Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/8467

To cite this version :

Amine AMMAR, Elias CUETO, Francisco CHINESTA SORIA - Reduction of the chemical master equation for gene regulatory networks using proper generalized decompositions - International Journal for Numerical Methods in Biomedical Engineering - Vol. 28, n°9, p.960-973 - 2012

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

Arts et Métiers Institute of Technology

Reduction of the chemical master equation for gene regulatory networks using proper generalized decompositions

Amine Ammar [1], Elías Cueto [2,*,†] and Francisco Chinesta [3]

[1] *Arts et Metiers Paris Tech, Angers, France*
[2] *Aragon Institute of Engineering Research, Universidad de Zaragoza, Zaragoza, Spain*
[3] *EADS Corporate Foundation International Chair, Ecole Centrale de Nantes, Nantes, France*

SUMMARY

The numerical solution of the chemical master equation (CME) governing gene regulatory networks and cell signaling processes remains a challenging task owing to its complexity, exponentially growing with the number of species involved. Although most of the existing techniques rely on the use of Monte Carlo-like techniques, we present here a new technique based on the approximation of the unknown variable (the probability of having a particular chemical state) in terms of a finite sum of separable functions. In this framework, the complexity of the CME grows only linearly with the number of state space dimensions. This technique generalizes the so-called Hartree approximation, by using terms as needed in the finite sums decomposition for ensuring convergence.

But noteworthy, the ease of the approximation allows for an easy treatment of unknown parameters (as is frequently the case when modeling gene regulatory networks, for instance). These unknown parameters can be considered as new space dimensions. In this way, the proposed method provides solutions for any value of the unknown parameters (within some interval of arbitrary size) in one execution of the program.

KEY WORDS: gene regulatory networks; chemical master equation; curse of dimensionality; proper generalized decomposition

1. INTRODUCTION

When chemically reacting species are present at very low concentrations (in the number of tens or hundreds of molecules, for instance) the resulting state can not be modeled accurately as deterministic, and the inherent randomness of the system should be taken into account. This is the case, for instance, when modeling gene regulatory networks. It is well known that small numbers of molecules can alter these networks significantly [1, 2].

It is also well known that under some circumstances (a well stirred mixture, fixed volume and fixed temperature), such a system can be considered Markovian, and that it is governed by the so-called chemical master equation (CME) [3], which is a set of linear ordinary differential equations.

The complexity of the CME comes from the fact that it involves one state space dimension for each species involved in the reactions. This means that very classic techniques such as finite elements involve a number of degrees of freedom (DOFs) growing exponentially with the number of reacting species. This difficulty is present in many other branches of science and engineering such as the Schroedinger equation, among others, and has been coined the *curse of dimensionality*.

*Correspondence to: Elías Cueto, Aragon Institute of Engineering Research, Universidad de Zaragoza, Edificio Betancourt, Maria de Luna, s.n., E-50018 Zaragoza, Spain.
†E-mail: ecueto@unizar.es

In practice, finite element-based techniques are limited by this reason to problems on the order of 20 state space dimensions [4], by using sparse grid techniques, for instance.

If we restrict ourselves to the field of cell signaling processes, or in general to the numerical solution of the CME itself, the so-called *stochastic simulation algorithm* [5] is the most extended algorithm to this end. Basically, the stochastic simulation algorithm is a Monte Carlo method, based upon the realization of individual trajectories within the chemical system. The method can become very computationally expensive when the system undergoes enormous numbers of individual reactions. Two different families of alternative methods exist. The first one can be categorized as time-leaping methods, with the τ-leaping method as the most extended one [6], whereas the second one separates slow and fast partitions of the system. For a more detailed analysis see, for instance [3].

It is well known, however, that for some simple systems, it is necessary to run on the order of 10^6 Monte Carlo simulations to achieve the needed precision [3]. On the contrary, models established for complex systems such as the cross-talk pathways of IL-1/NF-κB and TGF-β/Smad involve some 50 different reactions [2], giving a potential state space belonging to $\mathbb{N}^{50}$ ($\mathbb{N}$ represents the set of natural numbers). Considering an interval from 0 to 100 possible copies of each species, this would give a potential state space comprising the order of 10^{100} different states. Noting that the presumed number of elementary particles in the universe is on the order of 10^{80}, we clearly take notice of the formidable task of modeling such systems.

The approach followed herein is completely different. The essential variable of the problem, the probability of having a particular state on the system, provided an initial state, is approximated as a finite sum of separable functions. In this way, we avoid the burden associated with the curse of dimensionality, and all the variables will be approximated by a sum of tensor products.

This kind of representation is not new; it was widely employed in the last decades in the framework of quantum chemistry. In particular, the Hartree–Fock (that involves a single product of functions) and post-Hartree–Fock approaches (as the multi-configurational self-consistent field that involves a finite number of sums) made use of a separated representation of the wave function [7]. In the context of computational mechanics, a similar decomposition was proposed that was called radial approximation and that was applied for separating the space and time coordinates in thermomechanical models [8]. In fact, some prior works that employ Hartree-like approximations for the solution of models established in the form of a CME already exist, see [9] and references therein, for instance. These models assume a particular form of the functions employed along each spatial direction (usually, Poisson-like distributions, [10]).

The approach here presented has been introduced in one of the former works of the authors, [11,12] under the name of proper generalized decompositions and then applied in numerous contexts as follows: (i) quantum chemistry [13]; (ii) Brownian dynamics [14]; (iii) kinetic theory description of polymers solutions and melts [15]; and (iv) kinetic theory descriptions of rod suspensions [16]; among many others. See [17,18] for two recent reviews on the topic.

The outline of the paper is as follows. In the next section, we describe the basics of the stochastic modeling of chemical systems of very low number of molecules. In Section 3, we introduce the essential ingredients of the method, whereas in Section 4, some examples of the performance of the method are presented.

2. THE CHEMICAL MASTER EQUATION

When dealing with chemical system in which the different species are present at very low copy numbers, it is necessary to work with the number of molecules present at each time instant, instead of working, as usual, with the concentration of each species. Thus, consider a well mixed system, at constant volume and temperature, described by the state vector

$$\boldsymbol{Z}(t) = (\#A, \#B, \#C, \#D, \ldots)^T, \tag{1}$$

with initial state $\boldsymbol{Z}(t_0) = z_0$. Here, A, B, and so on, represent different chemical species. When some reaction r_j occurs, the system moves from z_0 to z^*, where the change in the number of

molecules is equal to its stoichiometry in the reaction r_j:

$$\begin{aligned} r_j : s^{\text{in}}_{1,j} x_1 + s^{\text{in}}_{2,j} x_2 + \cdots + s^{\text{in}}_{N,j} x_N \\ \xrightarrow{k_j} s^{\text{out}}_{1,j} x_1 + s^{\text{out}}_{2,j} x_2 + \cdots + s^{\text{out}}_{N,j} x_N, \end{aligned} \tag{2}$$

with k_j as the rate constant of reaction r_j. In turn, $s^{\text{in,out}}_{i,j}$ represent the stoichiometric coefficients of the ith species, and x_k represent the concentrations of each biochemical species.

This state transition depends on the probability that the changes owing to any reaction occur as described by the *propensity function*. For any reaction j,

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

This state transition results in a change of molecules of each species:

$$\boldsymbol{v}_{ij} = \boldsymbol{s}^{\text{in}}_{i,j} + \boldsymbol{s}^{\text{out}}_{i,j},$$

or, equivalently,

$$z - \boldsymbol{v}_j \xrightarrow{a_j(z-\boldsymbol{v}_j)} z, \tag{3}$$

and

$$z \xleftarrow{a_j(z)} z + \boldsymbol{v}_j. \tag{4}$$

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

$$P(z, t|z_0, t_0) \equiv \text{Prob}\{\boldsymbol{Z}(t) = z, \text{ given } \boldsymbol{Z}(t_0) = z_0\}.$$

The CME describes the time evolution of the probability, taking into account each propensity a_j:

$$\frac{\partial P(z, t|z_0, t_0)}{\partial t} = \sum_j \left[a_j(z - \boldsymbol{v}_j) P(z - \boldsymbol{v}_j, t|z_0, t_0) - a_j(z) P(z, t|z_0, t_0)\right]. \tag{5}$$

In compact notation, we will write the CME hereafter as

$$\frac{\partial P(z, t)}{\partial t} = \mathcal{A} P(z, t), \tag{6}$$

where operator $\mathcal{A}$ contains the propensities of each reaction:

$$\mathcal{A} = \sum_j \mathcal{A}_j \tag{7}$$

Note that the choice of the probability, instead of chemical concentrations as essential variable of the problem, eliminates the need of determining the rate constants of each reaction and translates the problem to finding the value of the propensities.

3. A METHOD BASED ON PROPER GENERALIZED DECOMPOSITIONS

The proposed method is constructed by assuming that the essential variable of the problem, the probability of having a particular chemical state, is given by a finite sum of separable functions (separated representation), that is,

$$P(z, t) = \sum_{j=1}^{n_F} \alpha^j F^j_1(z_1) \otimes F^j_2(z_2) \otimes \ldots \otimes F^j_N(z_N) \otimes F_t(t), \tag{8}$$

where, as mentioned before, the variables z_i represent the number of molecules of species i present at a given time instant. This particular choice of the form of the basis functions allows for an

important reduction in the number of DOFs of the problem, $n_N \times N \times n_F$ instead of $(n_N)^N$, where N is the number of dimensions of the state space and n_N the number of DOFs of each one-dimensional grid established for each spatial dimension. For this to be useful, one has to assume that the probability is negligible outside some interval, and therefore, substitute the infinite domain by a subdomain $[0, \ldots, m-1]^N$, m being the chosen limit number of molecules for any species in the simulation. A similar assumption is behind other methods in the literature, such as the finite state projection algorithm, for instance [3].

Another important point to be highlighted is the presence of a function depending solely on time, $F_t(t)$. This means that the algorithm is not incremental. Instead, it solves for the whole time history of the chemical species at each iteration of the method.

The CME is then written in a similar form, as expressed in Equation (7), by expanding the operator $\mathcal{A}$ in the form

$$\mathcal{A} = \sum_{j=1}^{n_A} \mathbb{A}_1^j \otimes \mathbb{A}_2^j \otimes \cdots \otimes \mathbb{A}_N^j \otimes \mathbb{I}, \tag{9}$$

where $\mathbb{A}_i$ represents the matrix form of each operator $\mathcal{A}_i$ involved in the CME and $\mathbb{I}$ represents the identity matrix, acting on the terms depending on time only. Their particular form will be seen readily.

An essential ingredient of the method is how to determine the coefficients α^j and also the particular form of the functions $F_i^j(z_i)$. This is carried out in a two-step algorithm consisting of a projection procedure to determine the coefficients α_i and an enrichment procedure to build the functions $F_i^j(z_i)$.

3.1. Projection algorithm

Assume that we know a set of suitable functions $F_i^j(z_i)$. To determine the coefficient α^j, we put the CME into a weighted form:

$$P^* \frac{\partial P}{\partial t} = P^* \mathcal{A} P \tag{10}$$

where the test function P^* is of the form

$$P^*(z,t) = \sum_{j=1}^{n_F} \alpha^{*j} F_1^j(z_1) \otimes F_2^j(z_2) \otimes \ldots \otimes F_N^j(z_N) \otimes F_t(t). \tag{11}$$

If we represent F_i^j by its nodal values collected in vectors $\boldsymbol{F}_i^j$, the weighted form of the CME will have the following matrix expression:

$$\sum_{i=1}^{n_F} \sum_{j=1}^{n_F} \alpha^{*i} G_{ij} \alpha^j = \sum_{i=1}^{n_F} \sum_{j=1}^{n_F} \alpha^{*i} H_{ij} \alpha^j, \tag{12}$$

where

$$G_{ij} = \left((\boldsymbol{F}_1^i)^T \boldsymbol{F}_1^j \right) \cdot \left((\boldsymbol{F}_2^i)^T \boldsymbol{F}_2^j \right) \cdot \ldots \cdot \left((\boldsymbol{F}_N^i)^T \boldsymbol{F}_N^j \right) \cdot (\boldsymbol{F}_t^i)^T \cdot (\boldsymbol{F}_t^{'j}) \tag{13}$$

with $\boldsymbol{F}_t^{'j}$, the backward finite difference time derivative of the function $\boldsymbol{F}_t^j$ and

$$H_{ij} = \sum_{k=1}^{n_A} \left((\boldsymbol{F}_1^i)^T \mathbb{A}_1^k \boldsymbol{F}_1^j \right) \cdot \left((\boldsymbol{F}_2^i)^T \mathbb{A}_2^k \boldsymbol{F}_2^j \right) \cdot \ldots \cdot \left((\boldsymbol{F}_N^i)^T \mathbb{A}_N^k \boldsymbol{F}_N^j \right) \cdot (\boldsymbol{F}_t^i)^T \cdot (\boldsymbol{F}_t^j). \tag{14}$$

Since Equation (12) must be verified independently of the value of the coefficients α^{*i}, we arrive at an $n_F \times n_F$ system of algebraic equations to be solved for the α^j. A normality condition of the probability function is enforced during the solution. To preserve the verification of the initial condition, we assume that $\alpha^1 = 1$, and the first component of F_t^j $(j \geq 2)$ is vanishing.

3.2. Enrichment algorithm

If, once solved through the system given by Equation (12), the accuracy of the result is adjudged not enough, a method for the enrichment of the basis should be employed. In this case, new functions will be added to the basis:

$$P(z,t) = \underbrace{\sum_{j=1}^{n_F} \alpha^{*j} \boldsymbol{F}_1^j \otimes \boldsymbol{F}_2^j \otimes \ldots \otimes \boldsymbol{F}_N^j \otimes \boldsymbol{F}_t}_{P_F} + \underbrace{\boldsymbol{R}_1 \otimes \boldsymbol{R}_2 \otimes \ldots \otimes \boldsymbol{R}_{N+1}}_{P_R}. \tag{15}$$

To determine the functions $\boldsymbol{R}_i$, we employed a fixed-point alternating directions algorithm, looking for one $\boldsymbol{R}_j$ at each iteration. Other methods such as Newton–Raphson algorithm could equally be employed. In our case, the weighting functions are of the following form:

$$P^* = \boldsymbol{R}_1 \otimes \boldsymbol{R}_2 \otimes \ldots \otimes \boldsymbol{R}_{j-1} \otimes \boldsymbol{R}_j^* \otimes \boldsymbol{R}_{j+1} \otimes \ldots \otimes \boldsymbol{R}_{N+1}, \tag{16}$$

when we look for the function $\boldsymbol{R}_j$. In this case, the different terms of the discrete representation are then given by

$$\begin{aligned} P^* \mathcal{A} P_R &= \sum_{k=1}^{n_A} \left(\left((\boldsymbol{R}_j^*)^T \mathbb{A}_j^k \boldsymbol{R}_j \right) \prod_{h=1, h\neq j}^{N} \left((\boldsymbol{R}_h)^T \mathbb{A}_h^k \boldsymbol{R}_h \right) \right) \cdot (\boldsymbol{R}_{N+1}^T \boldsymbol{R}_{N+1}) \\ &= (\boldsymbol{R}_j^*)^T \mathbb{K}_j \boldsymbol{R}_j \end{aligned} \tag{17}$$

and

$$\begin{aligned} P^* \mathcal{A} P_F &= \sum_{k=1}^{n_A} \left(\left((\boldsymbol{R}_j^*)^T \mathbb{A}_j^k \boldsymbol{F}_j \right) \prod_{h=1, h\neq j}^{N} \left((\boldsymbol{R}_h)^T \mathbb{A}_h^k \boldsymbol{F}_h \right) \right) \cdot (\boldsymbol{R}_{N+1}^T \boldsymbol{F}_t) \\ &= (\boldsymbol{R}_j^*)^T \boldsymbol{V}_j \end{aligned} \tag{18}$$

with their equivalents for the right-hand side

$$\begin{aligned} P^* \frac{\partial P_R}{\partial t} &= \left(\left((\boldsymbol{R}_j^*)^T \boldsymbol{R}_j \right) \prod_{h=1, h\neq j}^{N} \left((\boldsymbol{R}_h)^T \boldsymbol{R}_h \right) \right) \cdot (\boldsymbol{R}_{N+1}^T \boldsymbol{R}'_{N+1}) \\ &= (\boldsymbol{R}_j^*)^T \mathbb{M}_j \boldsymbol{R}_j \end{aligned} \tag{19}$$

with $\boldsymbol{R}'_{N+1}$, the backward finite difference time derivative of the function $\boldsymbol{R}_{N+1}$ and

$$P^* \frac{\partial P_F}{\partial t} = \left(\left((\boldsymbol{R}_j^*)^T \boldsymbol{F}_j \right) \prod_{h=1, h\neq j}^{N} \left((\boldsymbol{R}_h)^T \boldsymbol{F}_h \right) \right) \cdot (\boldsymbol{R}_{N+1}^T \boldsymbol{F}'_t) = (\boldsymbol{R}_j^*)^T \boldsymbol{W}_j \tag{20}$$

with $\boldsymbol{F}'_t$, the backward finite difference time derivative of the function $\boldsymbol{F}_t$. This gives rise to a system of equations of the form

$$(\mathbb{M}_j - \mathbb{K}_j) \boldsymbol{R}_j = \boldsymbol{V}_j - \boldsymbol{W}_j. \tag{21}$$

For the initial conditions to be verified at subsequent time instants, we take $\alpha^1 = 1$, with the first value of the new functions $\boldsymbol{F}_t^j$ (for $j \geqslant 2$) vanishing.

Remark 1
The sequence projection-enrichment is iterated until convergence is reached. Two stopping criteria are considered in our works. The simplest one consists of computing the norm of P_R. Despite its simplicity, the convergence is not guaranteed. The other possibility consists of computing the norm of the equation residual. This stopping criterion ensures the convergence, but it is more expensive from the computational viewpoint.

Remark 2
When convergence is reached, functions R_j are normalized and named $F_j^{n_F+1}$.

Remark 3
We also can use the separated representation introduced before in the framework of an incremental strategy. This enforces us to solve a problem in dimension N at several time instants. Assuming P_t is known in the solution at time t, the solution at the next time instant can be obtained after solving the following problem, written as a tensorial product of operators:

$$\left(\mathbb{I}_1 \otimes \ldots \otimes \mathbb{I}_N - \Delta t \cdot \sum_{j=1}^{n_A} \mathbb{A}_1^j \otimes \ldots \otimes \mathbb{A}_N^j \right) P^{t+\Delta t} = P^t \tag{22}$$

In this last equation, the right-hand-side term is known, and one only needs to find $P^{t+\Delta t}$. This discretization corresponds to an implicit time integration scheme.

4. NUMERICAL EXAMPLES

4.1. Simulation of a toggle switch

The behavior of the λ-phage virus is one the most studied and well-known examples in gene regulatory networks. When a bacteriophage λ infects a cell, it either stays dormant, or it reproduces until the cell dies. The resulting behavior depends crucially on two competing proteins that mutually inhibit each other, see a schematic representation in Figure 1. The so-called toggle switch is composed of a two-gene co-repressive network.

The operator form of the CME for this example is composed of two terms [19]: $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2$, given by

$$\begin{aligned} \mathcal{A}_1 P(z_1, z_2) = {} & \frac{\alpha\beta}{\beta + \gamma z_2} P(z_1 - 1, z_2) + \delta(z_1 + 1) \cdot P(z_1 + 1, z_2) \\ & - \left(\frac{\alpha\beta}{\beta + \gamma z_2} + \delta \cdot z_1 \right) P(z_1, z_2). \end{aligned} \tag{23}$$

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

The simulation started from a non-physiological state in which both proteins showed a very high probability, around $z_1 = z_2 = 15$. Despite this initial state, after $t = 100$ s (Figure 2), 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 [19], Figure 3.

4.2. A cascade of 20 genes

In order to show the potential of the method, we have solved a completely academic example, composed of a cascade of 20 terms. A cascading process occurs when (usually) adjacent genes

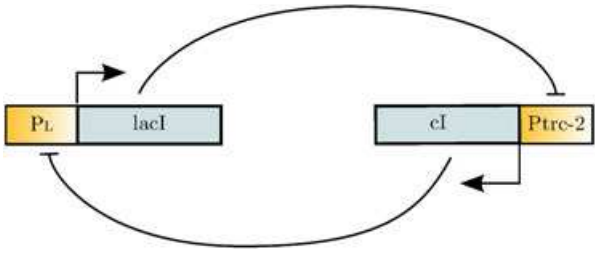


Figure 1. Schematic mechanism of the toggle switch [1]. The constitutive P_L promoter drives the expression of the $lacI$ gene, which produces the lac repressor tetramer. The lac repressor tetramer binds the lac operator sites adjacent to the $Ptrc-2$ promoter, thereby blocking transcription of cI. The constitutive $Ptrc-2$ promoter drives the expression of the cI gene, which produces the λ-repressor dimer. The λ-repressor dimer cooperatively binds to the operator sites native to the P_L promoter, which prevents transcription of $lacI$.

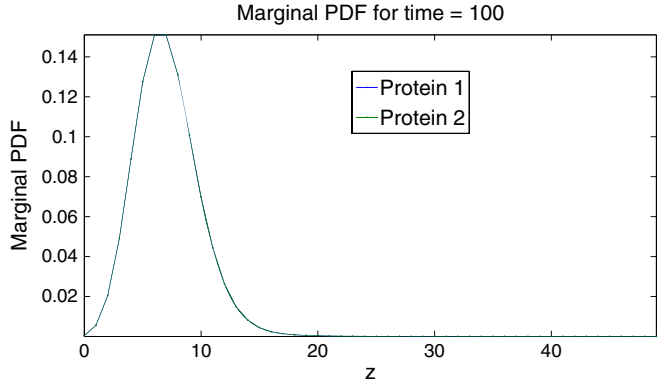


Figure 2. Marginal probability distribution function at $t = 100$ s. Axes denote the number of proteins 1 and 2.

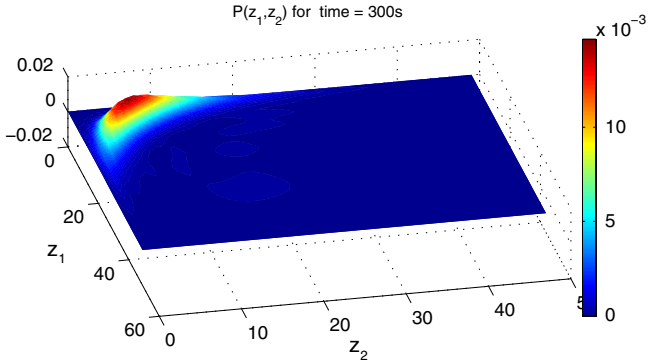


Figure 3. Solution at steady state ($t \approx 300$ s) by separation of variables. Axes denote the number of protein 1 (abscissa) and protein 2 (ordinate).

produce protein, which enhances the expression of the succeeding gene. One typical example can be found in the lytic phase of the λ-phage system [19], but is composed of substantially fewer terms.

This example is intended to show the potential of the proposed method, because methods based upon sparse grids, for instance, rarely can cope with more than 10 proteins [19]. Although completely academic, we believe that this example clearly demonstrates how a realistic number of different reactions can be handled without any particular difficulty (Figure 4).

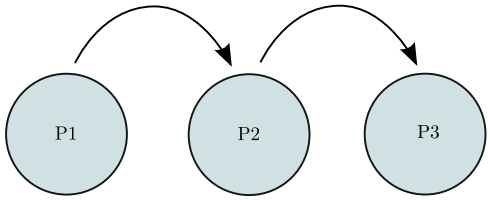


Figure 4. Schematic representation of a cascade.

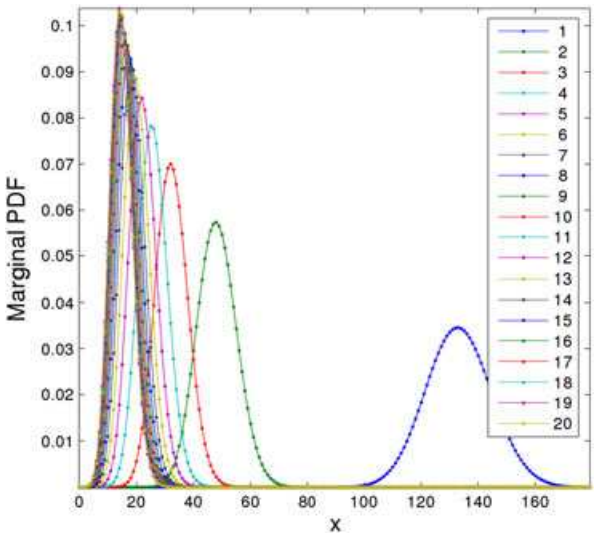


Figure 5. The simulation of a cascade composed of 20 terms shows the typical delay that has been observed for simpler examples.

The operator for this example is composed of 20 terms: $\mathcal{A} = \mathcal{A}_1 + \ldots + \mathcal{A}_{20}$. Operator $\mathcal{A}_1$ has the same form as that of the previous example, whereas the succeeding operators are

$$\begin{aligned}\mathcal{A}_i\, P(z) = &\frac{\beta z_{i-1}}{\beta z_{i-1}+\gamma} P(z-\boldsymbol{e}_i) + \delta(z_i+1) P(z+e_i) \\ &- \left(\frac{\beta z_{i-1}}{\beta z_{i-1}+\gamma} + \delta z_i\right) P(z), \quad \forall i > 1, \end{aligned} \tag{24}$$

where e_i is the ith standard basis of $\mathbb{R}^n$.

In the cascade, the expectations of the marginal distributions have a clear in-built delay. As an example, consider the case of three species with production of the first species $\alpha_0 = 0.7$, decay for all the species $\delta = 0.07$ and the production of succeeding species equal to $z_{i-1}/(5.0 + z_{i-1})$ for $i = 2, \ldots, 20$, respectively, see [19]. The results are depicted in Figure 5.

It can be noticed that the size of the bounded domain does not very much influence the resulting size of the model owing to the special separated representation of the variables. Note how, since only one-dimensional domains are 'meshed' for each species, a large domain can be considered without any special difficulty.

To compare the proposed method with Hartree-like methods, we conducted the same simulations, but employing only one summand in Equation (8), that is, $n_F = 1$. This is essentially the main difference of the proposed method with Hartree approaches. It must be noted, in addition, that no particular form of the separated function is assumed.

To see the implication of the use of one summand only, a convergence plot is shown in Figure 6 for the two cases described earlier.

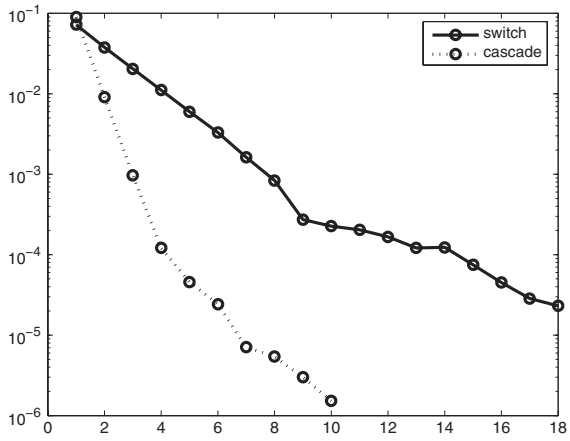


Figure 6. Convergence of the result for the toggle-switch and cascade problems (error versus number of functions employed in the approximation). Note how the use of one-summand approximation could lead to eventually high errors.

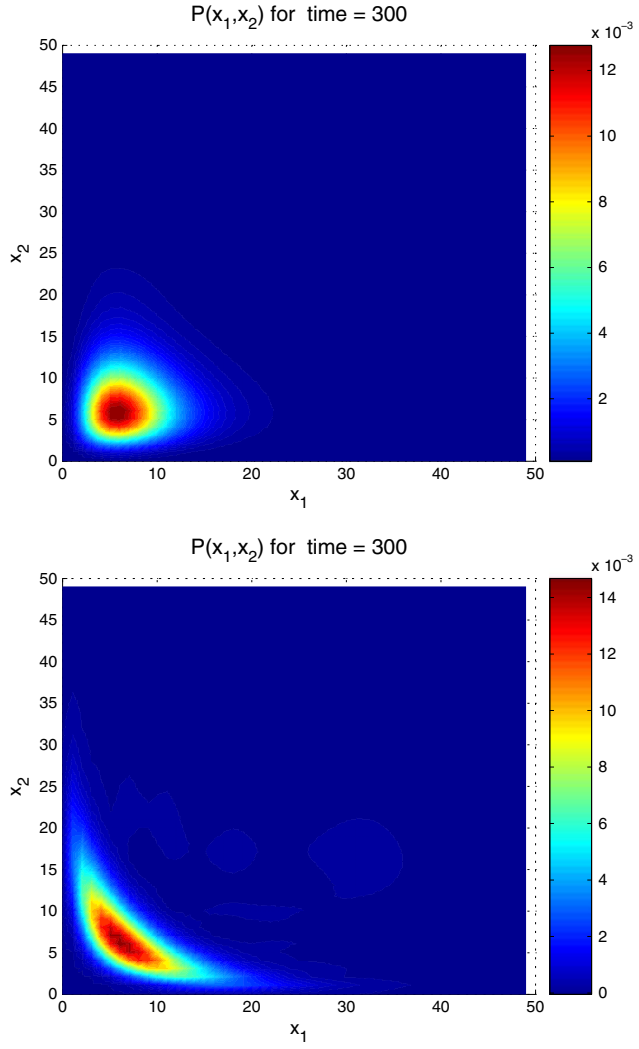
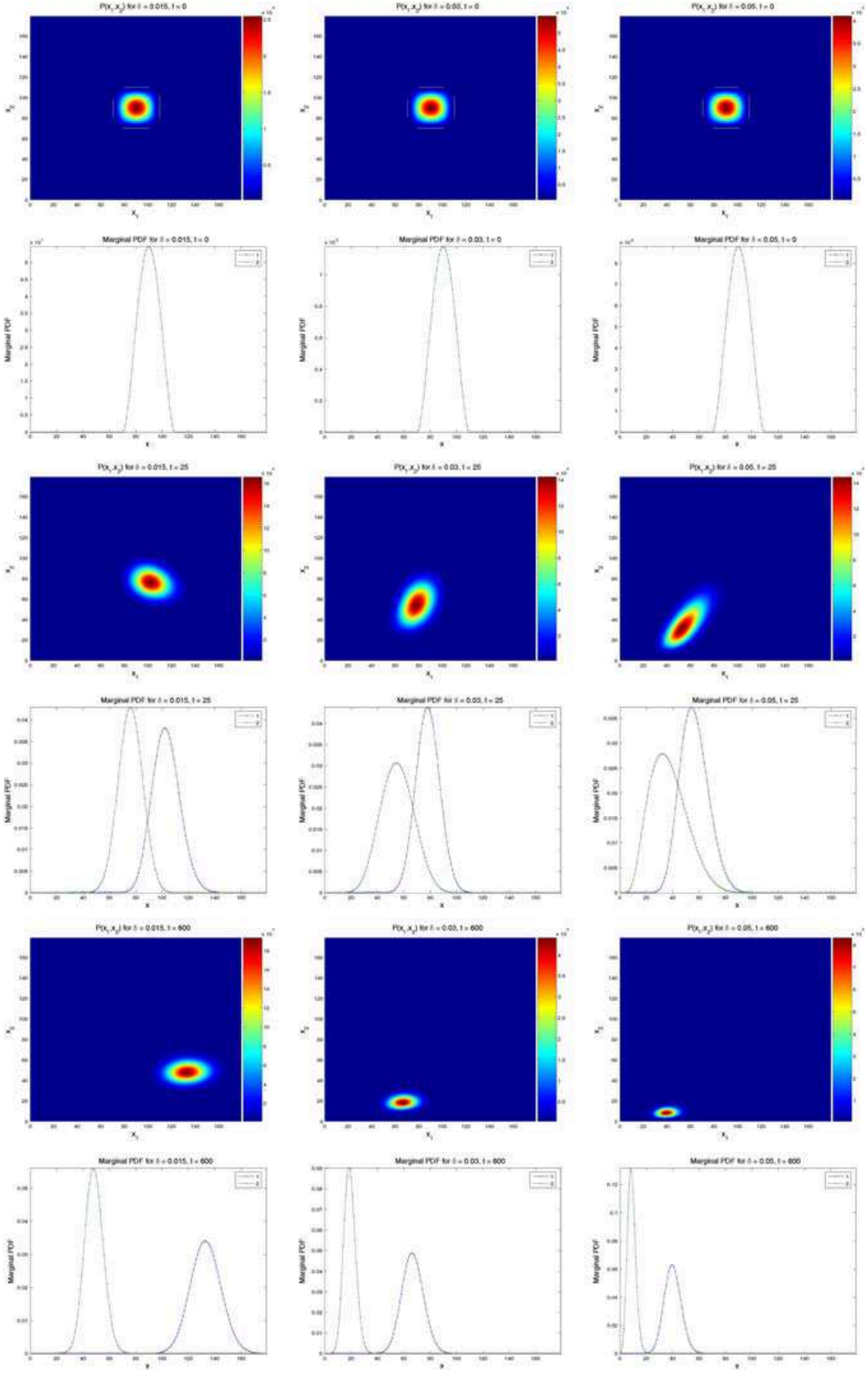


Figure 7. Final state of the toggle-switch model for a Hartree-like model (up), and the converged solution obtained by the proposed method (down). Axis represent the concentration of each competing protein.


P(x1,x2) for δ = 0.015, t = 0
P(x1,x2) for δ = 0.03, t = 0
P(x1,x2) for δ = 0.05, t = 0
Marginal PDF for δ = 0.015, t = 0
Marginal PDF for δ = 0.03, t = 0
Marginal PDF for δ = 0.05, t = 0
P(x1,x2) for δ = 0.015, t = 25
P(x1,x2) for δ = 0.03, t = 25
P(x1,x2) for δ = 0.05, t = 25
Marginal PDF for δ = 0.015, t = 25
Marginal PDF for δ = 0.03, t = 25
Marginal PDF for δ = 0.05, t = 25
P(x1,x2) for δ = 0.015, t = 600
P(x1,x2) for δ = 0.03, t = 600
P(x1,x2) for δ = 0.05, t = 600
Marginal PDF for δ = 0.015, t = 600
Marginal PDF for δ = 0.03, t = 600
Marginal PDF for δ = 0.05, t = 600
Marginal PDF
x

The final result could also be importantly affected. For comparison, we show the final state of the toggle-switch system using only one function and the converged one, as provided by the method, see Figure 7. As can be noticed, the result can be quite different, and this is particularly noteworthy for the toggle-switch system.

4.3. Simulations in the presence of uncertainty

The most important problem when modeling gene regulatory networks, however, comes from the fact that many reaction's parameters are unknown or very difficult to determine in an easy way [2]. The easy way in which the method copes with many state space dimensions allows it to handle unknown parameters in a natural way, by considering them as new state space dimensions, thus, opening the way to designing *in silico* experiments in which one could eventually observe the effect of different values of the unknown parameters on the final chemical state of the system.

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. The operator related to a cascade was introduced in the previous section. To check the proposed technique, and for the ease of illustration, we have considered a cascade of only 2 terms, with the parameter δ as an unknown. Note that the solution (obtained in one execution of the program), see Figure 8, provides the solution for different values of δ, that reproduces the ones in the literature [19].

In order to obtain the solution for a given value of the unknown parameter, one has to cut the hypercube of the solution by a hyperplane located at a particular value of the variable.

5. CONCLUSIONS

We have presented a technique for the numerical solution of the CME based on the approximation of the essential variable through a finite sum of separable functions. In this way, we can avoid the burden associated with the curse of dimensionality, that is, the exponentially growing number of DOFs with the number of state space dimensions.

This strategy allows for a simple method in which time is considered as another variable, such that all the time history of the system is solved for at each iteration, very much like space–time finite element techniques. Indeed, unknown parameters of the problem, such as reaction propensities, can be considered as new state space dimensions. This produces results for any value of the unknown variable. By cutting the hypercube of the solution field through any hyperplane, we readily obtain the behavior of the system for particular values of the unknowns.

This technique avoids the problems related to Monte Carlo-like techniques, such as the burden associated with high number of dimensions, the need for a large number of realizations of the system if high accuracy is needed or the lack of appropriate error bounds in the results.

Of course, it remains to be determined if the method can help clarify some very complex systems such as the cross-talk between interleukin-1-induced nuclear factor-κB and TGF-β-induced Smad pathways, for instance. This system has been modeled by using some 50 different reactions [2], with approximately one half of them not appropriately understood nor characterized. But, undoubtedly, the presented technique can help to understand the effect of some values of the unknown variables in the final state of the system.

Figure 8. Solution for the cascade problem with unknown propensities. Probability distribution function (top row) and marginal probability distribution function of each species (bottom row) at time $t = 0$, $t = 30$ s, and $t = 600$ s (approx. steady state). The left column presents the results for a value $\delta = 0.015$, whereas the central one is for $\delta = 0.03$ and the left one for $\delta = 0.05$. Note that all the results are obtained in one execution of the program. The four-dimensional hypercube containing the solution space, whose dimensions are the concentrations of the two proteins, the value of δ and time, is then cut by the hyperplanes defined by the different values of δ and time to give these plots.

APPENDIX A: DISCRETE FORMS

In order to make a discrete form of the equations of the cascade model, we need to define some matrices. These matrices are related to each species 'i', and their associated size is equal to $(m_i \times m_i)$, the number of operators in the model. The identity matrices will be denoted

$$\mathbb{I}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{25}$$

Two other matrices are also required to define the shift allowing to get the neighboring probability

$$\mathbb{I}_i^+ = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ & & & & 1 \\ 0 & \cdots & & 0 & 0 \end{pmatrix} \tag{26}$$

$$\mathbb{I}_i^- = \begin{pmatrix} 0 & 0 & \cdots & & 0 \\ 1 & 0 & \ddots & & \vdots \\ 0 & 1 & \ddots & & \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \tag{27}$$

According to these definitions, one can write each operator $\mathcal{A}_i$ in terms of tensorial product of operators applied on each dimension that consists of $\{0, \ldots, m_{i-1}\}$ components. The first operator writes

$$\begin{aligned} \mathcal{A}_1 = \Bigg(& \alpha \cdot \mathbb{I}_1^- + \delta \cdot \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & m_1 \end{pmatrix} \cdot \mathbb{I}_1^+ \\ & - \alpha \cdot \mathbb{I}_1 - \delta \cdot \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & m_1 - 1 \end{pmatrix} \Bigg) \otimes \mathbb{I}_2 \otimes \ldots \otimes \mathbb{I}_N \end{aligned} \tag{28}$$

that defines a single tensor product. For $i > 1$ the operator A_i consists of four-tensor products:

$$\mathcal{A}_i = \mathbb{I}_1 \otimes \ldots \otimes \mathbb{I}_{i-2} \otimes \begin{pmatrix} a(0) & 0 & \cdots & 0 \\ 0 & a(1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a(m_{i-1} - 1) \end{pmatrix} \otimes \mathbb{I}_i^- \otimes \mathbb{I}_{i+1} \otimes \ldots \otimes \mathbb{I}_N$$

$$+\delta\cdot\mathbb{I}_1\otimes\ldots\otimes\mathbb{I}_{i-1}\otimes\begin{pmatrix}1&0&\cdots&0\\0&2&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&0&m_i\end{pmatrix}\cdot\mathbb{I}_i^{+}\otimes\mathbb{I}_{i+1}\otimes\ldots\otimes\mathbb{I}_N-$$

$$-\ \mathbb{I}_1\otimes\ldots\otimes\mathbb{I}_{i-2}\otimes\begin{pmatrix}a(0)&0&\cdots&0\\0&a(1)&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&0&a(m_{i-1}-1)\end{pmatrix}\otimes\mathbb{I}_i\otimes\mathbb{I}_{i+1}\otimes\ldots\otimes\mathbb{I}_N-$$

$$-\delta\cdot\mathbb{I}_1\otimes\ldots\otimes\mathbb{I}_{i-1}\otimes\begin{pmatrix}0&0&\cdots&0\\0&1&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&0&m_i-1\end{pmatrix}\cdot\mathbb{I}_i\otimes\mathbb{I}_{i+1}\otimes\ldots\otimes\mathbb{I}_N \tag{29}$$

where

$$a(y)=\frac{\beta y}{\beta y+\gamma}. \tag{30}$$

ACKNOWLEDGEMENTS

This work has been partially funded by the Spanish Ministry of Science and Innovation (through grant number CICTY-DPI2011-27778-C02-010), whose help is gratefully acknowledged.

REFERENCES

1. Hasty J, McMillen D, Isaacs F, Collins JJ. Computational studies of gene regulatory networks: In numero molecular biology. *Nature Reviews Genetics* 2001; **2**:268–279.
2. Sreenath SN, Cho K-H, Wellstead P. Modelling the dynamics of signalling pathways. *Essays in biochemistry* 2008; **45**:1–28.
3. Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics* 2006; **124**(4):044104.
4. Bungartz H-J, Griebel M. Sparse grids. *Acta Numerica* 2004; **13**:1–123.
5. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 1977; **81**(25):2340–2361.
6. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics* 2001; **115**:1716–1733.
7. Cancès E, Defranceschi M, Kutzelnigg W, Le Bris C, Maday Y. Computational quantum chemistry: A primer. In *Handbook of numerical analysis*, Vol. X: North Holland, Amsterdam, 2003; 3–270.
8. Ladeveze P. *Nonlinear computational structural mechanics*. Springer: N.Y., 1999.
9. Kim K-Y, Wang J. Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *PLoS Computational Biology* 2007; **3**:0565–0577.
10. Sasai M, Wolynes PG. Stochastic gene expression as a many-body problem. *Proceedings of the National Academy of Sciences* 2003; **100**(5):2374–2379.
11. Ammar A, Mokdad B, Chinesta F, Keunings R. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. *Journal of Non-Newtonian Fluid Mechanics* 2006; **139**:153–176.
12. Ammar A, Mokdad B, Chinesta F, Keunings R. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. Part II: Transient simulation using space–time separated representations. *Journal Non-Newtonian Fluid Mechanics* 2007; **144**:98–121.
13. Chinesta F, Ammar A, Joyot P. The nanometric and micrometric scales of the structure and mechanics of materials revisited: An introduction to the challenges of fully deterministic numerical descriptions. *International Journal for Multiscale Computational Engineering* 2008; **6**:191–213.
14. Chinesta F, Ammar A, Falco A, Laso M. On the reduction of stochastic kinetic theory models of complex fluids. *Modeling and Simulation in Materials Science and Engineering* 2007; **15**:639–652.
15. Mokdad B, Pruliere E, Ammar A, Chinesta F. On the simulation of kinetic theory models of complex fluids using the Fokker–Planck approach. *Applied Rheology* 2007; **17**:1–14.

16. Pruliere E, Ammar A, El Kissi N, Chinesta F. Multiscale modelling of flows involving short fiber suspensions. *Archives of Computational Methods in Engineering* 2009; **16**:1–30.
17. Chinesta F, Ammar A, Cueto E. Recent advances and new challenges in the use of the proper generalized decomposition for solving multidimensional models. *Archives of Computational Methods in Engineering* 2010; **17**:327–350.
18. Chinesta F, Ladeveze P, Cueto E. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering* 2011; **18**:395–404.
19. Hegland M, Burden C, Santoso L, MacNamara S, Boothm H. A solver for the stochastic master equation applied to gene regulatory networks. *Journal of Computational and Applied Mathematics* 2007; **205**:708–724.