INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Markov chains with discontinuous drifts have differential inclusions limits. Application to stochastic stability and mean field approximation.

Nicolas Gast — Bruno Gaujal

N° 7315 — version 3

initial version June 2010 — revised version April, 2011

Thème NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7315--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche GRENOBLE - RHÔNE-ALPES

Markov chains with discontinuous drifts have differential inclusions limits. Application to stochastic stability and mean field approximation.

Nicolas Gast, Bruno Gaujal

Thème NUM — Systèmes numériques
Équipe-Projet MESCAL

Rapport de recherche n° 7315 — version 3 — initial version June 2010 — revised version April, 2011 — 26 pages

Abstract: In this paper, we study deterministic limits of Markov processes having discontinuous drifts. While most results assume that the limiting dynamics is continuous, we show that these conditions are not necessary to prove convergence to a deterministic system. More precisely, we show that under mild assumptions, the stochastic system is a stochastic approximation algorithm with constant step size that converges to a differential inclusion. This differential inclusion is obtained by convexifying the rescaled drift of the Markov chain.

This generic convergence result is used to compute stability conditions of stochastic systems, via their fluid limits. It is also used to analyze systems where discontinuous dynamics arise naturally, such as queueing systems with boundary conditions or with threshold control policies, via mean field approximations.

Key-words: Mean field, fluid limit, stability, differential inclusion, non-smooth dynamics, queueing systems.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Les chaînes de Markov ayant une dérive discontinue ont pour limite une inclusion différentielle.

Résumé : Ce document étudie des limites d'échelle de chaînes de Markov ayant une dérive discontinue. Alors que beaucoup de travaux sur le sujet supposent que la dérive est continue, nous montrons que cette condition n'est pas nécessaire pour obtenir une limite déterministe.

Nous montrons que sous des hypothèses faibles, un passage à l'échelle d'une chaîne de Markov peut être vu comme un algorithme d'approximation stochastique à pas constant, qui converge vers l'ensemble des solutions d'une inclusion différentielle. Cette inclusion est obtenue à partir d'une convexification de la dérive du processus initial.

Cette méthode est générique et permet de calculer la région de stabilité de nombreux systèmes stochastique, en étudiant leur limite fluide. Elle permet aussi d'étendre les techniques d'approximation champ moyen à des systèmes où les discontinuités apparaissent naturellement, comme des réseaux de files d'attente ou des systèmes contrôlés par une politique à seuil.

Mots-clés : Champ Moyen, Limite fluides, stabilité, Inclusions différentielles, dynamiques non-continues, réseaux de files d'attente.

1 Introduction

The use of ordinary differential equations has proved useful for performance evaluation of computing systems and communication networks. Here are a few sticking examples: Fluid limits have been used to prove stability of a large class of queuing systems [10, 9]; The performance of the wifi protocol 802.11b has been analyzed using a mean field approximation in [5, 6] and distributed algorithms such as work stealing [22, 15] have also been studied using the famed population dynamics approach introduced by Kurtz [19].

In this paper, we show that both scalings (fluid limit and mean field) can be studied within a common framework, by seeing a Markovian stochastic system as a *stochastic approximation* of a deterministic differential system driven by the rescaled *drift* of the initial system.

Under classical smoothness assumptions on the drift, there exist general results that show that the limiting system (when the scaling parameter goes to infinity) can be described by a system of deterministic ordinary differential equations

$$\dot{y}(t) = f(y(t)). \tag{1}$$

See [19, 4] and the references therein for examples of such convergence results. In most cases, the limiting drift function f in (1) is assumed to have a Lipschitz property. This strong condition restricts the applicability of these results in many practical cases, in particular, for systems exhibiting thresholds dynamics or with boundary conditions.

The purpose of this paper is to study the limiting behavior of such a system when the dynamics f is not continuous. Let us consider a simple queuing system with one buffer and many processors that can serve one packet per unit of time, on average. If y denotes the number of packets in the queue, then the average decrease of y is one packet per unit of time (under a proper rescaling of time) if the queue is non-empty (*i.e.* $y > 0$) and zero if the queue is empty. This leads to a deterministic limit behavior:

$$\dot{y}(t) = -1 \;\text{ if }\; y(t) > 0 \quad \text{and} \quad \dot{y}(t) = 0 \;\text{ if }\; y(t) = 0. \tag{2}$$

This dynamics is not continuous and therefore non Lipschitz which makes the classical approach inapplicable in that case.

In the case of a non-continuous right-hand side, the differential equation (2) is not well-defined since there exists no function y that is differentiable and that satisfies (2). The proper way to define solutions of (2) is to use *differential inclusions* (DI) instead. Equation (1) is replaced by the following equation

$$\dot{y}(t) \in F(y(t)), \tag{3}$$

where F is a set-valued mapping, defined as the convex hull of the accumulation points of the drift. In the running example, if $y \neq 0$ then $F(y) = \{-1\}$ and $F(0) = \{u : -1 \leq u \leq 0\}$. Of course a differential inclusion problem may (or may not) have multiple solutions. The main result of the paper is that over any finite time interval, the trajectory of the initial system converges to one element in the set of the solutions of the differential inclusion, when the scaling parameter N goes to infinity, (Theorem 1). This result is rather general and does not require any Lipschitz property on the function F. In particular, it shows that when (3) has a unique solution, the behavior of the system converges to it. Moreover, we also show that when F satisfies a one-sided Lipschitz condition (8), we can bound the gap with the limiting dynamics with explicit bounds (Theorem 3).

This generic results is put to practice in several frameworks. First (in Section 3), we show how it can be used to compute the fluid limit of a system and to provide sufficient conditions for the stability of the system. Several famed papers have established that the stability of the fluid limit implies the stability of the initial stochastic system, and this result has been used extensively to prove stability of many distributed stochastic systems where the main difficulty has often been to characterize the fluid limit. Our approach has two advantages: It provides a generic way to construct the limit even with non-continuous drifts, and this construction is explicit enough so that it can be used to give stability conditions in closed form. We illustrate this by establishing the stability condition of opportunistic scheduling policies whose original proofs are rather involved.

The second application framework concerns mean field limits (in Section 4). We show that a stochastic system made of N indistinguishable objects with a non-continuous drift can be seen as a stochastic approximation of a differential inclusion. This result is used to compute the mean field approximation of several systems that could not be studied this way before. We illustrate this by presenting several examples where discontinuities arise naturally: a volunteer computing system with non-smooth constraints, a system with a time-inhomogeneous behavior and a system under a threshold control policy. We also show a system whose mean field approximation satisfies the one-sided Lipschitz condition, corresponding to an energy-saving distributed computing system based on processors with self-adapting frequencies.

2 Stochastic Approximations and differential inclusions

This section presents a generic result that will be used as the methodological basis for the rest of the paper. We show that a Markov chain can be seen as a stochastic approximation with a constant step size of a differential inclusion and we state the main convergence result. A more precise convergence result is established when the limit differential inclusion has the OSL property (§2.1) and the result is also extended to the important case of continuous time Markov chains (§2.2).

Let us consider a *discrete time Markov chain* $Y^N(k)$ with values in $\mathbb{R}^d$. The index N is used to denote some scaling parameter of the chain and will have a clear meaning in the following (for example N could be the number of objects forming the system). The expected difference between $Y^N(k+1)$ and $Y^N(k)$ is called the *drift* of the chain and is denoted g^N:

$$g^N(y) \stackrel{\text{def}}{=} \mathbb{E}\left(Y^N(k+1) - Y^N(k) | Y^N(k) = y\right)$$

The main feature of the chains studied here, concerns their scaling with N. This translates as one essential assumption on the drift: we assume that as N goes to infinity, the drift goes to 0.

The *scaling factor* of the chain is a function of N, denoted $I(N)$ that gives the order of magnitude of the drift with respect to N. More precisely, this means that we assume that there exists a function $I(N)$ such that $\lim_{N\to\infty} I(N) = 0$ and such that for all y:

$$\left\| \frac{g^N(y)}{I(N)} \right\| \leq c(1 + \|y\|),$$

for some constant c. We also denote by $f^N(y)$ the drift rescaled by $I(N)$:

$$f^N(y) \stackrel{\text{def}}{=} \frac{g^N(y)}{I(N)}.$$

Using these definitions, one can write the evolution of the Markov chain $Y^N(k)$ as a *stochastic approximation* algorithm:

$$Y^N(k+1) = Y^N(k) + I(N)\left(f^N(Y^N(k)) + U^N(k+1)\right), \tag{4}$$

where $U^N(k+1) \stackrel{\text{def}}{=} \left(Y^N(k+1) - Y^N(k)\right)/I(N) - f^N(Y^N(k))$ is a process with zero mean[1] that captures the random innovation of the chain between steps k and $k+1$.

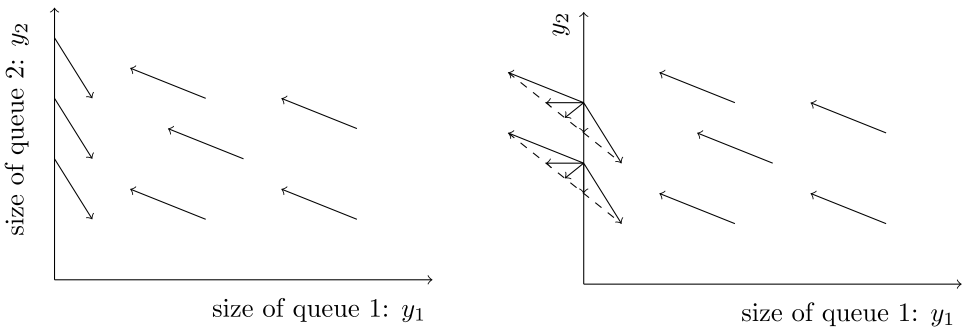
Under mild condition on U^N, when f^N converges uniformly to a Lipschitz-continuous function f, the behavior of $Y^N(t/I(N))$ defined by (4) is known to converge to the solution of an ODE $dy/dt = f(y)$ as N goes to infinity. However, the differential system $dy/dt = f(y)$ cannot be defined properly when f^N is not continuous. To deal with the general case, we introduce a set-valued function F to replace f and the ODE is replaced by the *differential inclusion* $dy/dt \in F(y)$ (see A for a brief introduction on differential inclusions). The set-valued function F associated to the

[1] *i.e.* U^N is a martingale difference sequence: $\mathbb{E}(U^N(k+1) \mid Y^N(k)) = 0$

rescaled drift f^N, at point y, is defined as the convex closure of the set of the accumulation points of $f^N(y^N)$ as N goes to infinity, for all sequences y^N converging to y:

$$F(y) \stackrel{\text{def}}{=} \operatorname{conv}\left(\left\{ \operatorname*{acc}_{N\to\infty} f^N(y^N) \text{ for all sequences } y^N \underset{N\to\infty}{\longrightarrow} y \right\}\right). \tag{5}$$

where $\operatorname*{acc}_{N\to\infty} x^N$ denotes the set of accumulation points of the sequence x^N as N goes to infinity and $\operatorname{conv}(A)$ is the convex hull of set A. The construction of F from f^N is illustrated in Figure 1 in $\mathbb{R}^2$.



(a) Drift of the system of size N: f^N (b) Corresponding set-valued drift F

Figure 1: Example of construction of the set-valued function F from the non-continuous drift f^N. This example is taken from the fluid limit of two queues with priority, developed in Section 3.2 with parameters $\lambda_1 = \lambda_2 = 1$ and $\mu_1 = \mu_2 = 3$. For all N, $f^N(y_1, y_2) = (-2, +1)$ if $y_1 > 0$ and $f(0, y_2) = (+1, 2)$ if $y_2 > 0$. Therefore, $f^N(y)$ is independent of N and is discontinuous in $y_1 = 0$. Since f^N is continuous for $y_1 > 0$, one has $F(y_1, y_2) = \{(-2, +1)\}$ for $y_1 > 0$. When $y_1 = 0$, $F(0, y_2)$ is the convex closure of $(-2, +1)$ and $(+1, -2)$.

We are now ready to state the main theorem of this section. Let us define the continuous function $\bar{Y}^N(t)$ as the piecewise-linear interpolation of $\{Y^N(k)\}_{k\in\mathbb{N}}$ whose time has been accelerated by $1/I(N)$: for all $k \in \mathbb{N}$, $\bar{Y}^N(k \cdot I(N)) = Y^N(k)$ and $\bar{Y}^N$ is linear on $[kI(N), (k+1)I(N)]$. Let us denote by $\mathscr{S}_T(y_0)$ the set of the solutions of the differential inclusion (DI)

$$\dot{y}(t) \in F(y(t)), \quad y(0) = y_0, \tag{6}$$

a solution of the DI (6) being an absolutely continuous function y such that $\dot{y}(t) \in F(y(t))$ almost everywhere.

Starting from y_0, we can show that $\bar{Y}^N(.)$ converges (in probability) to $\mathscr{S}_T(y_0)$. More precisely, the following theorem holds.

Theorem 1. *Let $Y^N()$ be a Markov process on $\mathbb{R}^d$ satisfying (4). Assume that*

- *The drift g^N vanishes with speed $I(N)$: there exists a sequence $I(N)$ and a constant c such that*

$$\lim_{N\to\infty} I(N) = 0 \quad \text{and} \quad \forall y \in \mathbb{R}^d : \left\|f^N(y)\right\| = \left\|\frac{g^N(y)}{I(N)}\right\| \le c(1 + \|y\|).$$

- *U^N is a martingale difference sequence uniformly integrable[2]:*

$$\mathbb{E}\left(U^N(k+1) \mid Y^N(k)\right) = 0 \quad \text{and} \quad \lim_{R\to\infty} \mathbb{E}\left(\left\|U^N(k+1)\right\| \mathbf{1}_{\|U^N(k+1)\ge R\|} \mid Y^N(k)\right) = 0. \tag{7}$$

[2]The uniform integrability can be replaced by a bound on the second moment: $\mathbb{E}(\left\|U^N(k+1)\right\|^2 \mid Y^N(k)) \le b$.

If $Y^N(0) \xrightarrow{\mathcal{P}} y_0$ (convergence in probability), then for all $T > 0$:

$$\inf_{y \in \mathscr{S}_T(y_0)} \sup_{0 \le t \le T} \left\| \bar{Y}^N(t) - y(t) \right\| \xrightarrow{\mathcal{P}} 0.$$

where $\mathscr{S}_T(y_0)$ is the set of solution of the DI (6) and F is defined by (5).

Proof. The proof is given in B.1. □

This theorem shows that if N is large enough, the trajectory of the stochastic system $\bar{Y}^N$ on $T/I(N)$ steps is close to a solution of the differential inclusion (6) over $[0,T]$. This theorem does not assume any regularity condition on the drift function f^N and only requires that the drift vanishes as N grows. It provides a constructive definition of the set-valued drift F. The price to be paid for this generality is that in general a differential inclusion may have multiple solutions. In that case, $\bar{Y}^N$ may converge to any solution of the DI, depending on its random innovations, making this result rather inefficient for performance evaluation. This result is of greater interest if the DI starting from y_0 has a unique solution: $\mathscr{S}_T(y_0) = \{y\}$. In that case, as a direct corollary of the preceding result, $\bar{Y}^N$ converges in probability to y.

Corollary 2. *Under the conditions of Theorem 1 and if the DI (6) has a unique solution y on $[0;T]$:*

$$\sup_{0 \le t \le T} \left\| \bar{Y}^N(t) - y(t) \right\| \xrightarrow{\mathcal{P}} 0.$$

In many cases, the limiting differential inclusion clearly has a unique solution which makes the preceding corollary directly applicable. In particular, this is the case for all the examples presented in this paper except the one of §3.4.

2.1 Bounding the speed of convergence

The main drawback of the previous theorem is that it does not give any insight on the speed of convergence of the stochastic system towards its limit. In fact, without further conditions, the convergence may be arbitrarily slow. This limitation can be overcome when the function F satisfies the one-sided Lipschitz (OSL) condition (8). A set-valued function F is said to be OSL if there exists a constant L such that for all points $y, y' \in \mathbb{R}^d$ and $z \in F(y), z' \in F(y')$:

$$\langle y - y', z - z' \rangle \le L \left\| y - y' \right\|^2. \tag{8}$$

It should be clear that if F is a single-valued Lipschitz function of constant L, F is also OSL with constant L. The name "one sided Lipschitz" comes the fact that a Lipschitz function F would satisfies $-L \left\| y - y' \right\|^2 \le \langle y - y', z - z' \rangle \le L \left\| y - y' \right\|^2$. A simple example of OSL function is $F(y) = -1$ if $y > 0$ and $F(0) = [-1; 0]$. In that case, F is OSL of constant zero.

The OSL condition ensures the uniqueness of the solution. Furthermore, one can get precise bounds on the gap between the stochastic system and its limit in that case. Although this is a natural condition that extends the Lipschitz-continuity to set-valued dynamics, many of the examples presented in this paper do not satisfy the OSL conditions. The applicability of this condition is discussed in Section 4.5 where examples and counter-examples are provided.

Theorem 3. *Let $Y^N(k)$ be a Markov chain on $\mathbb{R}^d$ satisfying (4). Assume that the assumptions of Theorem 1 hold and that*

- *$U^N(k+1)$ is bounded in second moment: $\mathbb{E}\left(\left\| U^N(k+1) \right\|^2 \mid Y^N(k) \right) \le b$.*
- *F is OSL of constant L.*

then the DI (6) has a unique solution y and there exist constants A_T, B_T, C_T depending only on T, L and c and a sequence $J(N)$ with $J(N) \to 0$ such that for all ϵ:

$$\mathcal{P}\left(\sup_{0\leq t\leq T}\left\|Y^N(t)-y(t)\right\| \geq \left\|Y^N(0)-y(0)\right\| e^{LT} + \min\left\{T, \frac{e^{LT}}{\sqrt{2L}}\right\}\sqrt{I(N)A_T{+}J(N)B_T{+}\epsilon C_T}\right) \leq \frac{I(N)bT}{\epsilon^2}.$$

Proof. The proof is given in B.2. □

The sequence $J(N)$ is a bound on the speed of convergence of f^N to F (see Lemma 14). The constants A_T, B_T, C_T and the sequence $J(N)$ are given in B.2. These constants are of a similar order than bounds that can be obtained in the case where f is Lipschitz (see [16]). However, the convergence speed with respect to N is in $O(\sqrt{I(N)})$ (compared with $O(I(N))$ is the Lipschitz case).

2.2 Density Dependent Population Processes

In this section, we show that our results can be adapted to the case of continuous time Markov chains and in particular to the famed model of density dependent population processes of Kurtz [19]. The two theorems 1 and 3 can be transposed in this case.

Let D^N be a continuous time Markov chain on $N^{-1}\mathbb{Z}^d$ ($d \geq 1$) for $N \geq 1$. D^N is called a *density dependent population process* if there exists a set $\mathcal{L} \subset \mathbb{Z}^d$ (with $0 \notin \mathcal{L}$), such that for each $\ell \in \mathcal{L}$ and $y \in N^{-1}\mathbb{Z}^d$, the rate of transition from y to $y + \ell/N$ is $N\beta_\ell(y) \geq 0$, where $\beta_\ell(.)$ does not depend on N. The ith component of $D^N(t)$, $D_i^N(t)$ can be seen as the density of individuals of a population that are in state i, hence the name, and a transition ℓ changes the number of individuals in state i by the quantity ℓ_i.

Let us assume that the transition rate from a state y is bounded: $\sup_{y\in\mathbb{Z}^d}\sum_{\ell\in\mathcal{L}}\beta_\ell(y) = \tau < \infty$ and that $\sum_{\ell\in\mathcal{L}}\|\ell\|\sup_y \beta_\ell(y) < \infty$. The expectation of the change of the system during a small interval dt is $f(y)dt$ where $f(y)$ is the drift of the system, defined by $f(y) = \sum_{\ell\in\mathcal{L}}\beta_\ell(y)\ell$. If f is Lipschitz, it is well-known that $D^N(.)$ goes to the solution of the ODE $\dot{y} = f(y)$ as N grows [19]. Using uniformization of the Markov chain and Theorems 1 and 3, we show that this convergence still holds for general drifts, replacing f by its set-valued counterpart F, defined in (5).

Theorem 4. *Assume that $\sup_{y\in\mathbb{Z}^d}\sum_{\ell\in\mathcal{L}}\beta_\ell(y) < \infty$ and that $\sum_{\ell\in\mathcal{L}}\|\ell\|\sup_y\beta_\ell(y) < \infty$. Let f be defined by $f(y) = \sum_\ell \ell\beta_\ell(y)$ and assume that f satisfies $\|f(y)\| \leq c(1 + \|y\|)$, then for all $T > 0$:*

$$\inf_{d\in\mathscr{S}_T(y_0)}\ \sup_{0\leq t\leq T}\left\|D^N(t) - d(t)\right\| \xrightarrow{\mathcal{P}} 0,$$

where $\mathscr{S}_T(y_0)$ is the solution set of the DI (6) starting in y_0.

Moreover, if F is OSL of constant L and $\sup_y \sum_{\ell\in\mathcal{L}}\|\ell\|^2 \sup_y \beta_\ell(y) \leq b$, then the differential inclusion (6) has a unique solution d and there exist constants A_T, B_T, C_T depending only on T, L and c and a sequence $J(N)$ with $J(N) \to 0$ such that for all ϵ:

$$\mathcal{P}\left(\sup_{0\leq t\leq T}\left\|Y^N(t){-}y(t)\right\| \geq \left\|Y^N(0){-}y(0)\right\| e^{LT} + \min\left\{T, \frac{e^{LT}}{\sqrt{2L}}\right\}\sqrt{\frac{A_T}{N}{+}J(N)B_T{+}\epsilon C'_T}\right) \leq \frac{b + 1/\tau}{N\epsilon^2}T.$$

Proof. Using uniformization of the Markov chain D, we construct a discrete time Markov chain Y^N that satisfies the assumptions of Theorem 1. A detailed proof is given in B.3. □

The constants A_T, B_T and the sequence $J(N)$ are the same as in Theorem 3. The constant C'_T is given in B.3.

3 Application 1: Fluid limits and Stability Issues

Fluid limits have become an important tool for studying stochastic stability of queuing networks. For a large class of queuing networks, when the initial state of the system is rescaled by a factor $N \to \infty$ and the time is accelerated by the same factor N, the system is shown to satisfy a system of deterministic equations, called the *fluid limit model*. The link between stability of fluid limit and stochastic stability has a long history of research. The results obtained can be mainly decomposed in two types. On the one hand, many people have studied specific queuing models with general arrival process and service rate and constructed explicitly the fluid model equations corresponding to these systems, see [10, 9] and the references therein. More recently, structural properties based on the drift have been studied but only for continuous drift f [14]. Theorem 6 makes the link between the two approaches by showing that generic results can be obtained even for non-continuous dynamics.

In Section 3.1, we define the fluid limit model of a system and show that the fluid limit model satisfies a differential inclusion. We further show that stability of this differential inclusion implies stability of the stochastic system. These results are illustrated in Sections 3.2 and Section 3.3 to study the stability of opportunistic routing policies. We conclude on some limitations of the approach in Section 3.4.

3.1 Definition of fluid limits and stability

Let $X(.)$ be a discrete time[3] Markov chain in $\mathbb{R}^d$. For any $y_0 \in \mathbb{R}^d$ and $N > 0$, we consider the rescaled process $\bar{Y}^N$ for which the state has been scaled by a factor $1/N$ and the time accelerated by N:

$$\bar{Y}^N(t) = \frac{1}{N} X\left(\lfloor N \cdot t \rfloor\right) \qquad\qquad \bar{Y}^N(0) = \frac{1}{N} X(0) = y_0.$$

The next result shows that fluid limits are solutions of a differential inclusion that can be constructed directly from the drift.

We say that a set E of functions from $\mathbb{R}^+$ to $\mathbb{R}^d$ *contains the fluid limits* of Y^N if for all $T > 0$:

$$\inf_{y \in E} \sup_{0 \le t \le T} \left\| \bar{Y}^N(t) - y(t) \right\| \xrightarrow{\mathcal{P}} 0. \tag{9}$$

Such a set E is a super-set of the limiting behaviors of X^N/N when the initial condition and the time are rescaled by a factor N. This definition is in accordance with classical definitions of fluid limits in the literature. For example in [14], a fluid limit is defined as a weak limit of the process $\bar{Y^N}$. The support of such a measure contains the fluid limits according to our definition. The definition of fluid limits in [9, 10] is less clear since it only concerns queuing networks. Translated in our framework, it corresponds to an accumulation point of $\bar{Y}^N$ conditioned on the fact that the noise $\sum_{i=1}^{T/I(N)} U^N(i)$ converges to zero as N grows.

The following theorem shows that the differential inclusion corresponding to (10) describes a super-set of the limiting behavior of $\bar{Y}^N$.

Proposition 5. *Assume that the drift $f(x) = \mathbb{E}\left(X(t+1) - X(t) \mid X(t) = x\right)$ is bounded and that $\lim_{R\to\infty} \mathbb{E}(\|X(t+1) - X(t)\| \mathbf{1}_{\|X(t+1)-X(t)\| \ge R} \mid X(t) = x) = 0$. Let F be a set-valued function defined as*

$$F(y) \stackrel{\text{def}}{=} \operatorname{conv}\left(\operatorname*{acc}_{N\to\infty} f(N \cdot y^N) \text{ with } \lim_{N\to\infty} y^N = y \right). \tag{10}$$

Then, the set of solution $\mathscr{S}_T(y_0)$ of the differential inclusion $\dot{y} \in F(y)$ starting in x contains the fluid limits of Y^N (in the sense of (9)).

[3] For readability, we restrict our presentation to discrete time models. However, these results can be extended directly to continuous time Markov chains using uniformization as in Section 2.2.

Proof. This result is a direct consequence of Theorem 1. To fit with the framework of Theorem 1, let us call $f^N(y) \stackrel{\text{def}}{=} f(Ny)$. For all $t \in \frac{1}{N} \cdot \mathbb{N}$, $\bar{Y}^N(t+\frac{1}{N})$ satisfies $\bar{Y}^N(t+\frac{1}{N}) = \bar{Y}(t) + \frac{1}{N}\left(f^N(\bar{Y}^N(t)) + U(t+\frac{1}{N})\right)$ with $\mathbb{E}\left(U(t+\frac{1}{N}) \mid X(t)\right) = 0$. The function F defined by Equation (10) is the same as with Equation (5). As f is bounded, f^N is bounded. Moreover, the assumption implies that $X(t+1) - X(t)$ is uniformly integrable. This shows that Y^N satisfies assumptions of Theorem 1. □

This theorem does not require any continuity assumption on f and provides a characterization of the fluid limit in term of differential inclusions. This theorem can be viewed as a generalization of Proposition 1.5 of [14] that assumes that f^N goes to some continuous function f. If the differential equation has a unique solution y on $[0;T]$, then y is called the fluid limit of Y^N and Proposition 5 implies that $\bar{Y}^N$ converges to y:

$$\sup_{0\le t\le T} \left\|\bar{Y}^N(t) - y(t)\right\| \xrightarrow{\mathcal{P}} 0.$$

In turn, this result can be viewed as a generalization of Theorem 1.6 of [14].

There are several ways to define the stability of the fluid limit model. We follow the definition of [25, 14] and say that the differential inclusion $\dot{y} \in F(y)$ is stable if there exists $T > 0$ and $\rho < 1$ such that:

$$\text{For any } y \text{ solution of } \dot{y} \in F(y) \text{ with } \|y(0)\| = 1: \quad \sup_{0\le t\le T} \|y(t)\| \le \rho < 1. \tag{11}$$

As expressed by the next proposition, stability of the fluid limit in the sense of (11) implies the stability of the stochastic model. Before stating the main theorem, we recall the definitions of ϕ-irreducibility and petite set that are useful to show stability of a Markovian process on a non-countable set. We refer to [21] for a more detailed presentation of these notions.

A discrete time Markov chain X on $\mathbb{R}^d$ is said to be ϕ-irreducible if there exists a σ-finite measure ϕ such that for any set $A \subset \mathbb{R}^d$, $\phi(A) > 0$ implies $\sum_{k\ge 0} \mathcal{P}(X(k) \in A \mid X(0) = x) > 0$. Moreover, a set $A \subset \mathbb{R}^d$ is said to be *petite* if for some fixed probability measure a on $\mathbb{Z}^+$ and some non-trivial measure ν on $\mathbb{R}^d$, $\nu(B) \le \sum_{k\ge 0} \mathcal{P}(X(k) \in B \mid X(0) = x)a(k)$ for all $x \in A$ and $B \subset \mathbb{R}^d$. Finally, X is said to be positive Harris recurrent if X has a unique stationary probability distribution π and $P^k(x,.)$ converges to π. In particular, if the state space of X is included in $\mathbb{Z}^d$ and if X is irreducible and aperiodic, then X is ψ-irreducible and all compact sets are petite.

Theorem 6. *Assume that X is an aperiodic, ψ-irreducible Markov chain such that all compact sets are petite. Assume that the drift $f(x) = \mathbb{E}\left(X(t+1) - X(t) \mid X(t) = x\right)$ is bounded and that $\lim_{R\to\infty} \mathbb{E}(\|X(t+1) - X(t)\| \mathbf{1}_{X(t+1)-X(t)\ge R} \mid X(t) = x) = 0$ and let F be defined as in Equation (10):*

$$F(y) \stackrel{\text{def}}{=} \operatorname{conv}\left(\operatorname*{acc}_{N\to\infty} f(N\cdot y^N) \quad \textit{for all } \{y^N\}_{N\in\mathbb{N}} \textit{ s.t. } \lim_{N\to\infty} y^N = y\right).$$

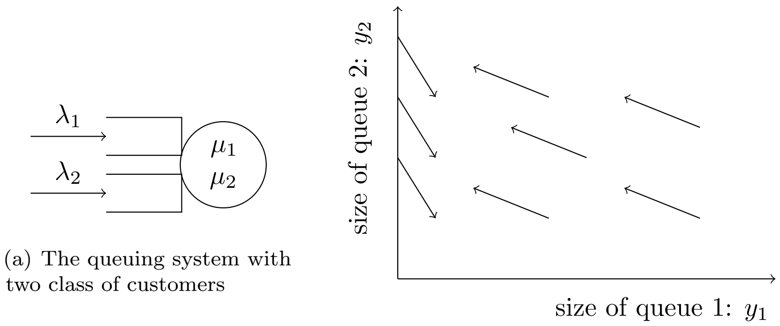
If the differential equation $\dot{y} \in F(y)$ is stable in the sense of Equation (11), then X is positive Harris recurrent.

Proof. Theorem 1.4 of [14] shows that if all function y of a set containing the fluid limits of $\bar{Y}^N$ are stable in the sense $\sup_{0\le t\le T} \|y(t)\| \le \rho$, then the process X is Harris recurrent. Proposition 5 shows that the solutions of the differential inclusion $\dot{y} \in F(y)$ contains the fluid limits. Therefore, the stability of the DI given by Equation (11) implies the Harris recurrence of X. □

This theorem shows that the constructive definition of F allows one to obtain sufficient conditions for stability. In the following, we illustrate this results on some examples. When the DI has a unique solution, the stability conditions obtained by Theorem 6 are generally also necessary. However, when the DI has multiple solutions, the DI may describe a super-set of the fluid-limits and stability conditions obtained may be too strong, as in the example of §3.4.

3.2 Fluid limit of a system of parallel queues with static priority.

We consider a time-slotted model of a queueing system composed of one server serving multiple classes of users. There are K classes of customers. At time step t, $A_k(t)$ customers of class k arrive. A_k are *i.i.d.* with $\mathbb{E}(A_k) = \lambda_k$. Let $X_k(t)$ be the number of customers of class k in the system at time t. For $k < k'$, customers of class k have preemptive priority over customers of class k'. When the system serves a customers of class k, it leaves the system in the same time slot with probability μ_k. This means that if there are one or more user of class 1 present in the system, a user of class 1 leaves the system with probability μ_1 and no other user departs in the same time slot. When there are no customer of class $1 \dots k-1$ and one or more users of class k, a user of class k departs with probability μ_k. This model with two classes of customers is depicted in Figure 2(a) where each queue corresponds to a class of customers.



(a) The queuing system with two class of customers

(b) Drift of the system

Figure 2: System and the corresponding drift

The drift of the system can be computed by:

$$f(x) = \begin{cases} (\lambda_1 - \mu_1, \lambda_2, \dots, \lambda_k) & \text{if } x_1 > 0 \\ (\lambda_1, \dots, \lambda_{k-1}, \lambda_k - \mu_k, \lambda_{k+1}, \dots, \lambda_k) & \text{if } x_1 = \dots = x_{k-1} = 0 \text{ and } x_k > 0. \end{cases}$$

For example, when $\lambda_1 = \lambda_2 = 1$ and $\mu_1 = \mu_2 = 3$, the drift is $f(x_1, x_2) = (-2, +1)$ if $x_1 > 0$ and $(+1, -2)$ if $x_1 = 0$ and $x_2 > 0$. This drift is depicted in Figure 2(b). As shown on Figure 2(b), the drift is constant for all $x_1 > 0$ but is discontinuous for $x_1 = 0$. Because of this discontinuity, there is no function x derivable almost everywhere such that $\dot{x}(t) = f(x)$: the axis $x_1 = 0$ both attracts the trajectories from $x_1 > 0$ and repulses the trajectories starting from $x_1 = 0$.

It should be clear that the model satisfies all the assumptions of Proposition 5. Let us compute the set-valued function F corresponding to the drift f defined as in Equation (10). For all k, let us define $u_k \stackrel{\text{def}}{=} (\lambda_1, \dots, \lambda_{k-1}, \lambda_k - \mu_k, \lambda_{k+1}, \dots, \lambda_k)$. When $x_1 > 0$, $f(x)$ is locally continuous and $F(x)$ is single-valued and $F(x) = \{u_1\}$. Because of the discontinuity in $x_1 = 0$, when $x_1 = 0$ and $x_2 > 0$, $F(x)$ is the convex hull of the vectors $\{u_1, u_2\}$. In the two classes case, this convex hull corresponds to the dashed line of Figure 3(a). Thus, the drift has the expression:

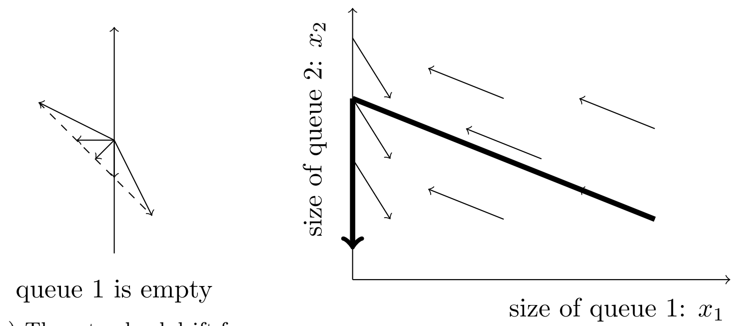
$$F(x) = \begin{cases} u_1 & \text{if } x_1 > 0 \\ \mathrm{conv}(u_1, \dots, u_k) & \text{if } x_1 = \dots = x_{k-1} = 0, x_k > 0. \end{cases} \tag{12}$$

Let us assume that $\sum_k \lambda_k/\mu_k < 1$ and let us show that the differential equation associated to F has a unique solution when starting from $x = (x_1 \dots x_K)$. Such a x is an absolutely continuous function such that the derivative of x at time t is $\dot{x}(t) \in F(x(t))$ almost everywhere. Assume that $x_1 > 0$. As long as $x_1 > 0$, the drift is u_1, then after a time $T_1 = x_1/(\mu_1 - \lambda_1)$, $x_1(T_1) = 0$ and $x_2(T_1) = x_2 + \lambda_2 T_1$. Since $f_1(x) < 0$ for $x_1 > 0$, this implies that $x_1(t) = 0$ for all $t \geq T_1$. Moreover, as long as $x_2 > 0$, the derivative of x at time t, $\dot{x}(t)$, is in the convex hull $\mathrm{conv}(u_1, u_2)$. These two conditions implies that there exists $\theta \in [0; 1]$ such that $\dot{x}(t) = \theta u_1 + (1-\theta) u_2$ and $\dot{x}_1(t) = 0$. This implies that $\theta(\lambda_1-\mu_1)+(1-\theta)\lambda_1$ and therefore $\theta=\lambda_1/\mu_1$. This shows that as long as $x_2(t)>0$, $\dot{x}(t)$ satisfies:

$$\dot{x}(t)=\frac{\lambda_1}{\mu_1}u_1+\left(1-\frac{\lambda_1}{\mu_1}\right)u_2=\left(0,\lambda_2-\left(1-\frac{\lambda_1}{\mu_1}\right)\mu_2,\lambda_3,\dots,\lambda_K\right).$$

The condition $\sum_k \lambda_k/\mu_k<1$ implies that $T_2<\infty$. Moreover, a similar computation shows that for all $k\leq$, there exists $T_k<\infty$ such that for all $t\in[T_{k-1},T_k]$, the derivative of x satisfies

$$\dot{x}(t)=\left(0,\dots,0,\lambda_k-\left(1-\sum_{i<K}\frac{\lambda_i}{\mu_i}\right)\mu_k,\lambda_{k+1},\dots,\lambda_K\right).$$



(a) The set-valued drift for $x_1=0$ is the set of vectors displayed by the dashed line.

(b) Unique solution of the differential inclusion.

Figure 3: Convex hull of the drift at $C_2=0$ and unique solution of the fluid limit.

Therefore, the differential inclusion corresponding to this system has a unique solution, depicted in Figure 3(b). Moreover, the solution of the differential inclusion goes to 0 in finite time and this system satisfies the assumptions of Theorem 6. This shows that the system is stable. Although this result can be shown directly, our framework provide an easy way to construct the fluid limit and prove the convergence of the original process.

3.3 Stability of opportunistic routing policies in wireless networks

In this section, we show how Proposition 5 and Theorem 6 can be used to characterize the stability of opportunistic routing policies in wireless setting with flow-level dynamics. The stability of such policies have been first studied in [2, 26]. Because of the discontinuity of the dynamic, generic approaches such as the ones introduced in [14] fail and ad-hoc methods have been developed. Our framework shows that a systematic generic approach can also be used in that case to compute easily the limiting dynamic and show stability.

We consider the model studied in [2]. Transmissions occur in a time-slotted channel. There are K classes of users. At time slot t, $A_k(t)$ new users of type k arrive in the system. The $A_k(t)$ are *i.i.d.* with $\mathbb{E}(A_k(t))=\lambda_k$, $\mathbb{E}(A_k^2(t))<\infty$. The condition of the channel is varying over time and at time slot t, a user of type k has condition $i\in\{1\dots I_k\}$ with probability $q_{k,i}\neq 0$. The channel condition of a user is independent of other users and of channel history. At each time slot, a server observe the channel condition of all user and chooses to serve one user. If this user is of type k and has a channel condition i, this user leaves the system with probability $\mu_{k,i}$. With no loss of generality, we may assume $\mu_{k,1}>\mu_{k,2}\dots$ The quantity $\mu_{k,i}$ represents the rate at which at user k with condition i is served. Under its best condition, a user of type k is served with rate $\mu_k^{\max}\stackrel{\text{def}}{=}\mu_{k,1}$.

The design of an efficient policy for scheduling the users have received a considerable attention in the past (see [2, 26] and the reference therein). An important feature of an efficient policy is to design a policy that stabilizes the system, *i.e.* such that the number of users in each class is a positive recurrent Markov chain. We next show how our framework can be used to prove the following results (originally proved in [2, 26] by ad-hoc arguments).

Proposition 7 (Theorem 5.2 of [2]). *There exists a scheduling policy that stabilizes the system if and only if*

$$\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k^{\max}} < 1. \tag{13}$$

Proof. It should be clear that (13) is a necessary condition for stability. Therefore, we only show that (13) is a sufficient condition and we assume that (13) holds.

Let us consider the following policy (called "Best Rate" policy in [2]):

- if there are n users $u_1 \dots u_n$ of class $k_1 \leq \dots \leq k_n$ that are in their best channel condition, serve the user with the smallest class (*i.e.* user u_1).
- if there are no user in their best channel condition, serve a user at random.

For all k, let $X_k(t)$ be the number of users in class k at time t when applying this policy. Since the channel conditions are independent, the process $X(.)$ is a Markov chain. Let us compute the drift $f(x) = \mathbb{E}(X(t+1) - X(t) \mid X(t) = x)$ of the system.

If the server is serving a user of type k which is in its best state, the drift of the system is $u_k = (\lambda_1, \dots, \lambda_{k-1}, \lambda_k - \mu_k^{\max}, \lambda_{k+1}, \dots, \lambda_K)$. Let $p_{i,k}$ the probability for a user of type k in state i to be served. In state $x = (x_1, \dots, x_k)$, the probability that no users are in their best state is $(1-q_{1,1})^{x_1} \dots (1-q_{K,1})^{x_K}$ which goes to 0 as $\|x\|$ goes to infinity. Therefore, as N goes to infinity, the drift of the system starting in $x.N$ is inside the convex closure of $u_1 \dots u_K$. Moreover, using the fact that a class k has priority over class $k' > k$, the set-valued drift defined in (10) is defined by:

$$F(x) = \begin{cases} u_1 & \text{if } x_1 > 0 \\ \operatorname{conv}(u_1, \dots, u_k) & \text{if } x_1 = \dots = x_{k-1} = 0, x_k > 0. \end{cases}$$

These equations are the same as (12). Therefore, the differential equation has a unique solution that goes to 0 in finite time. This shows that (13) implies the stability of the stochastic system. □

3.4 Limitations of the differential inclusion approach

In many cases, the differential inclusion approach allows one to characterize exactly what will be the fluid limits of a stochastic systems. This is the case in particular when the differential equation has a unique solution. However, in other cases, the construction of the set-valued function F by convexification leads to a super-set of the fluid limits, so that the stability conditions provided by this approach are only sufficient but not necessary.

Let us consider the example of three weakly coupled queues, presented in §5.1 of [7]. The system is composed of 3 queues. Customers arrive at queue i with rate λ_i. If x_i, x_j, x_k are the numbers of customers present in queues $i \neq j \neq k \in \{1, 2, 3\}$, a customer of queue i is served with rate $\phi_i(x)$:

$$\phi_i(x) = \begin{cases} a_i & \text{if } x_j = x_k = 0 \\ a_{ij} & \text{if } x_j > 0, x_k = 0 \\ 1 & \text{if } x_j > 0, x_k > 0, \end{cases}$$

where $a_i \leq a_{ij} \leq 1$.

The drift of the system is $f(x) = (\lambda_1 - \phi_1(x), \lambda_2 - \phi_2(x), \lambda_3 - \phi_3(x))$. We use the same method as in §3.2 to compute the solutions of the corresponding differential inclusion starting from a point

(x_1, x_2, x_3) with $x_1, x_2, x_3 > 0$. Let $x(.)$ be a solution of the differential inclusion. For t small enough, the derivative of x is $\dot{x}(t) = (\lambda_1 - 1, \lambda_2 - 1, \lambda_3 - 1)$. Let us assume (w.l.o.g.) that

$$\lambda_1 < 1, \tag{14}$$

and that $x_1(t)$ reaches 0 before $x_2(t)$ and $x_3(t)$.

Let T_1 be the time when $x_1(t)$ reaches 0. For $t > T_1$ and as long as $x_2(t) > 0$ and $x_3(t) > 0$, using the convex closure of the drift of the system implies that there exists $0 \le \theta \le 1$ such that $\dot{x}(t) = (\lambda_1 - \theta, \lambda_2 - \theta - (1-\theta)a_{23}, \lambda_3 - \theta - (1-\theta)a_{32})$. Since $\dot{x_1}(t) = 0$ for $t > T_1$ then $\theta = \lambda_1$ and the drift is:

$$\dot{x}(t) = (0, \lambda_2 - \lambda_1 - a_{23}(1-\lambda_1), \lambda_3 - \lambda_1 - a_{32}(1-\lambda_1)).$$

One of the two components of this drift has to be negative for the system to be stable. Thus, we may assume w.l.o.g. that

$$\lambda_2 < \lambda_1 + a_{23}(1-\lambda_1), \tag{15}$$

and that $x_2(t)$ reaches 0 before $x_3(t)$.

When $x_2(t)$ reaches 0, F is the convex closure of 4 vectors u_0, u_1, u_2, u_{12} corresponding respectively to the drift when $(x_1 > 0, x_2 > 0)$, $(x_1 = 0, x_2 > 0)$, $(x_1 > 0, x_2 = 0)$, and $(x_1 = x_2 = 0)$. Using the fact that the actual drift is in F and that $\dot{x_1} = \dot{x_2} = 0$, there exists $\theta_0, \theta_1, \theta_2 \in [0;1]$ with $\theta_1 + \theta_2 + \theta_0 \le 1$ such that:

$$0 = \lambda_1 - \theta_0 - a_{13}\theta_2 \tag{16}$$
$$0 = \lambda_2 - \theta_0 - a_{23}\theta_1 \tag{17}$$
$$\dot{x}_3(t) = \lambda_3 - \theta_0 - a_{31}\theta_2 - a_{32}\theta_1 - (1 - \theta_0 - \theta_1 - \theta_2)a_3. \tag{18}$$

In general, there is not a unique triplet $(\theta_0, \theta_1, \theta_2)$ such that (16–17) are satisfied. If for all $(\theta_0, \theta_1, \theta_2)$ such that (16–17) are verified, (18) is negative, then the system is stable. Conversely, if for all $(\theta_0, \theta_1, \theta_2)$ satisfying (16–17), (18) is positive, then the fluid limit is unstable. However, in general, one cannot compute the stability condition of the fluid system only using Equations (16–17–18) since the sign of (18) may depend on $(\theta_0, \theta_1, \theta_2)$.

In [7], the exact stability conditions are given. Equations (14-15) are similar while the conditions on θ (16–17) are expressed as a function of the stationary distribution of X_1, X_2 conditioned by the fact that $X_3 > 0$. However, the proof of this result is much more involved and these equations can not be solved in close form whereas the present approach gives upper bounds in closed-form. In [17], an simpler approximation method is also applied to the same problem. However, this leads to looser bounds than ours.

4 Application 2: Mean Field Limits

In this section, we show how our framework allows one to extend the expressive power of mean field model to study model with discontinuous dynamics. Convergence results of mean field model have received a considerable amount of attention in the past. However, except in particular cases where ad-hoc proof are presented, convergence results of mean field models in the literature [19, 4, 11] always assume the Lipschitz-continuity of the drift. Our framework shows that these models can be extended to characterize the limiting behavior of system with discontinuous dynamics and therefore simplify their study.

This section is organized as follows. We first present the model and show how our framework can be adapted in §4.1. Then we present several examples that illustrate the power of our model. The first example shows our this results allows one to show how to handle discontinuities that arise because of buffer. Then we show how to handle discontinuity caused by the time in §4.3 or by a control policy in §4.4. Finally, we study the OSL condition in §4.5.

4.1 Mean field model and its convergence

We consider of system composed of N *objects* evolving in a finite state space $\mathcal{S} = \{1 \dots S\}$. Time is discrete and the state of object n at time step k is denoted $X_n^N(k)$. The objects all evolve in a common environment, called the *context*. The context can represent a shared resource, like a buffer in which packets are stored. The state of the context at time step k is a vector of dimension d and is denoted by $C^N(k) \in \mathbb{R}^d$. The state of the global system at time k is $(X_1^N(k) \dots X_N^N(k), C^N(k))$. We denote by $M^N(k)$ the empirical measure associated with the N objects. Since an object has S possible states, $M^N(k)$ can be represented by a vector with S components, its ith component being the proportion of objects in state i:

$$M_i^N(k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{X_n^N(k)=i}.$$

The system $(M^N(k), C^N(k))_k$ is assumed to be a Markov chain. In particular, this is true if the objects all have a Markovian dynamics, if the evolution of the context is a deterministic function of itself and of M^N and if the law of the whole system is invariant by any permutation of the N objects. The state space of this Markov chain is included in $\mathbb{R}^{S+d}$. To fit with the notations introduced in Section 2, we let $Y^N(k) \stackrel{\text{def}}{=} (M^N(k), C^N(k))$. We denote by f^N the rescaled drift of Y^N and by F the convex hull of its accumulation points.

There are multiple situations in which assumptions of Theorem 1 are true. To make things more concrete, here are sufficient conditions under which the conditions of Theorem 1 hold:

- The number of objects that perform a transition at each time slot is bounded by a deterministic constant c. In that case, the intensity is $I(N) = 1/N$.
- The evolution of the context is deterministic and there exists a constant c_1 such that for all y, $f^N(y) \leq c_1$.
- $M^N(0), C^N(0) \xrightarrow{\mathcal{P}} m_0, c_0$.

These two assumptions imply that Theorem 1 holds and the behavior of M^N, C^N can be approximated by the solutions of the differential inclusion as N grows: for all $T > 0$:

$$\inf_{(m,c)\in\mathscr{S}_T} \sup_{0\leq t\leq T} \left\| \left(\bar{M}^N(t), \bar{C}^N(t)\right) - (m(t), c(t)) \right\| \xrightarrow{\mathcal{P}} 0,$$

where $\mathscr{S}_T$ denotes the set of solution of the DI $\dot{m,c} \in F(m,c)$ with $m(0) = m_0$ and $c(0) = c_0$.

4.2 Volunteer Computing

Here, we consider a model of a volunteer computing system, such as BOINC `http://boinc.berkeley.edu/`. The system is made of a single buffer and N desktop machines, offered by their owners (volunteers), that serve the packets of this buffer. However, as soon the owner of a processor wants to use it, she preempts it and the processor becomes unavailable for the computing system. As for the incoming packets, they are assumed to arrive in the buffer according to a Poisson process at rate λ. These kinds of systems are often called push/pull models: The distributed applications *push* jobs to a central server that stores them in a buffer and whenever a processor becomes available, it *pulls* a job from the buffer and executes it.

Such systems fit our density dependent population process framework. The context $C(t)$ represents the size of the buffer while the N objects represent both the applications sending jobs and the hosts executing them. The state of a host is its availability and its idleness (whether it is executing a job or not). The non-smooth part of the dynamics comes from the buffer size. When $C(t) > 0$, if a host asks for a job, it gets it with probability one while when $C(t) = 0$, a host asking for a job will get nothing. In that case, one can show that this dynamics satisfies the conditions of Theorem 4 that can be used to study the limiting behavior of the system when the number of hosts and applications grows.

In the simplest case, the intensity of the system is $I(N) = 1/N$ and an application sends a job to the system at rate λ while jobs are completed at rate μ by each server. To represent the communication delays, every host gets jobs at rate γ. It becomes unavailable with rate p_u, and available with rate p_a if $C(t) > 0$ and 0 otherwise. If b, a, u denote respectively the proportion of busy, available and unavailable hosts, the limiting system is described by a DI:

$$\begin{aligned} \dot{b}(t) &= -\mu b(t) + \gamma a(t)\mathbf{1}_{C(t)>0} \\ \dot{a}(t) &= \mu(t)b(t) + p_a u(t) - p_a a(t) - \gamma a(t)\mathbf{1}_{C(t)>0} \\ \dot{u}(t) &= -p_a u(t) + p_u a(t) \\ \dot{C}(t) &= -\gamma a(t)\mathbf{1}_{C(t)>0} + \lambda \mathbf{1}_{C(t)<C_{\max}}. \end{aligned}$$

The formal DI is obtained by replacing $a(t)\mathbf{1}_{C(t)>0}$ by the singleton $\{\gamma a(t)\}$ if $C(t) > 0$ and the interval $[0; \gamma a(t)]$ when $C(t) = 0$.

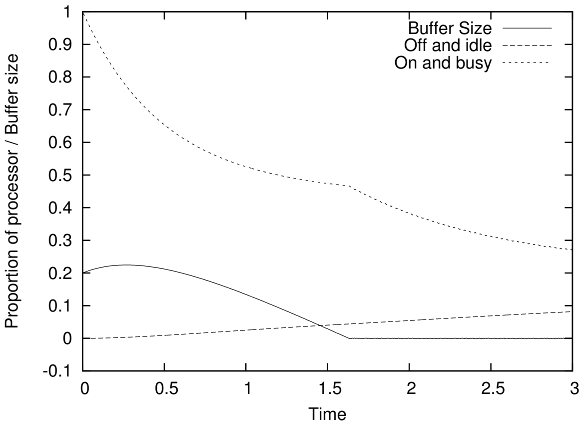


Figure 4: Limit dynamics of a volunteer computing system. A non-differentiable point occurs when the buffer becomes empty.

At time $t = 0$, we consider that the size of the buffer is $C(0) = .2$ and that all processors are available and are serving a job. The behavior of the system is represented in Figure 4. One can see that there is a point of non-differentiability in the behavior of the system hen the size of the buffer reaches 0.

4.3 Volunteer Computing: Day and Night Scenario

The next example shows that our DI approach can be used for systems whose dynamics is not time-homogeneous.

We consider a similar model as the previous one except that the processors follow a day and night behavior. We consider that some of the processors are turned off at night. Therefore, the availability is larger during the day (between 7am and 5pm) than at night.

The limiting dynamics can be represented by a differential inclusion that depends (not continuously) on time:

$$\frac{\partial y(t)}{\partial t} \in F(y(t), t). \tag{19}$$

In the general theorems, we assumed that the function F was time-homogeneous. There are three ways to tackle the case at hand. The first one is to adapt the proofs to the time dependent case. Another idea that can used here is the fact that for any finite interval of time, there is only a finite number of discontinuity points (7am, 5pm,...), and to apply the convergence results on a first sub-interval $[0; 7\text{am}]$. Using the fact that $Y^N(7) \to y(7)$, the convergence holds on $[7\text{am}, 5\text{pm}]$, and so forth. Yet another solution (adopted in the following) is to write $Z(t) = (Y(t), t)$. Then the differential inclusion (19) can be written:

$$\frac{\partial z(t)}{\partial t} \in (F(z(t)), 1)\,. \tag{20}$$

The fact that F is not continuous in t (and therefore in $z(t)$) is not a problem in our differential inclusion setting. It should be clear that in that case there is still a unique solution to the differential inclusion (20).

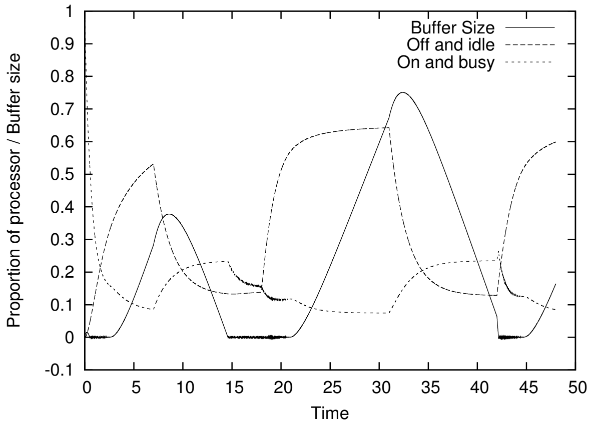


Figure 5: Limit dynamics for volunteer computing under day and night availability.

In Figure 5 we can observe two kinds of non-differentiable points. The first ones are the points representing the change from day time to night time. The other ones occur when the buffer becomes empty. The small oscillations of the buffer size around 0 and of the proportion of busy processors are just numerical integration artefacts (typical of numerical integration of differential inclusions). In both cases the exact solutions do not exhibit these jumps.

4.4 Join the Shortest Queue in Volunteer Computing

This example shows an application to a controlled system, exhibiting a threshold behavior.
This example is similar to the one of Section 4.2 but with two identical time-homogeneous volunteer systems. Each time a packet arrives, it is routed to the system with the smallest number of packets. Here, the routing of packets introduces a new cause of non-smoothness: there is a threshold in the dynamics of the system when both backlogs are equal.

Figure 6 shows the behavior of the limit differential inclusion. Once again, the limit behavior is unique once the initial condition is given.

As expected, new non-differential points appear when both buffers are equal.

4.5 OSL dynamics: energy-aware distributed computing system

The previous examples have dynamics for which the drift function did not satisfy an OSL condition. However, OSL conditions are commonly assumed in the non-smooth system literature [8, 20]. In this section, we recall sufficient condition for a dynamical system to be OSL and provide some examples of dynamics that are OSL. Recall that a set-valued function F is OSL with constant $L \geq 0$ if for all $y, \bar{y} \in \mathbb{R}^d$ and $u \in F(y), \bar{u} \in F(\bar{y})$: $\langle y - \bar{y}, u - \bar{u}\rangle \leq L\left\|y - \bar{y}\right\|^2$.

Let us first examine the example of Section 4.2. Let $y = (b, a, u, C)$ and $\bar{y} = (\bar{b}, a, u, 0)$, then $\langle y - \bar{y}, f(y) - f(\bar{(y)})\rangle = -\mu|b - \bar{b}|^2 + \gamma a(b - \bar{b}) - \gamma a$. If f was OSL, this would be less than $L\left\|y - \bar{y}\right\|^2$. However, when $b - \bar{b}$ is small enough and positive, this expression is of order $\gamma a(b - \bar{b})$ which is greater than $L\|y - \bar{y}\|^2 = L(|b - \bar{b}|^2 + C^2)$.

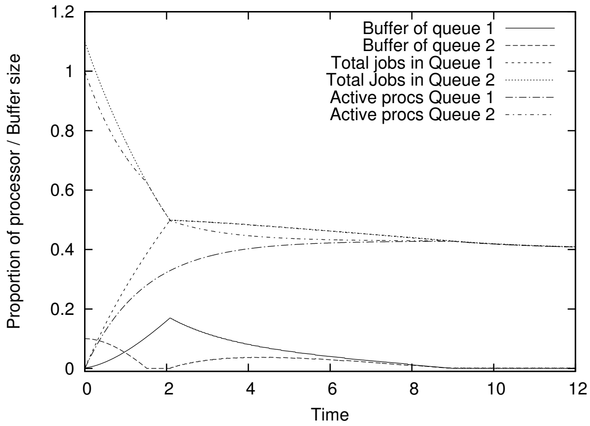


Figure 6: Join the shortest queue for volunteer computing

In fact, there are two types of non-smoothness in these equations. The first one is that the dynamics of C depends not continuously on the dynamics of C but in a OSL way (see below). The second type of discontinuity is that the dynamics of b depends non-smoothly on C. This latter discontinuity leads to a term of order $(b - \bar{b})$ which is greater than $L \left\|b - \bar{b}\right\|^2$ whenever $b - \bar{b}$ is small enough. This is a general problem for the applicability of OSL: whenever the derivative of one coordinate depends non-smoothly on a other coordinate, the dynamics is never OSL. This is actually the case in all the examples of Sections 3 to 4.4.

4.5.1 Sufficient conditions to prove OSL

The next lemma links conditions on the drift function f to its set-valued counterpart.

Lemma 8. *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a single valued function and F be the convex set-valued function associated to f, defined by*

$$F(y) = \mathrm{conv}(\underset{z \to y}{acc} f(z)).$$

Then:

(i) F is OSL with constant L iff *f is OSL with constant L.*

(ii) If f is Lipschitz of constant L, then F is OSL of constant L.

Sketch of proof. *(i)* – The proof of *(i)* is straightforward from the definition of OSL. The main idea is to use the fact that since R^d is a d dimensional space and by the definition of F, any $u \in F(y)$ can be written as the limit of a convex combination of d points $u_1 \dots u_d$ such that $u_i = f(y_i)$ and $\|y - y_i\| \leq \epsilon$. Using a similar combination for $\bar{u} \in F(\bar{y})$ and playing with the coefficient of the convex combination leads to *(i)*.

(ii) – The proof of *(ii)* comes from the Cauchy-Schwartz inequality. If f is Lipschitz with constant L, then for all $y, \bar{y} \in \mathbb{R}^d$: $\langle y - \bar{y}, f(y) - f(\bar{y})\rangle \leq \|y - \bar{y}\| \|f(y) - f(\bar{y})\| \leq L \|y - \bar{y}\|^2$. This shows that f is OSL. By *(i)*, this implies that F is OSL. □

In order to show that a particular dynamic is OSL, we can also use the fact that the sum of two OSL functions is OSL. However, unlike in the case of Lipschitz functions, the product or the composition of two OSL functions is not necessarily OSL.

Lemma 9. *If F_1 and F_2 are two OSL with constant L_1 and L_2, then $F_1 + F_2$ is OSL with constant $L_1 + L_2$, where $F_1 + F_2$ is defined by:*

$$(F_1 + F_2)(y) = \{u_1 + u_2 : u_1 \in F_1(y), u_2 \in F_2(y)\}.$$

4.5.2 Application to energy-aware distributed computing system

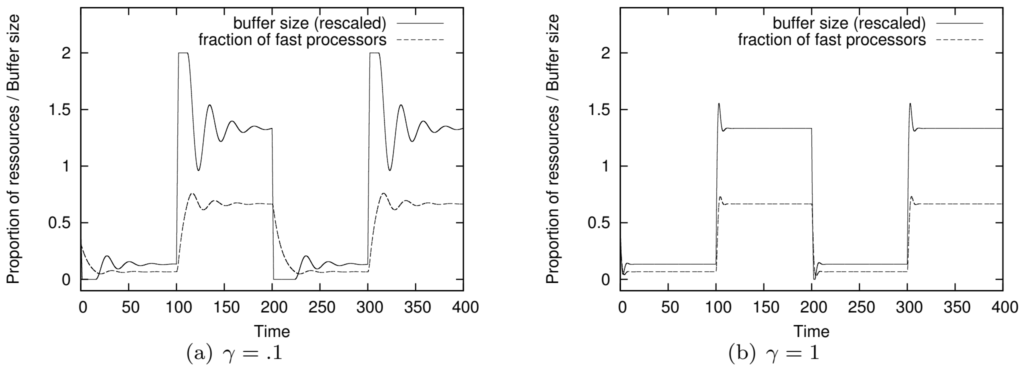
The basic example of a OSL function is the fluid limit of a single M/M/1 queue. Let us assume that jobs arrive in the system at rate $N\lambda > 0$ and are served by a server with rate $N\mu > 0$. Let $Y^N(t)$ be the number of packet at time t in the queue rescaled by $1/N$ and let assume that $Y^N(0) = x$. The drift of the system is independent of N ans is equal to $f(y) = \lambda - \mu$ if $y > 0$ and $f(0) = \lambda$ (that can be extended to $f(y) = \lambda$ for $y < 0$). The set-valued drift corresponding to f is $F(y) = \{\lambda - \mu\}$ if $y > 0$, $F(y) = \{\lambda\}$ if $y < 0$ and $F(0) = [\lambda - \mu; \lambda]$. This function F satisfies a OSL condition with constant 0. Indeed, let $y, y' \in \mathbb{R}$. If y and y' are both positive or both negative, $f(y) - f(y') = 0$. If $y > 0$ and $y' \leq 0$,

$$\langle y - y', f(y) - f(y') \rangle = \langle y - y', -\mu \rangle \leq 0 \leq 0 \left\| y - y' \right\|^2, \tag{21}$$

which is less than 0 since $y > y'$. Therefore, the dynamics of a M/M/1 queue is OSL with constant 0 and one can apply Theorem 3 with $K_T = \lambda$, $I(N) = 1/N$, $J(N) = 0$ and $L = 0$.

Let us now consider a more sophisticated computing model made of a single buffer and N computing resources. Jobs arrive at rate $N\lambda$ in the buffer (where λ may vary) and $C^N(t)$ is the number of jobs in the buffer rescaled by N. The buffer size is upper bounded by $C_{\max}$. In order to optimize energy consumption, the computing resource can compute at rate μ_1 or $\mu_2 > \mu_1$. If there is a proportion $M_1^N(t) = m_1$ resources at speed μ_1 and $M_2^N(t) = m_2 = 1 - m_1$ resources at speed μ_2, the drift of $C^N(t)$ is $f_c(c, m_1, m_2) = \lambda - m_1\mu_1 - m_2\mu_2$ if $0 < c < C_{\max}$, $g(0, m_1, m_2) = \lambda$ and $f_c(C_{\max}, m_1, m_2) = -m_1\mu_1 - m_2\mu_2$. A computation similar to (21) shows that f_c is OSL.

The computing resources can not communicate and each resource updates it speed independently of the others as a function of the load $C^N(t)$ of the system. To do so, each resource measures the size of the queue $C^N(t)$ at rate γ. If the size of the buffer is c and the speed of the resource is μ_i, the speed is set to *fast* (*i.e.* μ_2) with probability $p_i(c)$ (and is set to *slow* with probability $1 - p_i(c)$). Therefore, the drift of $M_1^N(t)$ is $f_1(c, m_1, m_2) = \gamma(-m_1 p_1(c) + m_2(1 - p_2(c)))$ and the drift of $M_2^N(t)$ is $f_2(c, m_1, m_2) = -f_1(c, m_1, m_2)$. If $b \mapsto p_i(c)$ is Lipschitz, the functions f_1 and f_2 are both Lipschitz. This shows that the drift of the whole system (which is the sum of f_c, f_1 and f_2) is a OSL function.



(a) $\gamma = .1$ (b) $\gamma = 1$

Figure 7: Differential inclusion dynamics of the energy-aware distributed systems for $\gamma = .1$ and $\gamma = 1$ and $C_{\max} = 2$. The y axis represents proportion of resources that are fast (*i.e.* $m_2(t)$) or the buffer size, depending on the curve.

In a practical situation, measuring c and changing power of a computing resource is costly in terms of time and energy. Thus, the choice of γ is a compromise between the responsiveness of the system and the time and energy spent to adapt the power to the load. Figure 7 shows two examples of the evolution of the differential inclusion dynamics for two values of γ: $\gamma = .1$ and $\gamma = 1$. The parameters of the system are $\mu_1 = .5$, $\mu_2 = 2$ and $C_{\max} = 2$. On each figure, $p_i(c) = c/C_{\max}$. When $t \in [0; 100]$ or $t \in [200; 300]$, we set $\lambda = .6$ while for $t \in [100; 200]$ or $t \in [300; 400]$, $\lambda = 1.5$.

Figure 7(a) shows that when γ is small, an abrupt change of the arrival rate leads to a long period of oscillation of the system. When going from $\lambda = .6$ to 1.5, some jobs are lost at the beginning. Then, the system overreacts before stabilizing. When $\gamma = 1$ as in Figure 7(b), these oscillations have disappeared, except for a small burst at the beginning.

A Differential inclusions

In this appendix, we recall the main concepts on differential inclusions. For a more complete description, the reader is referred to [1]. In all that follows, $\langle x, y\rangle$ denotes the classical inner-product on $\mathbb{R}^d$ and $\|x\| = \sqrt{\langle x, x\rangle}$ (L^2 norm) and $\|A\| = \sup_{x\in A}\|x\|$.

Definition 10. *Consider a differential inclusion problem:*

$$\dot{y}(t) \in F(y(t)), \quad y(0) = y_0,$$

where F is a set-valued function mapping each point $y \in \mathbb{R}^{d+S}$ to a set $F(y) \subset \mathbb{R}^{d+S}$. Let $I \subset \mathbb{R}$ be an interval with $0 \in I$. A function $y : I \to \mathbb{R}^{d+S}$ is a solution of the DI (6) with initial condition $y(0) = y_0$ if there exists a function $\varphi : I \to \mathbb{R}^{d+S}$ such that:

(i) for all $t \in I$: $y(t) = y_0 + \int_0^t \varphi(s)ds$;

(ii) for almost every (a.e.) $t \in I$: $\varphi(t) \in F(y(t))$.

In particular, *(i)* is equivalent to say that y is absolutely continuous. *(i)* and *(ii)* imply that y is differentiable at almost every $t \in I$ with $\dot{y}(t) \in F(y(t))$.

Definition 11 (Upper Semi-Continuous (USC))**.** *The function F is upper semi-continuous (USC) if for any $y \in \mathbb{R}^{d+S}$, $F(y)$ is a non-empty closed, convex and bounded set and if for any open set O containing $F(y)$, there exists a neighborhood V of y such that $F(V) \subset O$.*

Definition 12 (One-Sided Lipschitz (OSL))**.** *A set-valued function F is one-sided Lipschitz (OSL) with constant L if for all $y, \bar{y} \in \mathbb{R}^{d+S}$ and for all $u \in F(y)$ $\bar{u} \in F(\bar{y})$:*

$$\langle y - \bar{y}, u - \bar{u}\rangle \leq L\,\|y - \bar{y}\|^2 .$$

These two definitions give sufficient conditions for the existence (resp. uniqueness) of solutions for the differential inclusion (6). We recall the following results.

Proposition 13 (Theorems 2.2.1 and 2.2.2 of [18])**.**

- *If F is USC and if there exists c such that $\|F(x)\| \leq c(1 + \|x\|)$ then for any initial condition y_0, (6) has at least one solution on $[0;\infty)$ with $y(0) = y_0$.*
- *If F is OSL, then for all $T > 0$, there exists at most one solution of (6) on $[0;T]$.*

Of course (USC) and (OSL) combined insure that the DI has a unique solution.

B Proofs of Theorem 1 and Theorem 3

This section is devoted to the proof of Theorem 1 and Theorem 3. We first recall some notation of Section 2 before jumping into the proofs.

Let us recall that Y^N is defined by

$$Y^N(k+1) \quad = \quad Y^N(k) + I(N)\left(f^N(Y^N(k)) + U^N(k+1)\right). \tag{22}$$

This equation can be seen as an Euler discretization of the DI (6) $\dot{y} \in F(y)$ plus two error terms:

- A random error term caused by $U^N(k+1)$ which is such that $\mathbb{E}\left(U^N(k+1)\middle|\, Y^N(k)\right)=0$ and is either uniformly integrable (Theorem 1) or bounded in second moment (Theorem 3).
- A "deterministic" error term coming from the fact that $f^N(y)$ is not necessarily in $F(y)$ but converges to F in the sense of Equation 5 (see also Lemma 14):

$$F(y) \overset{\text{def}}{=} \operatorname{conv}\left(\underset{N\to\infty}{\operatorname{acc}} f^N(y^N) \ \text{ with } \lim_{N\to\infty} y^N = y\right).$$

Equation (22) is called a *stochastic approximation* algorithm with *constant step size* associated to the DI (6). The term *constant step size* comes from the fact that $I(N)$ does not vary with time. Both proofs of Theorem 1 and Theorem 3 are based on the convergence of such stochastic approximation (22) as N goes to infinity. However, the two proofs are radically different. The first one is based on compactness argument while the second one focusses on computing explicit error terms.

B.1 Proof of Theorem 1

The classical approach to prove convergence of a stochastic approximation to the solution of the associated differential system uses Gronwall's lemma [12]. Here, we use a different approach, based on compactness properties of the trajectories of the stochastic system. This proof is inspired by several results on differential inclusions, in particular the proof of Theorem 2.2.1 of [18]. However, it is different from Theorem 4.2 of [3] since we need to deal with *constant step sizes* instead of vanishing step sizes (often easier) and we are interested in the convergence over a finite time-horizon. Also, we do not need any *a priori* assumption on the boundedness of the stochastic process.

The idea of the proof is to show that for any sub-sequence of $\bar{Y}^N$, there exists a sub-sequence $\bar{Y}^{\sigma(N)}$ (of this sub-sequence) such that $d(\bar{Y}^{\sigma(N)}, \mathcal{D}_T(y_0)) \xrightarrow{\text{a.s.}} 0$. In all that follows, let $\bar{Y}^{\sigma(N)}$ be a sub-sequence of $\bar{Y}^N$. In order to simplify the notations and because we will take several sub-sequences of sub-sequences, we omit the σ in the notation and we denote all sub-sequences $\bar{Y}^N$. In the first part of the proof, we consider the problem from a probabilistic point of view to make sure that the random part of the process goes almost surely to 0. Then we consider the problem from a trajectorial point of view using analytic arguments.

We first start by two technical lemmas that show that f^N converges to F uniformly on all compact:

Lemma 14. *Let f^N be such that $\left\|f^N(y)\right\| \le c(1+\|y\|)$. Let F be defined by Equation 5 and for all $\epsilon>0$, define F^ϵ by:*

$$F^\epsilon(y) = \left\{z \quad \text{s.t } \exists u \in \mathbb{R}^d, \exists v \in F(u) \text{ with } \|u-y\| \le \epsilon \wedge \|v-z\| \le \epsilon\right\}. \tag{23}$$

Then:

(i) for all compact $K \subset \mathbb{R}^d$, there exists a sequence $J(N) \to 0$ such that, for all $N \ge N_0$ and for all $y \in K$: $f^N(y) \in F^{J(N_0)}(y)$.

(ii) F is USC, i.e., *for all y: $\bigcap_{\epsilon>0} F^\epsilon(y) \subset F(y)$.*

Proof. We prove *(i)* by contradiction. Assume that *(i)* does not hold. Then, there exists a compact K and $\epsilon>0$ such that for all N_0, there exists $N>N_0$ with $y_N \in K$ and $y_N \notin F^\epsilon(y_N)$. Since K is compact, there exists a sub-sequence of y_N that converges to some y. This implies that for N large enough, $\|y_N - y\| \le \epsilon$. Since we assumed that $f^N(y^N) \notin F^\epsilon(y_N)$ and by definition of $F^\epsilon(y^N)$, this implies that for all $v \in F(y)$, $\left\|f^N(y^N)-v\right\| \ge \epsilon$. This contradicts the definition of $F(y)$ which contains the set of limit points of $f^N(y^N)$.

Proof of (ii). Let $v \in \bigcap_{\epsilon>0} F^\epsilon(y)$. This implies that there exists a sequence $y_k \to y$ with $v_k \in F(y_k)$ and $v_k \to v$. By definition of F, v_k is a convex combination of points $\{w_{k,\ell}\}_\ell$ with $f^N(y^N_{k,\ell}) \to w_{k,\ell}$ and $y^N_{k,\ell} \to y_k$. By setting $z_{N,\ell} = y^N_{N,\ell}$, we have $z_{N,\ell} \to y$ and $f^N(z_{N,\ell})$ converges to $w_\ell = \lim_{k\to\infty} w_{k,\ell}$. This shows that $w_\ell \in F(y)$. Therefore, any convex combination of w_ℓ also belongs to $F(y)$. □

Lemma 15. *Let $(U^N(.))_{k\geq 0}$ be a uniformly integrable martingale difference sequence with respect to a filtration $\{\mathcal{F}_k\}$ and let $I(N)$ be a sequence with $I(N)\to 0$. Then for all $T>0$,*

$$\sup_{0\leq t\leq T}\left\|I(N)\sum_{k=0}^{T/I(N)}U^N(k)\right\|\xrightarrow{\mathcal{P}}0.$$

Proof. Let $\epsilon,\nu>0$ and let $V^N(i)=\sum_{k=0}^{i}U^N(k)$. We prove that for N large enough, we have $\mathcal{P}(\sup_{0\leq i\leq T/I(N)}\left\|V^N(i)\right\|\geq\epsilon)\leq\nu$.

Let $\delta=\nu\epsilon/8$. As U^N is uniformly integrable, there exists R such that $\mathbb{E}(U^N(k)\mathbf{1}_{U^N(k)\geq R})\leq\delta$. Define $V^N_+(k)$ and V^N_- as:

$$V^N_+(k)=U^N(k)\mathbf{1}_{U^N(k)\geq R}-\mathbb{E}\left(U(k)\mathbf{1}_{U^N(k)\geq R}\mid\mathcal{F}_{k-1}\right)$$
$$V^N_-(k)=U^N(k)\mathbf{1}_{U^N(k)<R}-\mathbb{E}\left(U(k)\mathbf{1}_{U^N(k)<R}\mid\mathcal{F}_{k-1}\right)=U^N(k)-U^N_+(k)$$

Let Applying Kolmogorov's inequality for martingales, we get:

$$\begin{aligned}\mathcal{P}\left(\sup_{0\leq i\leq T}\left\|V^N(i)\right\|\geq\epsilon\right)&\leq\mathcal{P}\left(\sup_{0\leq i\leq T}\left\|I(N)\sum_{k=0}^{i}U^N_-(k)\right\|\geq\frac{\epsilon}{2}\right)+\mathcal{P}\left(\sup_{0\leq i\leq T}\left\|I(N)\sum_{k=0}^{i}U^N_+(k)\right\|\geq\frac{\epsilon}{2}\right)\\&\leq\frac{4}{\epsilon^2}\mathbb{E}\left(\left\|I(N)\sum_{k=0}^{T/I(N)}U^N_-(k)\right\|^2\right)+\frac{2}{\epsilon}\mathbb{E}\left(\left\|I(N)\sum_{k=0}^{T/I(N)}U^N_+(k)\right\|\right)\\&\leq 16\frac{R^2}{N\epsilon^2}+\frac{4\delta}{\epsilon}\leq 16\frac{R^2}{N\epsilon^2}+\frac{\nu}{2}.\end{aligned}$$

Therefore, for all $N\geq 32R^2/(\epsilon^2\nu)$, this quantity is less than ν. □

Developing the recurrence (4), the value of $Y^N(k+1)$ is equal to:

$$Y^N(k+1)=Y^N(0)+\sum_{i=0}^{k}I(N)f^N(Y^N(i))+I(N)\sum_{i=0}^{k}U^N(i+1).\tag{24}$$

We define two functions $Z^N(t)$, and $V^N(t)$ to be piecewise linear functions such that for all $t=kI(N)$, $Z^N(t)=Y^N(0)+\sum_{i=0}^{k-1}I(N)f^N(Y^N(i))$ and $V^N(t)=\sum_{i=0}^{k-1}I(N)U^N(i+1)$.

By Lemma 15, since U^N is a martingale difference sequence uniformly integrable, $\sup_{0\leq t\leq T}\left\|V^N(t)\right\|$ converges in probability to 0. Therefore, there exists a sub-sequence of V^N such that $\sup_{t\leq T}\left\|V^N(t)\right\|$ converges almost surely to 0.

We now reason from a trajectorial point of view. Let us now consider a trajectory $\omega\in\Omega$ of the system such that $\sup_{t\leq T}\left\|V^N(t)\right\|$ converges to 0. In particular, this implies that $\left\|V^N(t)\right\|$ is bounded for all N and t: $\sup_{N,0\leq t\leq T}\left\|V^N(t)\right\|\leq d<\infty$. Using (24) and since $\left\|f^N(y)\right\|\leq c(1+\|y\|)$, for all $k\leq T/I(N)$, $\left\|Y^N(k+1)\right\|$ can be bounded by:

$$\begin{aligned}\left\|Y^N(k+1)\right\|&\leq\left\|Y^N(0)\right\|+\sum_{i=0}I(N)c(1+\left\|Y^N(i)\right\|)+\sup_{N,t}\left\|V^N(t)\right\|\\&\leq\left\|Y^N(0)\right\|+ckI(N)+d+\sum_{i=0}^{k}I(N)\left\|Y^N(i)\right\|\\&\leq\left(\left\|Y^N(0)\right\|+cT+d\right)\exp\left(cT\right)/c,\end{aligned}\tag{25}$$

where we used the discrete Gronwall's lemma and the fact that $kI(N)\leq T$.

Once we know that $\sup_{N,0\leq t\leq T}\left\|Y^N(t)\right\|$ is bounded, the rest of the proof can be adapted from classical results on the convergence of the Euler approximation for differential inclusions, see [18] for example. There exists $e>0$ such that $\sup_{N,0\leq t\leq T}\left\|Y^N(t)\right\|\leq e$. Thus $\left\|f(Y^N(k))\right\|<c(1+e)<\infty$. This shows that the functions Z^N are Lipschitz with constant $c(1+e)$. Thus the sequence of functions $(Z^N)_N$ are equicontinuous and bounded. Therefore by the Arzéla-Ascoli theorem, for all sub-sequence of $(Z^N)_N$, there exists a sub-sequence that converges to some $z:[0;T]\to\mathbb{R}^d$. In the following, we will show that z is a solution of (3) which shows that $d(Z^N,\mathscr{S}_T(y_0))\to 0$. As $\left\|Z^N-Y^N\right\|=\left\|V^N\right\|\to 0$, this implies that $d(Y^N,\mathscr{S}_T(y_0))\to 0$. To prove this, we will construct a function φ such that:

(i) for all t: $z(t)=z(0)+\int_0^t\varphi(s)ds$;

(ii) for almost every t: $\varphi(t)\in F(z(t))$.

Let $\varphi^N(t)$ be a step function, constant on the intervals $[kI(N),(k+1)I(N))$ and such that for $t=kI(N)$, $\varphi^N(t)=f(Y^N(k))$. Therefore, the sequence φ^N is bounded in $L_2([0;T],\mathbb{R}^d)$. Thus, there exists a sub-sequence of φ^N converging weakly in L_2 to a function φ. Since L_2 is a reflexive space, if a sequence of functions φ^N converges to φ, this means that for all function v, there exists a sub-sequence of φ^N such that $\left\langle v,\varphi^N\right\rangle\to\langle v,\varphi\rangle$. Let $\xi\in\mathbb{R}^d$ and $t\in[0;T]$. Let the function v be defined by $v(s)\stackrel{\text{def}}{=}\xi$ for $s<t$ and $v(s)\stackrel{\text{def}}{=}0$ for $t\geq s$. Since φ^N converges weakly to φ and $Z^N(t)\to z(t)$, we have:

$$\begin{aligned}
\left\langle Z^N(t),\xi\right\rangle &\to \langle z(t),\xi\rangle\\
\left\langle Z^N(t),\xi\right\rangle &= \left\langle Z^N(0),\xi\right\rangle+\left\langle\int_0^t\varphi^N(s)ds,\xi\right\rangle\\
&= \left\langle Z^N(0),\xi\right\rangle+\left\langle\varphi^N,v\right\rangle\\
&\to \langle z(0),\xi\rangle+\langle\varphi,v\rangle\\
&= \left\langle z(0)+\int_0^t\varphi(s)ds,\xi\right\rangle.
\end{aligned}$$

As this is true for all $\xi\in\mathbb{R}^d$, this shows that z is absolutely continuous: $z(t)=\int_0^t\varphi(s)ds$.

It remains to show that for a.e. t, $\phi(t)\in F(z(t))$. Let t^N denotes the greater multiple of $I(N)$ less than t ($t^N\stackrel{\text{def}}{=}\lfloor t/I(N)\rfloor I(N)$). Using that $f^N(Y^N(k))\leq c(1+e)$ and that z^N converges uniformly to y, for all $\delta>0$, there exists N_0 such that $N\geq N_0$ implies $\left\|z(t)-Y^N(t^N)\right\|\leq\delta$. By Lemma 14*(i)*, this shows that for N large enough, $\phi^N(t)\in F^{2\delta}(z(t))$. Since F is convex and z bounded, $\{\alpha\in L^2:\alpha(t)\in F^\delta(z(t))\}$ is convex and closed. This shows that this set is weakly closed (see [23], Theorem 3.12). Therefore, for all t, $\phi(t)\in F^\delta(t)$. As this is true for all δ and because of Lemma 14*(ii)*, this shows $\phi(t)\in\cap_{\delta>0}F^\delta(t)=F(z(t))$. Thus, z is a solution of the DI.

□

B.2 Proof of Theorem 3

The constants A_T,B_T and C_T of Theorem 3 are given by:

$$\begin{aligned}
A_T&=M_T\left(M_T^2+\frac{14M_T}{3}+2K_T\right)\\
B_T&=2M_T^2+4LJ(N)+12K_T\\
C_T&=2M_T^2+4L\epsilon+8K_T,
\end{aligned}$$

with $K_T=\left(\max\left\{\left\|Y^N(0)\right\|,\left\|y(0)\right\|\right\}+(cT+\epsilon)\right)e^{cT}/c$ and $M_T=\sup_{0\leq t\leq T}f^N(Y^N(t))\leq c(1+K_T)$. If $F(.)$ is bounded by some M, the constant M_T is just M and is in particular independent of T. This is true for example if Y^N is constrained to stay in a compact space of $\mathbb{R}^{d+S}$ or if the drift is bounded for all $y\in\mathbb{R}^{d+S}$. The existence of the sequence $J(N)$ is given by the definition of F in Equation 5 (see Lemma 14*(i)*).

By definition, $Y^N(k+1)$ can be written:

$$Y^N(k+1) = Y^N(0) + I(N)\sum_{i=0}^{k} f^N(Y^N(i)) + I(N)\sum_{i=0}^{k} U^N(i+1). \quad (26)$$

Let us define two random sequences Z and V by:

$$Z(k) \stackrel{\text{def}}{=} Y^N(0) + I(N)\sum_{i=0}^{k} f^N(Y^N(i)) \quad \text{and} \quad V(k) \stackrel{\text{def}}{=} I(N)\sum_{i=0}^{k} U^N(i+1).$$

We first start by two lemmas. The first one shows that $V(k)$ is small while the second one computes bounds on the growth of Y^N and the solution of the DI y.

Lemma 16. *For all T and all $\epsilon > 0$,*

$$\mathcal{P}\left(\sup_{i\leq T/I(N)} \left\|V^N(i)\right\| \geq \epsilon\right) \leq \frac{I(N)T}{\epsilon^2}.$$

Proof. Since $\mathbb{E}(U^N(k+1) \mid Y^N(k)) = 0$ and $\mathbb{E}(\left\|U^N(k+1)\right\|^2 \mid Y^N(k)) \leq b$, we have $\mathbb{E}(\|V(k)\|^2) \leq kI(N)^2 b \leq TbI(N)$ for all $k \leq T/I(N)$. Applying Kolmogorov's inequality for martingale to the martingale V leads to the bound of the lemma. ☐

Lemma 17. *Let Y^N be a sequence satisfying (26) with $\left\|f^N(y)\right\| \leq c\,\|1+\|y\|\|$. Let y denotes the solution of the differential equation associated to F.*

Then, if $\sup_{i\leq k}\left\|V^N(i)\right\| \leq \epsilon$, there exists a constant K_T such that

$$\max\left\{\sup_{0\leq k\leq T/I(N)} \left\|Y^N(k)\right\|, \sup_{0\leq t\leq T} \|y(t)\|\right\} \leq K_T.$$

The constant K_T is given by:

$$K_T \stackrel{\text{def}}{=} \left(\max\left\{\left\|Y^N(0)\right\|, \|y(0)\|\right\} + (cT+\epsilon)\right)e^{cT}/c.$$

Proof. By definition of $Y^N(k+1)$, we have:

$$\begin{aligned}\left\|Y^N(k+1)\right\| &\leq \left\|Y^N(0)\right\| + I(N)\sum_{i=0}^{k} c(1+\left\|Y^N(i)\right\|) + \epsilon \\ &= \left\|Y^N(0)\right\| + kI(N)c + \epsilon + I(N)c\sum_{i=0}^{k}\left\|Y^N(i)\right\|.\end{aligned}$$

Therefore, by the discrete Gronwall's lemma, we have $\left\|Y^N(k)\right\| \leq (\left\|Y^N(0)\right\| + (cT+\epsilon)e^{cT}/c$ for $k \leq T/I(N)$.

The proof for y is similar, replacing the discrete Gronwall's inequality by the continuous Gronwall's inequality. ☐

We now jump into the details of the proof.

Let $T > 0$ and $\epsilon > 0$. Assume that $\|V^N(k)\| \leq \epsilon$ for all $k \leq T/I(N)$ and let K_T be defined as in Lemma 17. Since F is OSL, there exists a unique solution y of the DI $\dot{y} \in F(y)$ with $y(0) = y_0$. Therefore, $y(t) = y(0) + \int_0^t f(s)ds$ with $f(s) \in F(y(s))$ *a.e.*

Let $k \leq T/I(N)$ and denote $t_N = kI(N)$.

$$\begin{aligned}
\left\|Z^N(k+1) - y(t_N + I(N))\right\|^2 &= \left\|Z^N(k) - y(t_N) + \int_0^{I(N)} f^N(Y^N(k)) - f(t_N+s)ds\right\|^2 \\
&= \left\|Z^N(k) - y(t_N)\right\|^2 + \left\|\int_0^{I(N)} f^N(Y^N(k)) - f(t_N+s)ds\right\|^2 \\
&\quad + \int_0^{I(N)} 2\left\langle Z^N(k) - y(t_N), f^N(Y^N(k)) - f(t_N+s)\right\rangle ds \\
&\leq \left\|Z^N(k) - y(t_N)\right\|^2 + I(N)^2 4M_T^2 + 2\int_0^{I(N)} w(s)ds,
\end{aligned}$$

where $w(s) \stackrel{\text{def}}{=} \left\langle Z^N(k) - y(t_N), f^N(Y^N(k)) - f(t_N+s)\right\rangle$. To prove the last inequality, we used Lemma 17 that shows that $\left\|Y^N(k)\right\|$ and $\|y(k)\|$ are bounded by K_T. Therefore, there exists a constant M_T such that $\left\|f^N\right\|$ and $\|f\|$ are bounded by M_T.

Because of Lemma 14 that guarantees the speed of convergence of f^N to F , there exists $u \in \mathbb{R}^d$ and $v \in F(v)$ with $\left\|u - Y^N(k)\right\| \leq J(N)$ and $\left\|v - f^N(Y^N(k))\right\| \leq \epsilon$. Thus, $w(s)$ is equal to:

$$\begin{aligned}
w(s) &= \left\langle Z^N(k) - u + u - y(t_N+s) + y(t_N) - y(t_N+s), f^N(Y^N(k)) - v + v - f(t_N+s)\right\rangle \\
&= \left\langle Z^N(k) - u + y(t_N) - y(t_N+s), f^N(Y^N(k)) - f(t_N+s)\right\rangle \\
&\quad + \left\langle u - y(t_N+s), f^N(Y^N(k)) - v\right\rangle + \langle u - y(t_N+s), v - f(t_N+s)\rangle,
\end{aligned}$$

where we developed the inner product using $\langle a+b+c, d+e\rangle = \langle a+c, d+e\rangle + \langle b,d\rangle + \langle b,e\rangle$.

By assumption on u and V, one has $\|Z^N(k) - u\| \leq \|Z^N(k) - Y^N(k)\| + \|Y^N(k) - u\| \leq \epsilon + J(N)$. Moreover, since $\|f\| \leq M_T$, one has $\|y(t_N) - y(t_N+s)\| \leq sM_T$. Combining with the fact that F is OSL of constant L, this gives:

$$w(s) \leq (\epsilon + J(N) + sM_T)M_T^2 + 2K_T J(N) + L\left\|u - y(t_N+s)\right\|^2.$$

Finally, $\|u - y(t_N+s)\|^2$ can be bounded by:

$$\begin{aligned}
\|u - y(t_N+s)\|^2 &= \left\|u - Z^N(k)\right\|^2 + \left\|Z^N(k) - y(t_N)\right\|^2 + \|y(t_N) - y(t_N+s)\|^2 \\
&\quad + 2\left\langle u - Z^N(k), Z^N(k) - y(t_N+s)\right\rangle + 2\left\langle Z^N(k) - y(t_N), y(t_N) - y(t_N+s)\right\rangle \\
&\leq \left\|Z^N(k) - y(t_N)\right\|^2 + (J(N)+\epsilon)^2 + s^2M_T^2 + 2(J(N)+\epsilon)2K_T + 2K_T s M_T.
\end{aligned}$$

This shows that $\int_0^{I(N)} w(s)ds$ can be bounded by $I(N)$ times:

$$\begin{aligned}
L\left\|Z^N(k) - y(t_N)\right\|^2 + (\epsilon + J(N) + \frac{I(N)}{2}M_T)M_T^2 + 2K_T J(N) + L(J(N)+\epsilon)^2 + \frac{I(N)^2 M_T^2}{3} \\
+2(J(N)+\epsilon)2K_T + K_T I(N) M_T.
\end{aligned}$$

Therefore, $I(N)^2 4M_T^2 + 2\int_0^{I(N)} w(s)ds$ is bounded by $2L\left\|Z^N(k) - y(t_N)\right\|^2$ plus $I(N)$ times

$$I(N)M_T\left(M_T^2 + \frac{14M_T}{3} + 2K_T\right) + J(N)\left(2M_T^2 + 4LJ(N) + 12K_T\right) + \epsilon\left(2M_T^2 + 4L\epsilon + 8K_T\right).$$

If a sequence a_k satisfies $a_{k+1} \leq (1 + 2I(N)L)a_k + b$ with $L \neq 0$, one has:

$$a_k = (1+2I(N)L)^k a_0 + \frac{(1+2I(N)L)^k - 1}{2I(N)L}b \leq \mathbf{e}^{2LI(N)k}a_0 + \frac{\mathbf{e}^{2LI(N)k}}{2I(N)L}b.$$

If $L = 0$ and a_k satisfies the recurrence, then $a_k \leq a_0 + kb$. This concludes the proof of the theorem.

B.3 Proof of Theorem 4

Since $\tau < \infty$, the rate of transition of $D^N(.)$ is bounded by $N\tau$. Using uniformization of continuous time Markov chain (see [24] for example), there exists a Poisson process Λ^N of rate $N\tau$ and a discrete time Markov chain $Y^N(.)$ such that $D^N(t) = Y^N(\Lambda^N(t))$ and Y^N and Λ^N are independent. Moreover, for all y and $\ell \in \mathcal{L}$,

$$\begin{aligned} \mathcal{P}\left(Y^N(k+1) = y + \frac{\ell}{N} | Y^N(k) = y\right) &= \frac{1}{\tau}\beta_\ell(y), \\ \mathcal{P}\left(Y^N(k+1) = y | Y^N(k) = y\right) &= 1 - \frac{1}{\tau}\sum_{\ell\in\mathcal{L}}\beta_\ell(y). \end{aligned}$$

For all $y \in \mathbb{R}^d$, the drift of $Y^N(.)$ is $\mathbb{E}\left(Y^N(k+1) - Y^N(k) | Y^N(k) = y\right) = (N\tau)^{-1}f(y)$ and $Y^N(k+1)$ can be written $Y^N(k+1) = Y^N(k) + (N\tau)^{-1}(f(y) + U^N(k+1))$. By assumption $\sum_{\ell\in\mathcal{L}} \|\ell\| \sup_y \beta_\ell(y) < \infty$, U^N is uniformly integrable. Therefore, $Y^N(k)$ satisfies the conditions of Theorem 1. This shows that $\inf_{y\in\mathscr{S}_T(y_0)} \sup_{t\le T} \left\|Y^N(tN) - y(t)\right\| = 0$.

As Λ^N is a Poisson process of rate $N\tau$, $|\Lambda^N(t) - tN\tau|^2$ is a sub-martingale and by Doob's inequality ([13] p 250), $\mathcal{P}(\sup_{t\le T} |\Lambda^N(t) - tN\tau| \ge N\tau\epsilon) \le \mathbb{E}(|\Lambda^N(T) - TN\tau|^2)/(N\tau\epsilon)^2 = (TN\tau)/(N\tau\epsilon)^2 = T/(N\tau\epsilon^2)$. If y is a solution of the DI (6) on $[0;T]$, for all $t, s \in [0,T]$, $\|y(t) - y(s)\| \le c(1 + K_T)\,|t - s|$ where K_T is defined in Lemma 17. This shows that if y is a solution of the differential inclusion, with probability greater than $T/(N\tau\epsilon^2)$, we have:

$$\begin{aligned} \left\|D^N(t) - y(t)\right\| &= \left\|Y^N(\Lambda^N(t)) - y(t)\right\| \\ &\le \left\|Y^N(\Lambda^N(t)) - y\left(\frac{\Lambda^N(t)}{N\tau}\right)\right\| + \left\|y\left(\frac{\Lambda^N(t)}{N\tau}\right) - y(t)\right\| \\ &\le \left\|Y^N(\Lambda^N(t)) - y\left(\frac{\Lambda^N(t)}{N\tau}\right)\right\| + c(1 + K_T)\epsilon. \end{aligned}$$

By Theorem 1, for all $\epsilon > 0$, for all N large enough, there exists a solution y of the DI such that the first term of the last inequality is less than ϵ.

In the OSL case, the constant C'_T is given by $C'_T = C_T + c(1 + K_T)\epsilon$ where C_T is the same as in the previous section (B.2): $C_T = 2M_T^2 + 4L\epsilon + 8K_T$.

References

[1] V. Acary and B. Brogliato. *Numerical methods for nonsmooth dynamical systems: applications in mechanics and electronics*. Springer Verlag, 2008.

[2] U. Ayesta, M. Erausquin, M. Jonckheere, and IM Verloop. Scheduling in a random environment: stability and asymptotic optimality. *Arxiv preprint arXiv:1101.5794*, 2011.

[3] M. Benaim, J. Hofbauer, and S. Sorin. Stochastic approximation and differential inclusions. *SIAM J. Control and Optimization*, 44(1):328–348, 2005.

[4] M. Benaim and J.-Y. Le Boudec. A Class Of Mean Field Interaction Models for Computer and Communication Systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[5] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on selected areas in communications*, 18(3):535–547, 2000.

[6] C. Bordenave, D. Mcdonald, and A. Proutière. Random multi-access algorithms: A mean field analysis. In *Proceedings of Allerton conference*. Citeseer, 2005.

[7] S. Borst, M. Jonckheere, and L. Leskelä. Stability of parallel queueing systems with coupled service rates. *Discrete Event Dynamic Systems*, 18(4):447–472, 2008.

[8] F. Bouchut and F. James. One-dimensional transport equations with discontinuous coefficients. *Nonlinear Analysis*, 32(7):891, 1998.

[9] M. Bramson. Stability of queueing networks. *Probability Surveys*, 5(169-345):1, 2008.

[10] J.G. Dai. On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *The Annals of Applied Probability*, 5(1):49–77, 1995.

[11] RWR Darling and J.R. Norris. Differential equation approximations for Markov chains. *Probability surveys*, 5:37–79, 2008.

[12] J. Dieudonne. *Foundations of modern analysis*. Academic Press, 1960.

[13] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, 1996.

[14] G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for stability of skip-free Markov chains with applications to MCMC. *Annals of Applied Probability*, 18(2):664–707, 2008.

[15] N. Gast and B. Gaujal. A mean field model of work stealing in large-scale systems. *ACM SIGMETRICS Performance Evaluation Review*, 38(1):13–24, 2010.

[16] N. Gast, B. Gaujal, and J.-Y. Le Boudec. Mean field for Markov Decision Processes: from Discrete to Continuous Optimization. *INRIA RR*, 7239, 2010.

[17] M. Jonckheere and S. Shneer. *Stability of multi-dimensional birth-and-death processes with state-dependent 0-homogeneous jumps*. EURANDOM report 2009-029, 2009.

[18] M. Kunze. *Non-smooth dynamical systems*. Springer Berlin Heidelberg, 2000.

[19] T.G. Kurtz. *Approximation of population processes*. Society for Industrial Mathematics, 1981.

[20] F. Lempio. Euler's method revisited. *Proceedings of the Steklov Institute of Mathematics*, 211(4):429–449, 1995.

[21] S.P. Meyn, R.L. Tweedie, and JL Hibey. *Markov chains and stochastic stability*. Springer London et al., 1993.

[22] M. Mitzenmacher. Studying balanced allocations with differential equations. *Combinatorics, Probability and Computing*, 8(5):473–482, 1999.

[23] W. Rudin. *Functional Analysis*. McGraw-Hill Science, 1991.

[24] W.J. Stewart. *Numerical solution of Markov chains*. CRC, 1991.

[25] A.L. Stolyar. On the stability of multiclass queueing networks: a relaxed sufficient condition via limiting fluid processes. *Markov Processes and Related Fields*, 1(4):491–512, 1995.

[26] P. van de Ven, S. Borst, and S. Shneer. Instability of MaxWeight scheduling algorithms. In *INFOCOM 2009, IEEE*, pages 1701–1709. IEEE, 2009.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399