkader LAKMECHE, 199
 Patrice LANGLOIS, 116, 121, 297
 Christian LAVALT, 53
 Michel LE VAN QUYEN, 347
 Mustapha LEBBAH, 53
 Marc LEFRANC, 24
 Klaus LEHNERTZ, 331
 Gaëtan LESAUVAGE, 290
 Annick LESNE, 24
 Christophe LETELLIER, 1, 13, 42, 333
 Hua LI, 258
 Xin LI, 213
 Xiaoran LIN, 258
 Patrick LIONS, 151
 Max LITTLE, 353
 Alexander LOSKUTOV, 202
 François LUSSEYRAN, 42

M. MAHMOUDIAN, 242
 Aude MAIGNAN, 376
 Stéphane MAINGOT, 174
 Paul MANNEVILLE, 40
 F. MARIN, 48
 Jérôme MARTIN, 41, 45
 Ahmed E. MATOUK, 232, 253
 D. MAYER, 195
 Ahmed MEDEGHRI, 174
 Salim MESSAOUDI, 157, 174
 Stanislaw MIGORSKI, 164
 P. MILANI, 347
 Hoang Ngoc MINH, 97
 G. MINO, 32
 Fatimetou MINT AGHRABATT, 175
 Jean Pierre MÜLLER , 140
 Shaher MOMANI, 218, 237, 242, 247

Roberto MONETTI, 343
 Edouard MONNIER, 75
 Florian MORMANN, 331
 V. MOSSER, 39
 Richard MOUSSA, 369
 Mohand MOUSSAOUI, 174
 J.-F. MUIR, 13
 Innocent MUTABAZI, 25, 36, 38, 49

 Michel NABAA, 151
 R. NAECK, 13
 S. NAKAMAE, 39
 E. NASERI, 237, 242
 V. NAVARRO, 347
 Mounira NEKRI, 379
 T. NGUYEN, 29
 Stavros NIKOLOPOULOS, 347

 Anna OCHAL, 164
 Zaid M. ODIBAT, 205, 208
 Damien OLIVIER, 102, 151, 290
 I. O. OLUSOLA, 195
 Sidi Mohamed OULD MOHAMED MOCTAR,
 308

 D. PAILLARD, 33
 Rodrigo PAIVA, 273
 Luc PASTUR, 42
 Karol A. PENSON, 60
 Xavier PERROT, 193
 R. PERZYNSKI, 39
 Michel PETITOT, 91
 Artyom PETROSSYAN, 31
 Nélio D. PIZZOLATO, 273, 281
 Laurent POINSOT, 60
 L. PONSON, 30
 A. PORTMANN, 13
 Arnaud PRIGENT, 25, 38
 Eliane PROPECK, 125
 Damienne PROVITOLLO, 99, 140

 Marc RABAUD, 43
 Marat RAFIKOV, 20
 N. RAKOTOMALALA, 45
 A. RANJBAR, 237, 242, 247
 Ana Margarida RIBEIRO, 169
 Nicoleta ROGOVSKI, 53
 Cindy ROUNTREE, 44
 A. ROUSSELET, 32
 S. ROUX, 44
 Márcio ROZENTAL, 281

 Luo RUNZI, 228

 Juan SABUCO, 192
 Thierry SAINT-GÉRARD, 125
 Dominique SALIN, 41, 45
 Miguel A. F. SANJUÁN, 177, 182, 188, 192,
 194, 196, 200, 202
 S. SANTUCCI, 30
 V.L. SAVELIEV, 50
 Luiz F. SCAVARDA, 273
 Julien SCHEIBERT, 46
 Abhijit SEN, 186
 Jesús M. SEOANE, 187
 Frédéric SERIN, 134
 M. SHAHIRI T., 247
 Deng SHUCHENG, 228
 Sitabhra SINHA, 203
 Michael SMALL, 347
 V.M. SOKOL, 204
 Stephan SPRINGER, 343

 Hanane TAFAT, 97
 Y. TAHRI, 39
 M. TALAMALI, 44
 L. TALON, 45
 Moncef TEMANI, 302
 Caroline THIBIERGE, 39, 47
 Pierrick TRANOUEZ, 116, 297

 Mario VALDERRAMA, 347
 D. VANDEMBROUCQ, 34, 44
 Ricardo Luiz VIANA, 183
 Samuel VIDAL, 91
 E. VINCENT, 39
 U. E. VINCENT, 195

 Alexandre WAGEMAKERS, 196
 Lei WANG, 159
 Ana Paula P. WYSE, 20

 Xiaoke XU, 347

 Adnan YASSINE, 271, 302
 James A. YORKE, 181

 Samuel ZAMBRANO, 182, 192, 194, 196, 200
 Mahdi ZARGAYOUNA, 302
 Bisma ZEDDINI, 302
 Quanxin ZHANG, 159
 Wei ZHENGMIN, 228
 Shangbo ZHOU, 258

ICCSA 2009

June 29 - July 02

3rd International Conference on Complex Systems and Applications

The first International Conference on Complex Systems and Applications (ICCSA) was held in Huhhot during June 15-18, 2006. The second ICCSA was held in Jinan, China, during June 8-10, 2007. These two first ICCSA were huge successes. The third ICCSA is held in France, Normandy, Le Havre on the Seine river estuary, during June 29 - July 2, 2009.

This conference provides a unique international forum, where exciting interactions and communications take place among researchers, and it brings fruitful cooperation and collaborations to the world community. The third ICCSA focuses on recent advances in complex systems and applications in all fields of science and engineering. There are several invited expository addresses covering recent trends and many invited lectures on problems of current interest and important applications in various disciplines.

