Separated Representations: A Powerful Strategy for Model Reduction
Amine Ammar, Elías Cueto, Francisco Chinesta

To cite this version:
Amine Ammar, Elías Cueto, Francisco Chinesta. Separated Representations: A Powerful Strategy for Model Reduction. 9e Colloque national en calcul des structures, CSMA, May 2009, Giens, France. hal-01413980

HAL Id: hal-01413980
https://hal.archives-ouvertes.fr/hal-01413980
Submitted on 12 Dec 2016
**Separated Representations: A Powerful Strategy for Model Reduction**

A. Ammar\(^1\), E. Cueto\(^2\), F. Chinesta\(^3\)

\(^1\) Laboratoire de Rhéologie  
Université Joseph Fourier  
Amine.Ammar@ujf-grenoble.fr

\(^2\) Aragon Institute of Engineering Research  
Universidad de Zaragoza, Espagne  
ecueto@unizar.es

\(^3\) GeM UMR CNRS-ECN  
École Centrale de Nantes  
Francisco.Chinesta@ec-nantes.fr

**Résumé** — In this paper we discuss about the features of a novel numerical method based on the use of separated representations for the variables of interest. This separated representation allows for an easy treatment of problems defined in a space of high number of dimensions, or in which some parameters are unknown or uncertain. As an example of such a problem we study the problem of cell signalling under a stochastic framework. The method also enables to solve problems in which one of the variables within is not known, by considering these unknowns as new model dimensions.

**Mots clés** — Separated representation, curse of dimensionality, cell signalling.

### 1 Introduction. A method of separation of variables.

It is frequent in Science and Engineering that models are defined in spaces with a high number of dimensions. This characteristic makes the numerical solution of these problems an intricate procedure, since a standard finite element discretisation of them involves an exponentially-growing number of degrees of freedom. This is a well-known phenomenon that has received the name of *curse of dimensionality*.

Several attempts have been made to overcome this curse. The employ of Monte Carlo procedures somewhat alleviates this burden, but it is well known that it converges with low velocity. Sparse grid techniques [1] also help to overcome this difficulty, but are limited to a number of dimensions on the order of 20.

In this paper we overview a new attempt to overcome this difficulty that has proven to be very efficient for several problems [2]. The method is based upon the assumption that the essential variable of the problem can be approximated as a finite sum of separable functions, i.e.,

\[
\Psi(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{NF} c^i F_1^i(x_1) \cdot F_2^i(x_2) \cdot \ldots \cdot F_N^i(x_N)
\]  

(1)
It is readily seen that since each function $F_j$ can be discretised in a one-dimensional space by standard finite elements, the approximation grows only linearly —and not exponentially— with the number of space dimensions.

In order to find the final approximation of the solution, the method is composed of two steps:

- A projection stage in which the best set of $\alpha$ coefficients is sought.
- An enrichment procedure, in which new terms are added to Eq. (1).

$$\Psi = \sum_{j=1}^{n_F} \alpha_j F_j$$

We usually employ an alternated directions approach, in which we are looking at each iteration to only one vector $R_j$ assuming known all the other vectors. Thus, in the weak formulation the test functions given by

$$\Psi^* = R_1 \otimes \ldots \otimes R_{j-1} \otimes R_j^* \otimes R_{j+1} \otimes \ldots \otimes R_N.$$  

For a detailed derivation of the method, the interested reader can consult [2].

### 2 Modelling of gene regulatory networks

In this contribution we are concerned, as an example of potential use of the proposed technique, with a problem of great scientific interest and social relevance. Gene regulatory networks are usually described in terms of the number of molecules of each biochemical species (e.g., $A$, $B$, $C$ and $D$ in Eq. (4)) that are present at a given location $Z(t) = (\#A, \#B, \#C, \#D)^T$, with initial state $Z(t_0) = z_0$.\(^{(4)}\)

This approach is usually preferred for the simulation of cell signaling processes since the number of molecules of interest present at a given region of a cell is usually on the order of dozens or hundreds. In this framework, the random ingredient of the reaction becomes evident and a stochastic differential equation is preferred to describe the process.

For any reaction $R_j$, the propensity function will be given by

$$a_j(z) dt \equiv \text{the probability, given } Z(t) = z, \text{ that } R_j \text{ occurs in } [t, t+dt]. \quad (5)$$

Once reaction $R_j$ has occurred, a change in the number of molecules of the state variables follows, in the amount given by the stoichiometry of the reaction:

$$v_j = s_{i,j}^{in} + s_{i,j}^{out},$$

where $s_{i,j}$ represents the stoichiometric coefficients of molecules of species $i$ involved in reaction $j$. Equivalently,

$$a_j(z - v_j) \rightarrow z,$$  

and

$$a_j(z) \rightarrow z + v_j.$$  

Let us define the probability that each species exists in $z$ number of molecules at any time $t$:

$$P(z, t|z_0, t_0) \equiv \text{Prob}\{Z(t) = z, \text{ given } Z(t_0) = z_0\}. \quad (9)$$
The Chemical Master Equation describes the time evolution of the probability taking into account each propensity $a_j$:

$$\frac{\partial P(z,t|z_0,t_0)}{\partial t} = \sum_j [a_j(z-v_j)P(z-v_j,t|z_0,t_0) - a_j(z)P(z,t|z_0,t_0)].$$

As explained before, the main problem with the simulation of cell signaling processes by the Chemical Master Equation (CME) is that the CME grows exponentially with the number of species [3]. Consider, for instance, a situation in which there are $N$ species with $n$ copies of each specie. The number of possible states in that case is $n^N$. In an eukaryotic cell, for instance, the number of species, on average, can be of the order of $10^4$ different proteins, while there will be $10^6$ copies of each protein on average. This gives a situation in which we have $10^{6000}$ possible states of the cell. But note that the presumed total number of elementary particles in the universe is on the order of $10^{80}$, thus making impossible to deal with such a problem with any —existing or not— computer. In the case of mRNA, or proteines, however, the most common situation is to have on the order of 10 copies per cell of each specie. Even in the case of such a simplification, the problem is of a tremendous magnitude.

The most extended algorithm for the simulation of chemical reaction under the assumption of few copies of each molecule is that of D.T. Gillespie [4]. It is based on the calculation of individual trajectories of the species, instead of solving for the individual state transition probabilities. The method has a slow convergence rate and no guaranteed error bounds. It also requires a large number of realizations [5].

On the contrary, following the separation of variables algorithm described before, consider the CME as given by

$$\frac{\partial P(z,t)}{\partial t} = AP(z,t),$$

where operator $A$ contains the propensities:

$$A = \sum_i A_i$$

One can write each operator $A_i$ in terms of tensorial products applied on each direction of the state space $z$:

$$A = \sum_{j=1}^{n_1} A_{1j} \otimes A_{2j} \cdots \otimes A_{Nj}$$

and, equivalently, the probability:

$$P \approx \sum_{j=1}^{n_1} \alpha^j F_{1j} \otimes F_{2j} \cdots \otimes F_{Nj}.$$  

This algorithm provides a means of overcoming the curse of dimensionality associated with the CME. In the next sections we include some examples to illustrate the performance of the method.

### 3 Numerical examples

#### 3.1 λ-phage switch

This constitutes the simplest model in cell signalling processes. When a bacteriophage λ infects a cell, either stays dormant or it reproduces until the death of the cell. The resulting behaviour depends on two competing proteins that inhibit mutually each other, see Fig. 1. The so-called toggle switch is composed of a two-gene co-repressive network. The constitutive $P_L$ promoter
drives the expression of the \textit{lacI} gene, which produces the lac repressor tetramer. The \textit{lac} repressor tetramer binds the \textit{lac} operator sites adjacent to the \textit{Ptrc} – 2 promoter, thereby blocking transcription of \textit{cl}. The constitutive \textit{Ptrc} – 2 promoter drives the expression of the \textit{cl} gene, which produces the \textit{\lambda}-repressor dimer. The \textit{\lambda}-repressor dimer cooperatively binds to the operator sites native to the \textit{P}_L promoter, which prevents transcription of \textit{lacI}.

The operator for this example is composed by two terms: \( \mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 \):

\[ \mathcal{A}_1 P(z_1, z_2) = \frac{\alpha \beta}{\beta + \gamma z_2} p(z_1 - 1, z_2) + \delta (z_1 + 1) \cdot P(z_1 + 1, z_2) - \left( \frac{\alpha \beta}{\beta + \gamma z_2} + \delta \cdot z_1 \right) P(z_1, z_2), \]

and \( \mathcal{A}_2 \) equivalent with \( z_1 \) and \( z_2 \) interchanged. We computed the solution for \( \delta = 0.05 \), \( \alpha = 1.0 \), \( \gamma = 1.0 \) and \( \beta = 0.4 \).

It can be shown that the corresponding deterministic model in this case leads to a monostable point. The simulations show that after \( t = 5s \) one has a case where both average values of both proteins and small levels of the one protein combined with higher level of the other protein are quite likely, and this remains the case for the stationary distribution as well [6], Fig. 2.

3.2 Simulations in the presence of uncertainty

The true problem when modeling the kinetics of gene networks is the determination of the propensities for all the reactions involved in the network. Although many databases are nowadays
available (see, for instance [3]), for complex reactions, such as the TGF-β, for instance, whose activity has been related to cancer, around half of the propensities have not been determined experimentally.

For this cases, the separation of variables can provide a valuable insight into some problems, by just considering the unknown propensities as new space dimensions. A way to overcome this difficulty is to take the unknown propensity as a new coordinate belonging to the uncertainty interval. The transient solution for a particular value of the propensity can then be computed by restricting the general solution to each particular value of this extra-coordinate. Obviously, the price to be paid is the increase of the model dimensionality. However, this is not a serious issue when one proceeds within the separated representation framework just described.

To illustrate this feature, we have simulated a cascade of two terms. A cascade occurs when a gene produces proteins that enhance the expression of the succeeding gene, see Fig. 3. The operator related to a cascade of length \( k \) writes:

\[
\mathcal{A} = \sum_{i=1}^{k} \mathcal{A}_i
\]

where \( \mathcal{A}_i \) is given by:

\[
\mathcal{A}_i(z)P(z) = \alpha \cdot p(z - e_1) + \delta \cdot (z_1 + 1) \cdot P(z + e_1) - (\alpha + \delta \cdot z_1) p(z)
\]

In order to check the proposed technique, and for the ease of illustration, we have considered a cascade of only 2 terms, with the parameter \( \delta \) as an unknown. Note that the solution (obtained in one execution of the program), see Fig. 4, provides the solution for different values of \( \delta \), that reproduce the ones in the literature [6].

Références

Figure 4 – Solution for the cascade problem with unknown propensities. Solution at time $t = 0$, $t = 30s$, and $t = 600s$ (approx. steady state).