Chemical Master Equation Empirical Moment Closure

A. Ammar

LAMPA, ENSAM Angers, France & UMSSDT, ENSIT Tunis, Tunisie.

M. Magnin, O. Roux

IRCCyN, Ecole Centrale de Nantes.
1 rue de la Noe, BP 92101, F-44321 Nantes cedex 3, France.

E. Cueto

I3A, Universidad de Zaragoza. Zaragoza, Spain.

F. Chinesta*

GeM, Ecole Centrale de Nantes.
1 rue de la Noe, BP 92101, F-44321 Nantes cedex 3, France.
* *Corresponding author Francisco Chinesta, Francisco.Chinesta@ec-nantes.fr*

Abstract

The numerical solution of the Chemical Master Equation (CME) governing gene regulatory networks and cell signaling processes remains a challenging task due to its complexity, exponentially growing with the number of species involved. When considering separated representations of the probability distribution function within the Proper Generalized Decomposition – PGD –framework the complexity of the CME grows only linearly with the number of state space dimensions. In order to speed up calculations moment-based descriptions are usually preferred, however these descriptions involve the necessity of using closure relations whose impact on the calculated solution is most of time unpredictable. In this work we propose an empirical closure, fitted from the solution of the chemical master equation, the last solved within the PGD framework..

Keywords: Chemical Master Equation, Proper Generalized Decomposition, Curse of dimensionality, Closure relations

1. Introduction

Simulating the behavior of gene regulatory networks is a formidable task for several reasons. At this level of description, only a few molecules (maybe dozens or hundreds) of each species involved in the regulation process is present, and this fact limits the possibility of considering the process as deterministic, as is done very often in most chemical applications. Here, the concept of concentration of the species does not make sense [10] [16]. On the contrary, under some weak hypotheses the system can be considered as Markovian, and can be consequently modeled by the so-called Chemical Master Equation (CME) [15], which is in fact no more than a set of ordinary differential equations stating the conservation of the probability distribution function – pdf – P in time:

$$\frac{\partial P(\boldsymbol{z}, t|\boldsymbol{z}_0, t_0)}{\partial t} = \sum_j \Big[a_j(\boldsymbol{z} - \boldsymbol{v}_j) P(\boldsymbol{z} - \boldsymbol{v}_j, t|\boldsymbol{z}_0, t_0) - a_j(\boldsymbol{z}) P(\boldsymbol{z}, t|\boldsymbol{z}_0, t_0) \Big], \qquad (1)$$

where $P(\boldsymbol{z}, t|\boldsymbol{z}_0, t_0)$ represents the probability of being at a state in which there are a number of molecules of each species stored in the vector $\boldsymbol{z}$ at time t when we started from a state $\boldsymbol{z}_0$ at time t_0. a_j represents the *propensity* (i.e., the probability) of reaction j to occur, while $\boldsymbol{v}_j$ represents the change in the number of molecules of each species if reaction j takes place. This change is given, of course, by the stoichiometry of the reaction at hand.

What is challenging, however, in this set of equations is that they are defined in a state space which possesses as many dimensions as the number of different species involved in the regulatory network. Under this framework, if we consider N different species, present at a number n of copies, the number of different possible states of the system is n^N. This number can take the astronomical value of 10^{6000} if we consider some types of proteins, for instance [15]. This phenomenon is known as the *curse of dimensionality* in many branches of science.

To overcome this difficulty, most of the authors employ Monte Carlo-like algorithms (the so-called stochastic simulation algorithm, SSA [15] [6] [7]).

However, Monte Carlo techniques need for as many as possible individual realizations of the problem, leading to excessive time consuming simulations, together with great variance in the results.

Separated representations involved in the Proper Generalized Decomposition described below allow circumventing the issues related to the high-dimensional character of the CME as was successfully proved in our former works discussed later. However, and despite the fact of being able to solve the CME, its solution requires a significant amount of computation, and thus, the simulation of a variety of scenarios remains a challenging issue because its computational complexity. For alleviating such a computational complexity an appealing route consists of calculating the moments of the probability distribution function instead of the pdf itself. Moments constitute a valuable description of great interest in many practical applications and then moment-based descriptions represent an appealing alternative to pdf-based descriptions. However, as discussed later, when deriving the equations that govern the time evolution of the pdf-moments, usually the one related to a moment of a certain order depends on higher-order moments and so-on. In order to close the model at a certain order, we must approximate higher order moments as a function of the ones lower or equal to the one considered. Such an approached constitutes the closure-based description.

As discussed later different closures has been proposed however, no closure relation is general enough to represent any possible scenario with the required accuracy. In this paper we propose using the expensive but very accurate CME solution efficiently obtained by invoking the PGD technology for fitting empirical polynomial closures. These closures are then used for obtaining moment-based solutions in an efficient way, because the integration of the evolution equations governing the time evolution of these moments can be performed almost in real-time, for scenarios that slightly differ from the ones that served to construct the empirical closure relation.

In any case it is important to note that the validity and accuracy of the computed closure-based solutions is never assured but in many cases is a valuable

tool for pre-analysis. As soon as the main tendencies are obtained using the fast closure-based simulations, by performing a lot of simulations, all them very fast, finer analysis can be performed by solving the CME in a few selected scenarios of special interest extracted from the previous fast closure-based simulations.

1.1. Proper Generalized Decomposition for alleviating the curse of dimensionality

Dealing with the problem of the curse of dimensionality in a very different context, the authors presented in a previous work a technique that is now known under the name of Proper Generalized Decomposition (PGD) based on the use of separated representations [2] [3]. Essentially, to avoid the exponentialy-growing complexity of the problem with the number of state space dimensions, the method approximates the variable of interest, say u, as a finite sum of separable functions:

$$u(x_1, x_2, \ldots, x_D, t) \approx \sum_{i=1}^{N} F_1^i(x_1) \cdot F_2^i(x_2) \cdot \ldots \cdot F_D^i(x_D) \cdot T^i(t). \tag{2}$$

The reason for this particular choice motivated the method itself, that is conceived as a greedy algorithm that computes one sum at a time and one product at a time, within a fixed-point, alternating directions algorithm. This leads to a sequence of one-dimensional (low-dimensional, in general) problems, one for each function F_j^i that can be solved using your favorite technique (finite elements, finite volumes, finite differences, colocation, ...).

If M nodes are used to discretize each coordinate, the total number of PGD unknowns is $N \times M \times D$ instead of the M^D degrees of freedom involved in standard mesh-based discretizations. Moreover, all numerical experiments carried out to date with the PGD show that the number of terms N required to obtain an accurate solution is not a function of the problem dimension D, but it rather depends on the regularity and separability of the exact solution as well as on the solution constructor itself. The PGD thus avoids the exponential complexity with respect to the problem dimension.

1.2. A PGD approach to gene regulatory networks simulation

The PGD approach to the problem of efficiently simulating gene regulatory networks begins by assuming that the probability of being at a particular state $\boldsymbol{z}$ at time t can be approximated as a finite sum of separable functions, i.e.,

$$P^N(\boldsymbol{z},t) = \sum_{j=1}^{N} F_1^j(z_1) \cdot F_2^j(z_2) \cdot \ldots \cdot F_D^j(z_D) \cdot F_t^j(t), \tag{3}$$

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 degrees of freedom of the problem, $N \times n_{\text{nod}} \times (D+1)$ instead of $(n_{\text{nod}})^D$, where D is the number of dimensions of the state space and n_{nod} the number of degrees of freedom 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 $[0,\ldots,m-1]^D$, 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 [15].

Another important point to be highlighted is the presence of a function depending solely on time, $F_t^j(t)$. 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. If one then assumes that n terms of the sum given by Eq. (3) are already known,

$$P^{n+1}(\boldsymbol{z},t) = P^n(\boldsymbol{z},t) + F_1^{n+1}(z_1) \cdot F_2^{n+1}(z_2) \cdot \ldots \cdot F_D^{n+1}(z_D) \cdot F_t^{n+1}(t), \tag{4}$$

and look for the $n+1$-th term, by substituting Eq. (4) into the CME, Eq. (1) gives a non-linear problem in $F_1^{n+1},\ldots,F_D^{n+1},F_t^{n+1}$ that is solved by means of a fixed-point, alternating directions algorithm, see [4] [5].

The separated representation just considered does not involve any assumption. Any solution defined in a high-dimensional space can be written, if it is regular enough, in a separated form if the number of terms in the finite sum

decomposition is high enough. A polynomial of any order is no more than a sum of functional products, each depending on a different coordinate. Thus, solutions can be approximated with the desired accuracy by using a separated representation and an adequate constructor as the one described above.

1.3. Moments-based descriptions

Even if the use of separated representations allows circumventing the curse of dimensionality the computational cost remains considerable. This fact motivated in many other disciplines the replacement of the pdf by some of its moments [8] [9] [1] [11], since many times the last suffice for having a view rich enough on the dynamics of the systems. The use of moment-based description was of major interest in different areas of statistical mechanics and it is being the more and more considered as an alternative to the discretization of the CME.

A moment represents the expected value of a random variable, z, raised to a certain power. An "expectation" is a specifically defined function in statistics, $E[f(z)] = \int f(z)P(z)dz$ when in continuous spaces or $\sum f(z)P(z)$ in discrete spaces. In general, we can talk about the ith moment as:

$$\mu_i(t) = E\left[z^i\right] = \sum_{z=0}^{\infty} P(z,t)z^i$$

A probability distribution is uniquely defined by its full set of moments. Having access to these moments could eliminate the need to solve for the full distribution, depending on what information would be considered important. A special function, called the Moment Generating Function, is specifically intended for this purpose:

$$M(\theta,t) = \sum_{z=0}^{\infty} e^{\theta z}P(z,t)$$

By taking the Taylor expansion of $e^{\theta z} = 1 + \frac{(\theta z)^1}{1!} + \frac{(\theta z)^2}{2!} + \frac{(\theta z)^3}{3!} + \cdots$, we can see the moments emerging from this function, the i-th moment associated with

the i-th power of θ:

$$M(\theta,t) = \mu_0(t) + \frac{\mu_1(t)\theta}{1!} + \frac{\mu_2(t)\theta^2}{2!} + \cdots = \sum_{z=0}^{\infty} \frac{\mu_z(t)\theta^z}{z!}$$

The following equations will be used extensively in the following derivation, so it will be useful to define them now:

$$M(\theta,t) = \sum_{z=0}^{\infty} e^{\theta z} P(z,t) = \sum_{z=0}^{\infty} \frac{\mu_z(t)\theta^z}{z!} \tag{5}$$

$$\frac{\partial M(\theta,t)}{\partial t} = \sum_{z=0}^{\infty} e^{\theta z} \frac{\partial P(z,t)}{\partial t} \tag{6}$$

$$\frac{\partial^i M(\theta,t)}{\partial \theta^i} = \sum_{z=0}^{\infty} e^{\theta z} P(z,t) z^i = \sum_{z=0}^{\infty} \frac{\mu_z(t)\theta^{z-i}}{(z-i)!} \tag{7}$$

2. From the Chemical Master Equation to moments based descriptions

Since we will be uniquely considering the structure of the Chemical Master Equation, we would like to derive a general version of the Moment Generating Function which can be used for any system. The CME for l reactions with stoichiometric change v_l is:

$$\frac{\partial P(z,t|z_0,t_0)}{\partial t} = \sum_l a_l(z - v_l)P(z - v_l,t) - a_l(z)P(z,t)$$

As we will see later on, the kind of rate laws associated with the system dramatically impact the complexity of the overall problem. We begin with the most simple case of kinetic mass action laws, following the derivation from [8]. However, we would eventually like to take rational rate laws, such as Hill functions, as is seen in [14]. An example of a mass action rate law is $a_l(z) = \frac{\lambda z_1(z_1-1)}{2} = \frac{\lambda}{2}z_1^2 - \frac{\lambda}{2}z_1 = \sum_i c_{l,i} a\prime_{l,i}$, where the law can be rewritten as a sum of coefficients $c_{l,i}$ and variables $a\prime_{l,i}$. This expanded, polynomial form will be exploited in our derivation.

Since we would like to talk about moments of the CME rather than probabilities, our first priority is to write this equation in terms of M, rather than in

terms of P. We multiply both sides by $e^{\theta z}$ and sum over all possible values of z:

$$\sum_{z=0}^{\infty} e^{\theta z} \frac{\partial P(z,t)}{\partial t} = \sum_{z=0}^{\infty} \sum_{l} e^{\theta z} a_l(z-v_l) P(z-v_l,t) - e^{\theta z} a_l(z) P(z,t)$$

that taking into account the previous definitions results

$$\begin{aligned}
\frac{\partial M(\theta,t)}{\partial t} &= \sum_{z=0}^{\infty} \sum_{l} \left(\sum_{i} c_{l,i} a\prime_{l,i}(z-v_l) e^{\theta z} P(z-v_l,t) - \sum_{i} c_{l,i} a\prime_{l,i}(z) e^{\theta z} P(z,t) \right) \\
&= \sum_{z=0}^{\infty} \sum_{l} \left(\sum_{i} c_{l,i} a\prime_{l,i}(z-v_l) e^{\theta(z-v_l)} e^{\theta(v_l)} P(z-v_l,t) - \sum_{i} c_{l,i} a\prime_{l,i}(z) e^{\theta z} P(z,t) \right) \\
&= \sum_{l} \left(\sum_{i} c_{l,i} \frac{\partial^i M}{\partial \theta^i} e^{\theta(v_l)} - \sum_{i} c_{l,i} \frac{\partial^i M}{\partial \theta^i} \right) = \sum_{l} \sum_{i} c_{l,i} \frac{\partial^i M}{\partial \theta^i} \left(e^{\theta(v_l)} - 1 \right)
\end{aligned}$$

Now, we can take the second definition of $\frac{\partial^i M}{\partial \theta^i}$ and expand $e^{\theta(v_l)}$ into its Taylor series. Notice that the summation now begins at $j = i$. When $j < i$, the index will be out of bounds and not correspond to any physical state

$$\begin{aligned}
\frac{\partial M(\theta,t)}{\partial t} &= \sum_{l} \sum_{i} c_{l,i} \frac{\partial^i M}{\partial \theta^i} \left(e^{\theta(v_l)} - 1 \right) \\
&= \sum_{l} \sum_{i} c_{l,i} \sum_{j=i}^{\infty} \frac{\mu_j(t) \theta^{j-i}}{(j-i)!} \left(\sum_{k=0}^{\infty} \frac{(\theta v_l)^k}{k!} - 1 \right)
\end{aligned}$$

Remember that the initial goal was to isolate the coefficients of θ^n in order to obtain the nth moments:

$$\begin{aligned}
\frac{\partial M(\theta,t)}{\partial t} &= \sum_{l} \sum_{i} c_{l,i} \sum_{j=i}^{\infty} \frac{\mu_j(t) \theta^{j-i}}{(j-i)!} \left(\sum_{k=0}^{\infty} \frac{(\theta v_l)^k}{k!} - 1 \right) \\
&= \sum_{l} \sum_{i} c_{l,i} \left(\frac{\mu_i}{0!} + \frac{\mu_{i+1}\theta}{1!} + \frac{\mu_{i+2}\theta^2}{2!} + \cdots \right) \left(\frac{v_l \theta}{1!} + \frac{(v_l\theta)^2}{2!} + \frac{(v_l\theta)^3}{3!} + \cdots \right) \\
&= \sum_{l} \sum_{i} c_{l,i} \left(\left[\frac{\mu_i}{0!} \frac{v_l}{1!} \right] \theta + \left[\frac{\mu_i}{0!} \frac{{v_l}^2}{2!} + \frac{\mu_{i+1}}{1!} \frac{v_l}{1!} \right] \theta^2 + \cdots \right) \\
&= \sum_{l} \sum_{i} c_{l,i} \sum_{n=0}^{\infty} \theta^n \sum_{k=1}^{n} \mu_{i+(n-k)} \frac{1}{k!(n-k)!}
\end{aligned}$$

Our next step will be isolate just the coefficients of θ in order to achieve a form in which we are creating ODE's of μ rather than $M(\theta,t)$. Since $\frac{\partial M(\theta,t)}{\partial t} =$

$\sum_n \frac{\partial \mu_n}{\partial t} \frac{1}{n!} \theta^n$, we will have to multiply both sides by $n!$ in order to isolate μ. Thus, it finally results

$$\frac{\partial \mu_n(t)}{\partial t} = \sum_l \sum_i c_{l,i} \sum_{k=1}^{n} v_l^k \mu_{i+(n-k)} \frac{n!}{k!(n-k)!} \tag{8}$$

3. Closures

It is easy to note from the previous expression (8) that the equation that governs the time evolution of the moments up to a certain order implies, in general, higher order moments, and then, before solving all them, higher order moments must be written as a combination of those involved in the considered time evolution equations. These relations have in most of cases an approximate character and are known as closure relations [13] [1]. Before describing the technique that we are following for defining such closures for a given system, we introduce some notation.

When considering multicomponent systems involving D components, the state becomes a vector $\mathbf{z}^T = (z_1, z_2, \cdots, z_D)$ as previously discussed. Now the first moment also becomes a vector $\boldsymbol{\mu}_1$, of size D, defined by

$$\boldsymbol{\mu}_1(t) = \sum \mathbf{z}\ P(\mathbf{z}, t) \tag{9}$$

The second order moment $\boldsymbol{\mu}_2$ results a $D \times D$ matrix

$$\boldsymbol{\mu}_2(t) = \sum \mathbf{z} \otimes \mathbf{z}\ P(\mathbf{z}, t) \tag{10}$$

$\boldsymbol{\mu}_3$ a third order tensor

$$\boldsymbol{\mu}_3(t) = \sum \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}\ P(\mathbf{z}, t) \tag{11}$$

and so on. These expressions involve many symmetries, e.g. $\boldsymbol{\mu}_2(i, j) = \boldsymbol{\mu}_2(j, i)$.

In what follows and without loss of generality, we consider reactions involving linear propensities. Thus, when considering the equations governing the time evolution of the first two moments $\boldsymbol{\mu}_1(t)$ and $\boldsymbol{\mu}_2(t)$ the third order moment $\boldsymbol{\mu}_3(t)$ remains in these equations, and it needs to be expressed from both lower moments.

The simplest closure writes:

$$\boldsymbol{\mu}_3 = \alpha_1 \mathbf{I} \otimes \mathbf{I} \otimes \mathbf{I} + \alpha_2 \mathcal{S}(\boldsymbol{\mu}_1, \mathbf{I}, \mathbf{I}) + \alpha_3 \mathcal{S}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_1, \mathbf{I}) + \alpha_4 \boldsymbol{\mu}_1 \otimes \boldsymbol{\mu}_2 \otimes \boldsymbol{\mu}_1 +$$

$$\alpha_5 \mathfrak{S}(\boldsymbol{\mu}_2, \mathbf{I}) + \alpha_6 \mathfrak{S}(\boldsymbol{\mu}_2, \boldsymbol{\mu}_1) \tag{12}$$

where

$$\mathcal{S}(\boldsymbol{\mu}_1, \mathbf{I}, \mathbf{I}) = \boldsymbol{\mu}_1 \otimes \mathbf{I} \otimes \mathbf{I} + \mathbf{I} \otimes \boldsymbol{\mu}_1 \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{I} \otimes \boldsymbol{\mu}_1,$$

$$\mathcal{S}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_1, \mathbf{I}) = \boldsymbol{\mu}_1 \otimes \boldsymbol{\mu}_1 \otimes \mathbf{I} + \boldsymbol{\mu}_1 \otimes \mathbf{I} \otimes \boldsymbol{\mu}_1 + \mathbf{I} \otimes \boldsymbol{\mu}_1 \otimes \boldsymbol{\mu}_1,$$

$$\mathfrak{S}(\boldsymbol{\mu}_2, \mathbf{I}) = \boldsymbol{\mu}_2 \otimes \mathbf{I} + \mathbf{I} \otimes \boldsymbol{\mu}_2$$

and

$$\mathfrak{S}(\boldsymbol{\mu}_2, \boldsymbol{\mu}_1) = \boldsymbol{\mu}_2 \otimes \boldsymbol{\mu}_1 + \boldsymbol{\mu}_1 \otimes \boldsymbol{\mu}_2$$

Thus, the third order closure relation (12) involves 6 coefficients to be determined. For this purpose, and for a given system, we solve the CME by using the PGD in order to circumvent the curse of dimensionality and then evaluate the third order moment according to Eq. (11) and then we choose the alpha parameters in (12) to provide the best fitting (in a least squares sense).

As soon as the alpha parameters are empirically fitted, the CME is substituted by the two ordinary differential equations governing the time evolution of $\boldsymbol{\mu}_1(t)$ and $\boldsymbol{\mu}_2(t)$, when considering the solution of the same systems for any other initial condition or slightly different kinetic rates.

4. Numerical results

4.1. Lotka model

Fist, we consider the so-called Lotka model. This model consists of:

$$A + X \xrightarrow{\lambda_1} 2X$$

$$X + Y \xrightarrow{\lambda_2} 2Y$$

$$Y \xrightarrow{\lambda_3} B$$

where the number of molecules of species A and B are enforced to be constant.

We consider the two simulation cases:

- *Case 1.* At the initial time $t = 0$, the state of the systems consists of $\mathbf{z}_0^T = (\#X_0, \#Y_0) = (120, 50)$, the reaction rates being $\lambda_1 = 1$, $\lambda_2 = 0.012$ and $\lambda_3 = 1$.

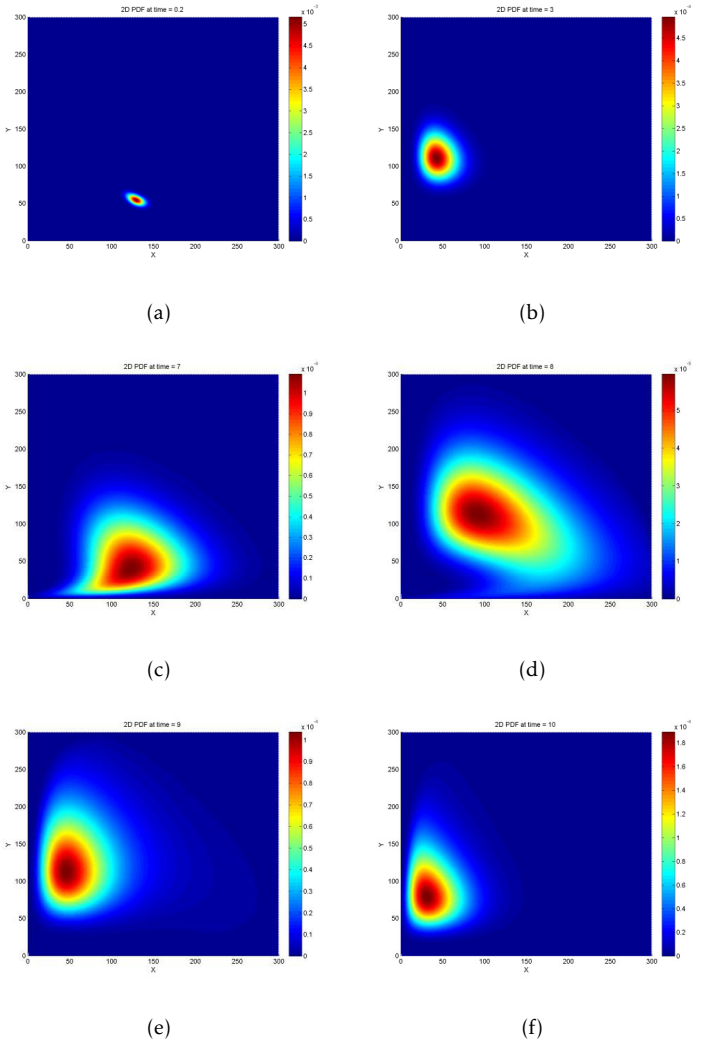
- *Case 2.* At the initial time $t = 0$, the state of the systems consists of $\mathbf{z}_0^T = (\#X_0, \#Y_0) = (60, 60)$, the reaction rates being $\lambda_1 = 1$, $\lambda_2 = 0.008$ and $\lambda_3 = 1$.

First, we solve the CME with the *Case 1* conditions. The probability distribution function at 6 different times is depicted in Fig. 1.

Now from the pdf $P(\mathbf{z}, t)$, $\mathbf{z}^T = (\#X, \#Y)$ and $t = [0, 10]$, we compute the three first moments $\boldsymbol{\mu}_1(t)$, $\boldsymbol{\mu}_2(t)$ and $\boldsymbol{\mu}_3(t)$ respectively from Eqs. (9-11) that will be considered as reference moment solutions. The parameters alpha involved in the empirical closure relation (12) are then determined. In the present case, and taking into account the symmetries, $\boldsymbol{\mu}_1$ is of size two, $\boldsymbol{\mu}_2$ has three independent components and $\boldsymbol{\mu}_3$ four. Fig. 2 shows the different independent components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Figure 3 compares the reference third moment with the one fitted with the empirical closure relation, from which we can conclude that both are in perfect agreement.

Now, by integrating in time the equations governing the time evolution of the components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, using the just identified empirical closure relation, we obtained the curves depicted in Fig. 4 that are very close to those obtained from the probability distribution function that were depicted in Fig. 2.

Now, with the closure relation obtained from the analysis of *Case 1* we are addressing *Case 2* without modifying the closure relation. For that purpose the equations governing the time evolution of the different components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are integrated in time by considering the closure relation fitted in *Case 1*. In order to check the accuracy of those solutions we solve again the CME and compute the reference moments from the resulting probability distribution function. Figure 5 compares the moment-based and the pdf-based moments. Even if non negligeable deviations in the second order moment are noticed at

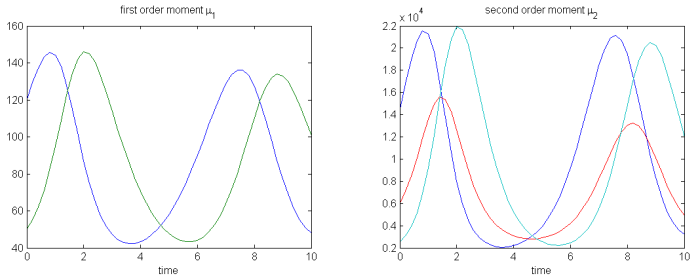


(a) (b)

(c) (d)

(e) (f)

Figure 1: Probability distribution function $P(\mathbf{z}, t)$ at different times: (a) $t = 0.2$, (b) $t = 3$, (c) $t = 7$, (d) $t = 8$, (e) $t = 9$ and (f) $t = 10$



(a) (b)

Figure 2: Time evolution of the components of $\boldsymbol{\mu}_1$ (left) and $\boldsymbol{\mu}_2$ (right)

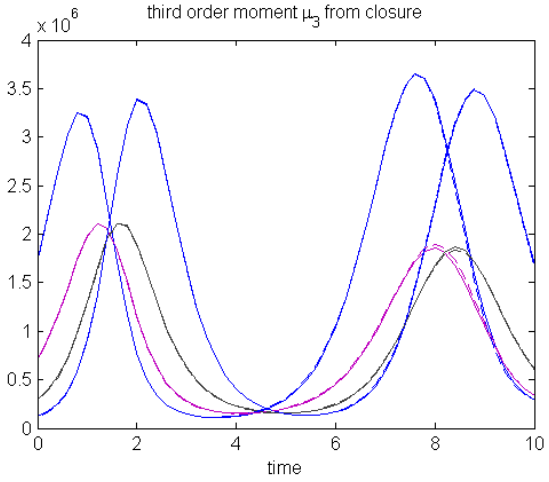
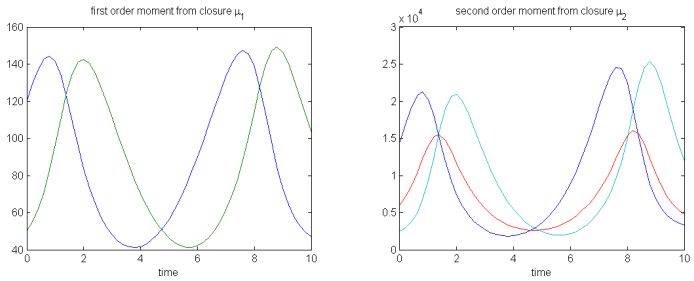


Figure 3: Time evolution of the third order moment $\boldsymbol{\mu}_3(t)$: reference components (continuous line) versus closure-based approximation (broken line)



(a) (b)

Figure 4: Time evolution of the components of $\boldsymbol{\mu}_1$ (left) and $\boldsymbol{\mu}_2$ (right) obtained from the moment-based description

the final time, results are qualitatively quite good.

4.2. Exclusive switch 5D model

We consider a gene regulatory network called exclusive switch. It describes the dynamics of two genes with an overlapping promoter region, and the corresponding proteins X_1 and X_2. Both X_1 and X_2 are produced if no transcription factor is bound to the promoter region. However if a molecule of type X_1 (X_2) is bound to the promotor then it inhibits the expression of the other protein, i.e. molecules of type X_2 (X_1) can not be produced. Only one molecule can be bound to the promotor region at a time which gives three possibilities for the state of the promoter region (free, X_1 bound, X_2 bound). The model is infinite in two dimensions (X_1 and X_2) and the reactions are given by:

$$X_5 \xrightarrow{\lambda_1} X_5 + X_1$$

$$X_5 \xrightarrow{\lambda_2} X_5 + X_2$$

$$X_1 \xrightarrow{\lambda_3} 0$$

$$X_2 \xrightarrow{\lambda_4} 0$$

$$X_1 + X_5 \xrightarrow{\lambda_5} X_3$$

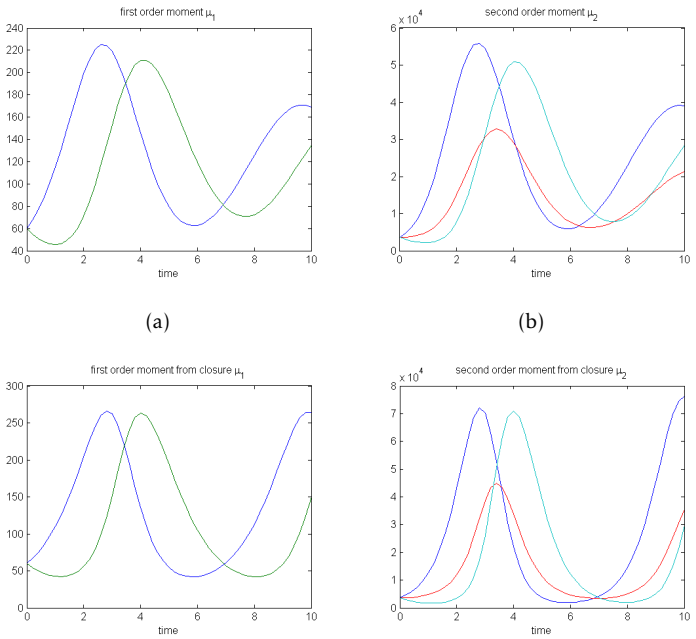


Figure 5: Components of $\boldsymbol{\mu}_1$ (left) and $\boldsymbol{\mu}_2$ (right), computed from the pdf-based description (top) and the moment-based description (bottom)

$$X_2 + X_5 \xrightarrow{\lambda_6} X_4$$
$$X_3 \xrightarrow{\lambda_7} X_5 + X_1$$
$$X_4 \xrightarrow{\lambda_8} X_5 + X_2$$
$$X_3 \xrightarrow{\lambda_9} X_3 + X_1$$
$$X_4 \xrightarrow{\lambda_{10}} X_4 + X_2$$

where $(\lambda_1, \lambda_2, \cdots, \lambda_{10}) = (2, 5, 0.005, 0.005, 0.005, 0.002, 0.02, 0.02, 2, 5)$. The initial conditions are such that only one DNA molecule is present in the system while the molecular counts for the rest of the species are zero.

First, we solve the CME. The probability distribution function at 6 different times is depicted in Fig. 6.

Now from the pdf $P(\mathbf{z}, t)$, $\mathbf{z}^T = (\#X_1, \cdots, \#X_5)$ and $t = [0, 60]$, we compute the three first moments $\boldsymbol{\mu}_1(t)$, $\boldsymbol{\mu}_2(t)$ and $\boldsymbol{\mu}_3(t)$ respectively from Eqs. (9-11) that will be considered as reference moment solutions. The parameters alpha involved in the empirical closure relation (12) are then determined. Fig. 7 shows the different independent components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Figure 8 depicts the fitted third moment that makes use of the empirical closure relation.

With the closure relation just identified we are addressing a new scenario consisting of a different initial condition $\mathbf{z}^T(t = 0) = (\#X_1^0, \cdots, \#X_5^0) = (50, 100, 0, 0, 1)$ that produces at time $t = 10$ the pdf depicted in figure 9. Now, the equations governing the time evolution of the different components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are integrated in time by considering the closure relation just identified fitted. In order to check the accuracy of those solutions we solve again the CME and compute the reference moments from the resulting probability distribution function. Figure 10 compares the moment-based and the pdf-based moments, proving that the moment approach based on the use of an empirical closure produces a quite reasonable agreement.

4.3. The toggle

Mutually repressing gene pair, or gene toggle, can be found in biological systems as discussed in [12]. As in their example, here we focus on protein

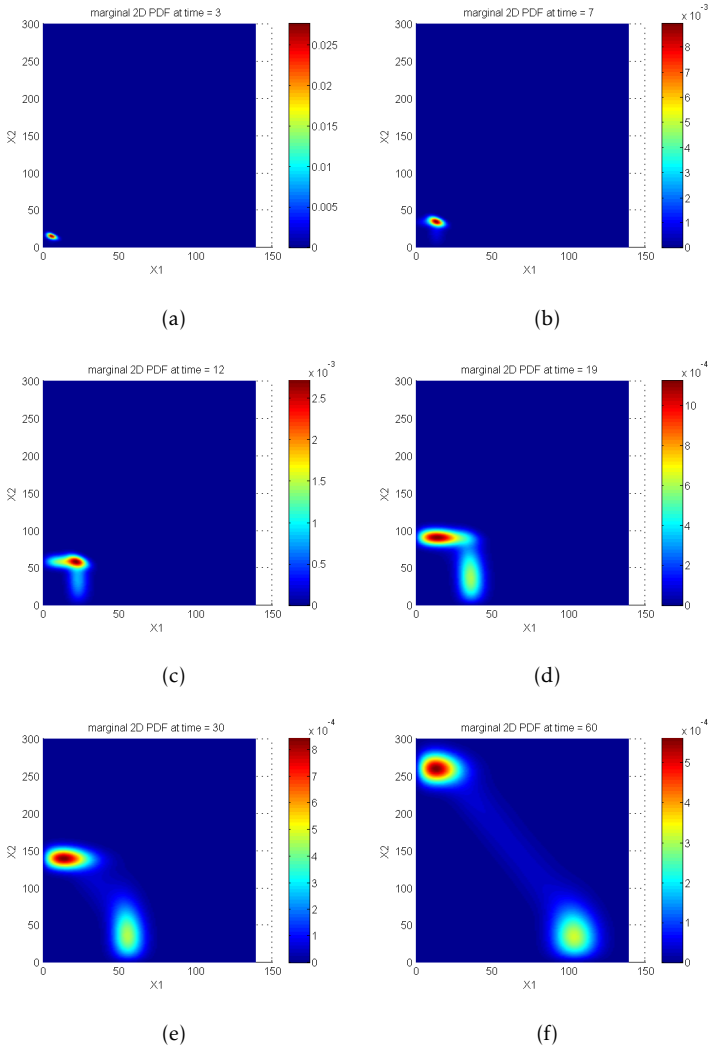
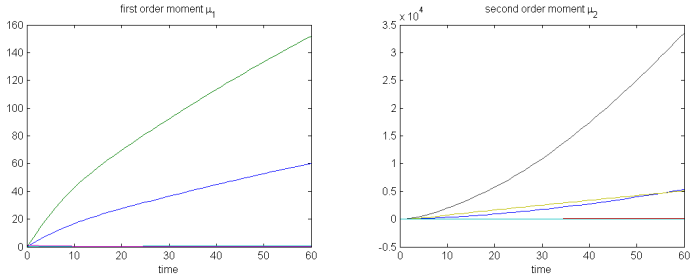


Figure 6: Probability distribution function $P(\mathbf{z}, t)$ at different times: (a) $t = 3$, (b) $t = 7$, (c) $t = 12$, (d) $t = 19$, (e) $t = 30$ and (f) $t = 60$



(a) (b)

Figure 7: Time evolution of the components of $\boldsymbol{\mu}_1$ (left) and $\boldsymbol{\mu}_2$ (right)

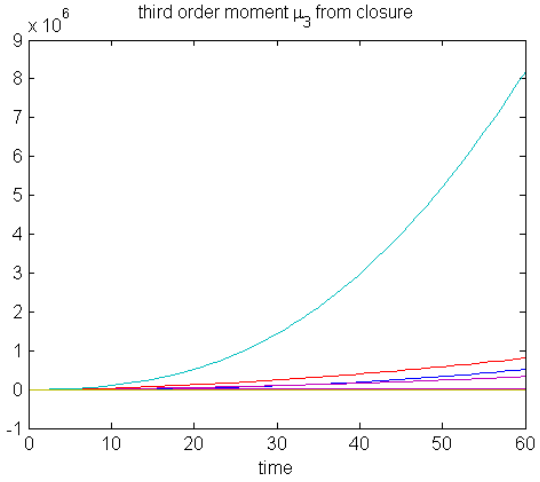


Figure 8: Time evolution of the third order moment $\boldsymbol{\mu}_3(t)$

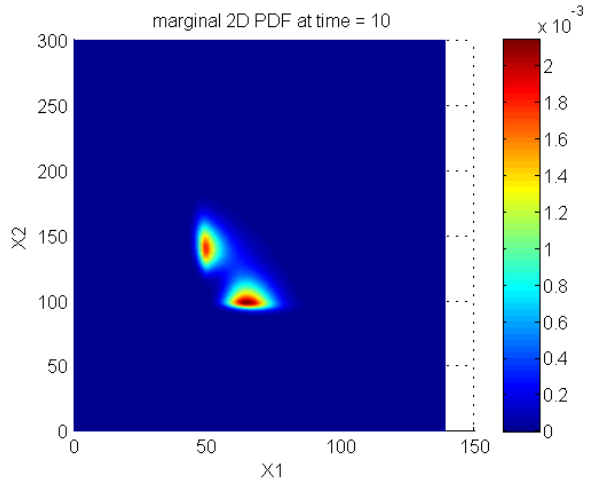
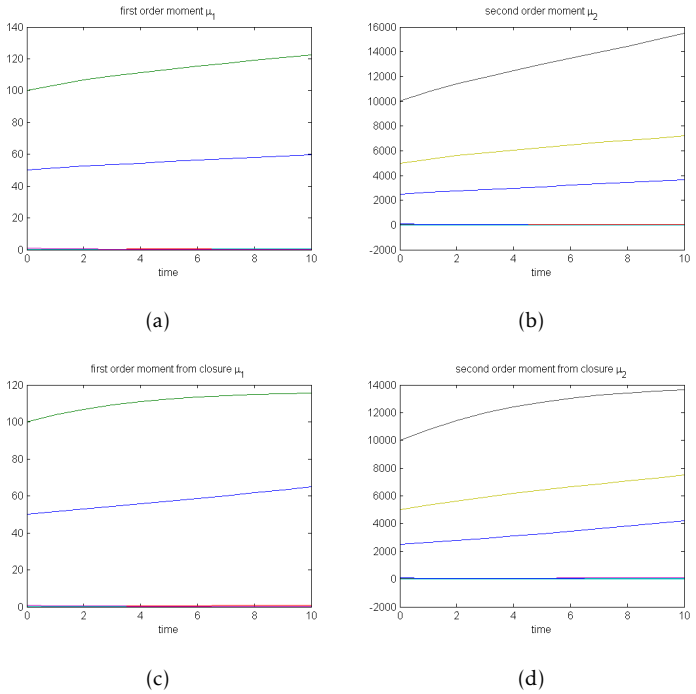


Figure 9: probability distribution function at time $t = 10$



(a) (b)

(c) (d)

Figure 10: Components of $\boldsymbol{\mu}_1$ (left) and $\boldsymbol{\mu}_2$ (right), computed from the pdf-based description (top) and the moment-based description (bottom)

dynamics, and more particularly in the toggle-switch network. The reactions consist of:

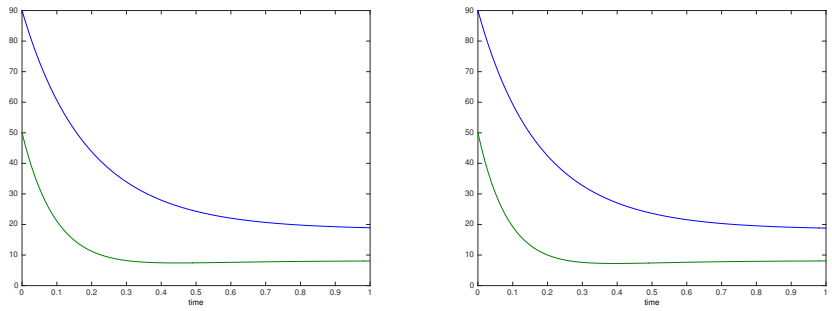
$$X_2 \xrightarrow{\lambda_1} X_1 + X_2$$
$$X_1 \xrightarrow{\lambda_2} 0$$
$$X_1 \xrightarrow{\lambda_3} X_1 + X_2$$
$$X_2 \xrightarrow{\lambda_4} 0$$

and the following polynomial properties are considered:

$$\begin{cases} a_1(X_1, X_2) = \lambda_1(A - X_2) \\ a_2(X_1, X_2) = \lambda_2 X_1 \\ a_3(X_1, X_2) = \lambda_3(A - X_1) \\ a_4(X_1, X_2) = \lambda_4 X_1 \end{cases} \tag{13}$$

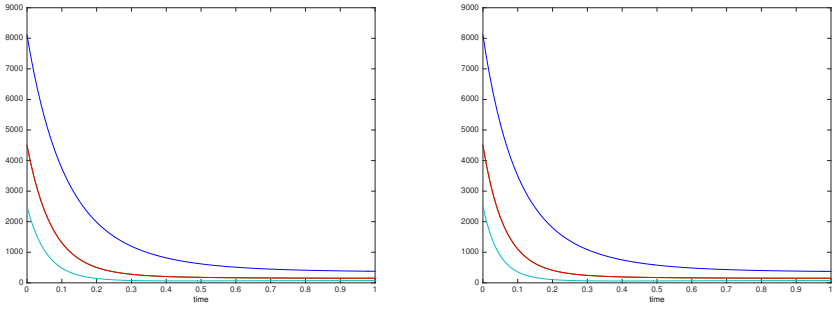
Now, we consider the initial condition $(\#X_1^0, \#X_2^0) = (90, 50)$ as well as the parameters $\lambda_1 = 1$, $\lambda_2 = 5$, $\lambda_3 = 1$ and $\lambda_4 = 10$, and solve the associated chemical master equation for calculating the joint probability distribution function $P(\mathbf{z}, t)$. Now from the pdf $P(\mathbf{z}, t)$, $\mathbf{z}^T = (\#X_1, \#X_2)$ and $t = [0, 1]$, we compute the three first moments $\boldsymbol{\mu}_1(t)$, $\boldsymbol{\mu}_2(t)$ and $\boldsymbol{\mu}_3(t)$ respectively from Eqs. (9-11) that will be considered as reference moment solutions. The parameters alpha involved in the empirical closure relation (12) are then determined. Figs. 11, 12 and 13 compare respectively the different independent components of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ calculated from chemical master equation solution and from the closure-based description. A very good agreement can be noticed.

With the closure relation just identified we are addressing a new scenario consisting of a different initial condition $\mathbf{z}^T(t = 0) = (\#X_1^0, \#X_2^0) = (1, 1)$. Now, the equations governing the time evolution of the different components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are integrated in time by considering the closure relation just identified fitted. The third moment is calculated from the fitted closure from the first two moments. In order to check the accuracy of those solutions we solve again the CME and compute the reference moments from the resulting probability



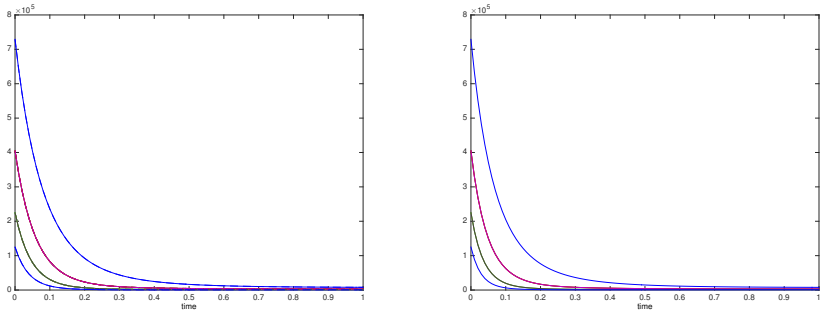
(a) (b)

Figure 11: Time evolution of the components of $\boldsymbol{\mu}_1$ calculated from the CME solution (left) and from the closure-based description (right)



(a) (b)

Figure 12: Time evolution of the components of $\boldsymbol{\mu}_2$ calculated from the CME solution (left) and from the closure-based description (right)



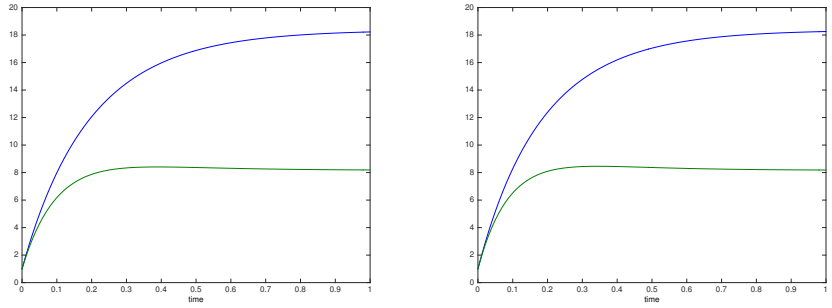
(a) (b)

Figure 13: Time evolution of the components of $\boldsymbol{\mu}_3$ calculated from the CME solution (left) and from the closure-based description (right)

distribution function. Figs. 14, 15 and 16 compare respectively the different independent components of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ calculated from chemical master equation solution and from the closure-based description. Again a very good agreement can be noticed.

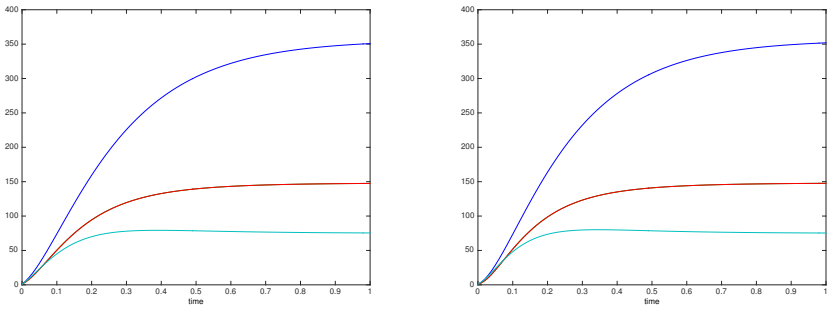
5. Conclusions

In this work we revisited the modeling of regulatory networks described within the chemical master equation framework. Deterministic solutions of the CME were performed by using the separated representation involved in the PGD, allowing circumventing the so-called curse of dimensionality. In order to improve the computational efficiency a moment-based description is derived, however such description involves higher order moments that must be described from the ones of lower order. In that sense we proposed a simplest closure relation, of empirical nature, that can be fitted numerically from the probability distribution function, the last coming from the PGD solution of the CME. As soon as the closure relation is fitted, it can be used for solving similar problems to the one that served to identify the closure relation.



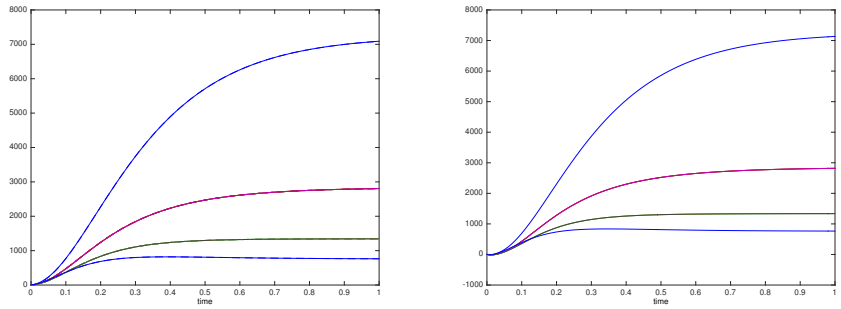
(a) (b)

Figure 14: Time evolution of the components of $\boldsymbol{\mu}_1$ calculated from the CME solution (left) and from the closure-based description (right)



(a) (b)

Figure 15: Time evolution of the components of $\boldsymbol{\mu}_2$ calculated from the CME solution (left) and from the closure-based description (right)



(a) (b)

Figure 16: Time evolution of the components of $\boldsymbol{\mu}_3$ calculated from the CME solution (left) and from the closure-based description (right)

[1] A. Andreychenko, L. Mikeev, V. Wolf. Reconstruction of multimodal distributions for hybrid moment-based chemical kinetics. Journal of Coupled Systems and Multiscale Dynamics, 2014, arXiv:1410.3267

[2] A. Ammar, B. Mokdad, F. Chinesta, R. Keunings. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. J. Non-Newtonian Fluid Mech., 139, 153-176, 2006.

[3] A. Ammar, B. Mokdad, F. Chinesta, R. Keunings. 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. J. Non-Newtonian Fluid Mech., 144, 98121, 2007.

[4] A. Ammar, E. Cueto, F. Chinesta. Reduction of the Chemical Master Equation for gene regulatory networks using Proper Generalized Decompositions. International Journal for Numerical Methods in Biomedical Engineering, 28/9, 960-973, 2012.

[5] F. Chinesta, M. Magnin, O. Roux, A. Ammar, E. Cueto. Kinetic theory

modeling and efficient numerical simulation of gene regulatory networks based on qualitative descriptions. Entropy, 17/4, 1896-1915, 2015.

[6] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81/25, 23402361, 1977.

[7] D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. The Journal of Chemical Physics, 115, 1716-1733, 2001.

[8] C.S. Gillespie. Moment-closure approximations for mass-action models. IET Syst. Biol., 3/1, 52-58, 2009.

[9] A. Gupta, C. Briat, M. Khammash. (2014) A Scalable Computational Framework for Establishing Long-Term Behavior of Stochastic Reaction Networks. PLoS Comput Biol., 10/6, e1003669, 2014

[10] J. Hasty, D. McMillen, F. Isaacs, J.J. Collins. Computational studies of gene regulatory networks: in numero molecular Biology. Nature Reviews Genetics, 2, 268-279, 2001.

[11] J. Hasenauer, V. Wolf, A. Kazeroonian, F.J. Theis. Method of conditional moments (MCM) for the Chemical Master Equation. Journal of Mathematical Biology, Volume 69/3, 687-735, 2014.

[12] M. Hegland, C. Burden, L. Santodo, S.MacNamara, H. Booth. A solver for the stochastic master equation applied to gene regulatory networks. Journal of Computational and Applied Mathematics, 205/2, 708-724, 2007.

[13] M. Kroger, A. Ammar, F. Chinesta. Consistent closure schemes for statistical models of anisotropic fluids. Journal of Non-Newtonian Fluid Mechanics, 149, 40-55, 2008.

[14] P. Milner, C.S. Gillespie, J.W. Darren. Moment closure approximations for stochastic kinetic models with rational rate laws. Mathematical Biosciences, 231/2, 99-104, 2011.

[15] B. Munsky, M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. The Journal of Chemical Physics, 124/4, 044104, 2006.

[16] S.N. Sreenath, K.H. Cho, P. Wellstead. Modelling the dynamics of signaling pathways. Essays in biochemistry, 45, 1-28, 2008.