Robust Moment Closure Method for the Chemical Master Equation

Mohammad Naghnaeian and Domitilla Del Vecchio

***Abstract*— The Chemical Master Equation (CME) is used to stochastically model biochemical reaction networks, under the Markovian assumption. The low-order statistical moments induced by the CME are often the key quantities that one is interested in. However, in most cases, the moments equation is not closed; in the sense that the first n moments depend on the higher order moments, for any positive integer n. In this paper, we develop a moment closure technique in which the higher order moments are approximated by an affine function of the lower order moments. We refer to such functions as the affine Moment Closure Functions (MCF) and prove that they are optimal in the worst-case context, in which no a priori information on the probability distribution is available. Furthermore, we cast the problem of finding the optimal affine MCF as a linear program, which is tractable. We utilize the affine MCFs to derive a finite dimensional linear system that approximates the low-order moments. We quantify the approximation error in terms of the l_∞ induced norm of some linear system. Our results can be effectively used to approximate the low-order moments and characterize the noise properties of the biochemical network under study.**

I. Introduction

Biomolecular reaction networks are mostly studied in two frameworks, deterministic or stochastic [1]. In the former, the system is modeled by a set of Ordinary Differential Equations (ODEs) whose states represent the concentration of the species. Such models have proved to be useful in explaining and predicting the behavior of the system especially in the high concentration regime. They, however, fail to accurately explain the characteristics of the system in the low concentration regime [2]. In fact, when the number of molecules in the network is low, the inherent randomness in the interactions and the discreteness of the system's state play an important role towards the overall behavior [3]. This necessitates the use of stochastic models.

In stochastic framework, the Chemical Master Equation (CME) is used to model biochemical reaction networks [4]. It is a popular modeling framework in the systems biology community, in which it has been widely employed to study the impact of intrinsic noise on a network's behavior and to capture the behavior of networks characterized by low molecule counts [5]. Although the CME is a linear system, its explicit solution cannot be obtained, in general. This is due to the fact that, except in very idealistic situations, the dimension of the CME is very large or often infinite. A reasonable approach to cope with this curse of dimensionality is to study the statistical moments.

Low-order statistical moments, particularly the first and second moments, are often the key quantities that one is interested in as they provide indication on standard noise quantifications, such as the coefficient of variation. One difficulty that arises in this approach is that, in most cases, the moments equation induced by the CME is not closed [6]; in the sense that the first n moments depend on the higher order moments, for any positive integer n. This is challenging since the low-order statistical moments of the system cannot be studied without knowing the higher-order moments due to this coupling. For analysis and simulation purposes, one can close the system of moments by approximating the higher-order moments. In the literature, such a procedure is referred to as moment closure. Any moment closure technique consist of two steps [6]:

(a) The statistical moments higher than n are approximated as a function (possibly nonlinear) of the first n moments. This function is called the Moment Closure Function (MCF).

(b) The high-order moments in the low-order moments equations are replaced by the MCF. This results in a closed system for the low-order moments.

There are various moment closure methods proposed in the literature. Most of them assume an underlying probability distribution. For example, in [7], [8], and [9] the probability distribution is assumed to be normal, log-normal, and beta-binomial, respectively. There are also techniques that are not distribution based. For instance, [10] uses cumulant truncation and [11] uses the derivative-matching. Upon utilizing any moment closure method, the resulting closed system serves as an approximation to the low-order moments. Hence, it is important to quantify the error of this approximation. To the best of authors knowledge, no such quantifications are available in the literature. Accordingly, the development of a moment closure method with quantifiable error bounds deems necessary and this is what this paper aims to address.

In this paper, we develop the Robust Moment Closure (RMC) method for which we can exactly quantify the approximation error. In this method the higher order moments are approximated by an affine function of the lower order moments. We mathematically prove that affine MCFs are optimal in the worst-case context, in which we do not have a priori information on the probability distribution. In this case, no (possibly nonlinear) MCF can outperform the affine ones. We show that finding the optimal affine MCF is a

M. Naghnaeian is a postdoctoral associate with the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, USA mongh@mit.edu

D. Del Vecchio is with the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, USA ddv@mit.edu

This work was supported by the Air Force Office of Scientific Research under grant FA9550-14-1-0060

Linear Program (LP) and hence tractable [12]. Consequently, utilizing the affine MCFs, we derive a set ODEs of finite dimension that approximates the time evolution of the lower order moments. Furthermore, we quantify the error in this approximation in terms of the l_∞ induced norm of some linear system. Our results allow for the explicit simulation and analytical computation of approximate moments, which can be used to characterize the noise properties of the biomolecular reaction networks. To show the utility of this approach, we consider an academic birth-death process and an application example. In these examples, we approximate the low order moments, which in turn can be used to compute various dispersion indices [13]. The proofs of the results are given in the Appendix.

II. Preliminaries

The following notations are used throughout this paper: $\mathbb{Z}_{\geq 0}$ and $\mathbb{R}_{\geq 0}$ is the set of nonnegative integers and real numbers, respectively. For a positive integer n, $\mathbb{Z}_{\geq 0}^n$ ($\mathbb{R}_{\geq 0}^n$) denotes the set of n-dimensional vectors with entries in $\mathbb{Z}_{\geq 0}$ ($\mathbb{R}_{\geq 0}^n$). Given an n-dimensional vector $X = [x_1, x_2, ..., x_n]^T$ and a nonnegative integer I, define $\Psi_I(X)$ to be the vector composed of entries of the form $x_1^{k_1} x_2^{k_2} ... x_n^{k_n}$ where $k_i \in \mathbb{Z}_{\geq 0}$, for $i = 1, 2, ..., n$, and $\sum_{i=1}^n k_i = I$. For example, for $X = [x_1, x_2]^T$, we have

$$\begin{aligned} \Psi_1(X) &= [x_1, x_2]^T, \Psi_2(X) = \left[x_1^2, x_1 x_2, x_2^2\right]^T, \\ \Psi_3(X) &= \left[x_1^3, x_1^2 x_2, x_1 x_2^2, x_2^3\right]^T. \end{aligned}$$

Also, define $\Psi_0(X) = 1$. Given a positive integer n, define the vector

$$\bar{\Psi}_n(X) = \left[\Psi_1^T(X), \Psi_2^T(X), ..., \Psi_n^T(X)\right]^T, \quad (1)$$

and, for $i = 1, 2, ..., n$, let c_i be a matrix whose multiplication with $\bar{\Psi}_n(X)$ isolates $\Psi_i(X)$, i.e.

$$\Psi_i(X) = c_i \left[\Psi_1^T(X), \Psi_2^T(X), ..., \Psi_n^T(X)\right]^T. \quad (2)$$

The l_∞ and l_1 norms of a vector $X = [x_1, x_2, ..., x_n]^T$ are defined as $\|X\|_{l_\infty} := \max_i |x_i|$ and $\|X\|_{l_1} = \sum_{i=1}^n |x_i|$. We use $\|X\|$ without any subscript to mean $\|X\|_{l_\infty}$. A vector $P \in \mathbb{R}_{\geq 0}^p$ is called a probability vector if $\|P\|_{l_1} = 1$. The set of all probability vectors with dimension p is denoted by $\mathbb{P}^p$. We omit the superscript p when the dimension is irrelevant or obvious from the context. Given a matrix $M = [m_{ij}] \in \mathbb{R}^{m \times n}$, by $\mathcal{R}[M]_i$ and $\mathcal{C}[M]_j$ we mean the i^{th} row and j^{th}column of M, respectively. That is,

$$\begin{aligned} \mathcal{R}[M]_i &= \begin{bmatrix} m_{i1} & m_{i2} & \cdots & m_{in} \end{bmatrix}, \\ \mathcal{C}[M]_j &= \begin{bmatrix} m_{1j} & m_{2j} & \cdots & m_{mj} \end{bmatrix}^T, \end{aligned}$$

for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$. The l_1, l_∞, and l_1 to l_∞ induced norms of M are defined as $\|M\|_{l_1 - ind} = \max_j \sum_{i=1}^m |m_{ij}|$, $\|M\|_{l_\infty - ind} = \max_i \sum_{j=1}^n |m_{ij}|$, and $\|M\|_{l_1 - l_\infty} = \max_{i,j} |m_{ij}|$. Furthermore, the null space of M and its perpendicular complement (perp) are denoted respectively by $\mathcal{N}(M)$ and $\mathcal{N}^\perp(M)$ and defined as

$$\begin{aligned} \mathcal{N}(M) &= \{x : Mx = 0\}, \\ \mathcal{N}^\perp(M) &= \left\{y : y^T x = 0, \forall x \in \mathcal{N}(M)\right\}. \end{aligned}$$

Also, define $N(M)$ and $N^\perp(M)$ to be matrices whose columns form orthonormal basis for $\mathcal{N}(M)$ and $\mathcal{N}^\perp(M)$, respectively. The following lemma holds:

Lemma 1: Given a matrix $M = [m_{ij}] \in \mathbb{R}^{m \times n}$, we have

$$\sup_{P \in \mathbb{P}} \|MP\| = \|M\|_{l_1 - l_\infty}.$$

Markov processes can be used to describe the dynamics of chemical reaction networks. Each state of this Markov process represents the aggregated molecule counts of the species. A transition from one state to another state occurs when a chemical reaction fires and, as a result, the molecule counts of species change. More precisely, suppose a reaction network with q number of species and J number of reactions. Let s_i, for $i = 1, 2, ..., q$, be the count of each species and let $S = [s_1, s_2, ..., s_q]^T$. Associated with each reaction $j \in \{1, 2, .., J\}$, there are a propensity function $a_j(t, S)$ and a stoichiometry vector γ_j defined as

$$\Pr\left(S(t + dt) = S(t) + \gamma_j | S(t)\right) = a_j(t, S(t))\, dt + O\left(dt^2\right),$$

with γ_j is the change in species count upon firing of reaction j. In this case, for any $k \in \mathbb{Z}_{\geq 0}^q$, the probability vector satisfies

$$\begin{aligned} \frac{d}{dt} \Pr(S(t) = k) = \sum_{j=1}^J \Big\{ &-a_j(t, k) \Pr(S(t) = k) \\ &+ a_j\left(t, k - \gamma_j\right) \Pr\left(S(t) = k - \gamma_j\right) \Big\}. \end{aligned} \quad (3)$$

This equation is referred to as the Chemical Master Equation [14][15]. Throughout this paper, we make the following assumptions.

Assumption 2: There exist nonnegative integers U_i such that

$$0 \leq s_i \leq U_i,$$

for $i = 1, 2, ..., q$.

Assumption 3: The propensity functions are polynomial in S [16][15]. That is, for $j = 1, 2, ..., J$,

$$a_j(t, S) = \sum_{i=0}^l \theta_i^j(t) \Psi_i(S),$$

for some $l \in \mathbb{Z}_{\geq 0}$, where $\theta_i^j(t)$'s are matrices with appropriate dimensions.

Assumption 2 states that we have an upper bound on the number of molecules for each species. This assumption is readily satisfied for species that are conserved in the biochemical reaction network, such as DNA copy number or total protein concentrations in enzymatic reactions [4]. In the presence of species that are not conserved, one can use the methods given e.g. in [17] to truncate the system and find an upper bound on the species count such that the

truncated (finite dimensional) system is arbitrarily close to the infinite dimensional CME. Regarding Assumption 3, we refer the reader to [16], [15], and [18] where the polynomial propensity functions are derived under suitable conditions such as well-mixedness.

III. Basic Setup

Consider the CME given in (3) with Assumptions 2 and 3. This is a linear system of ODEs describing the time evolution of the probability distribution vector of the underlying Markov process. Based on Assumption 2, the CME is of order p, where

$$p := \prod_{i=1}^{q} (1 + U_i). \tag{4}$$

Clearly, the order of a CME grows exponentially with respect to the number of species present in the system. In most cases, the CME is a high dimensional system and hence solving it is a computationally challenging task. Thus, instead of directly solving the CME, one can consider the low-order statistical moments. While the statistical moments of a probability distribution are informative quantities to consider, they contain less information than the distribution itself. Hence, intuitively, one hopes for a less complex problem if only the moments are considered. In the next proposition we derive the moments equation induced by the CME in (3).

We denote the i^{th} moment of the random variable S by μ_i. Recall that $\mu_i := \mathbb{E}[\Psi_i(S)] = \sum_{k\in\mathbb{Z}^q} \Psi_i(k) \Pr(S = k)$, where $\Psi_i(S)$ is a vector composed of the entries of the form $s_1^{k_1} s_2^{k_2} ... s_q^{k_q}$ with $k_1 + k_2 + ... + k_q = i$. The following proposition holds (see e.g. [6] for the proof):

Proposition 4: For the chemical master equation in (3) with Assumptions 2 and 3, for $i = 1, 2, ...,$

$$\frac{d}{dt}\mu_i(t) = \beta_{i,0}(t) + \sum_{n=1}^{i+l-1} \beta_{i,n}\mu_n(t), \tag{5}$$

with initial condition $\mu_i(0) = \Psi_i(S(0))$, for some properly defined matrices $\beta_{i,n}(.)$ with appropriate dimension.

When $l > 1$, the system of moments in (5) is not closed in the sense that the lower-order moments depend on the higher-order moments. This introduces a certain degree of complexity into the system. More precisely, one cannot consider the low-order moments decoupled from the high-order ones. Therefore, one needs to study the full system of moments including all the moments up to order p, where p is defined in 4 and generally is very large as it scales exponentially with respect to the number of species. Therefore, the full system of moments up to order p, although closed, but is a high dimensional set of ODEs whose study is as difficult as that of CME. Therefore, in the literature, there has been a great deal of effort to approximate the higher-order moments, $(\mu_{n+1}, \mu_{n+2}, ..., \mu_{n+l})$, by a possibly nonlinear function of low-order moments. This procedure is referred to as moment closure. Unfortunately, the lack of error quantification prevails amongst the moment closure methods. In the next section, we introduce the *Robust Moment Closure* technique for which we exactly quantify the error between the true system (5) and the resulting closed system of moments.

IV. Robust Moment Closure

Any moment closure method revolves around the idea of approximating the higher order moments by a possibly nonlinear function of lower order moments. This allows for closing the system of moments which in turn can be more easily analyzed. In this section, we first discuss on the optimal MCF in the worst-case setting; that is, when no a priori information on the probability distribution is available. For the rest of this paper, we assume that $l = 2$ in (5). This assumption is made for two reasons. First, any biochemical reaction, with more than two reactants , can be written as a series of mono- or bi-molecular reactions that result in propensity functions of order at most two [18]. Second, our results can be easily extended to the case $l > 3$ as remarked later. Therefore, without loss of generality, we assume $l = 2$ and obtain

$$\frac{d}{dt}E_n(t) = A(t) E_n + b(t) \mu_{n+1} + r(t); \text{ with } E_n(0) \text{ given} \tag{6}$$

where $E_n := \left[\mu_1^T, \mu_2^T, ..., \mu_n^T\right]^T$ is the aggregation of all moments up to order n, and $A(t)$ and $b(t)$ are matrices with appropriate dimension [6]. Define matrices H_n and V_n such that

$$\mu_{n+1}(t) = H_n P(t), \tag{7}$$

$$E_n(t) = V_n P(t), \tag{8}$$

where $P \in \mathbb{R}^p_{\geq 0}$ is the vector composed of entries $\Pr(S = k)$ with $k \in \mathbb{Z}^q_{\geq 0}$ and $k \leq U$. For example, for a one dimensional random variable S,

$$H_n = \begin{bmatrix} 0 & 1^{n+1} & 2^{n+1} & \cdots & U^{n+1} \end{bmatrix},$$

$$V_n = \begin{bmatrix} 0 & 1 & 2 & \cdots & U \\ 0 & 1^2 & 2^2 & \cdots & U^2 \\ \vdots & \vdots & & & \\ 0 & 1^n & 2^n & \cdots & U^n \end{bmatrix}.$$

Remark 5: Notice that if $l > 2$ in (5) the moments equation takes the form

$$\frac{d}{dt}E_n(t) = A(t) E_n + b(t) \left[\mu_{n+1}^T, ..., \mu_{n+l-1}^T\right]^T + r(t).$$

In this case, we modify the definition of H_n given in (7). We define H_n^l as matrix such that

$$\left[\mu_{n+1}^T, \mu_{n+2}^T, ..., \mu_{n+l-1}^T\right]^T = H_n^l P.$$

Then, the results of this paper hold valid with H_n replace by H_n^l. Hence, without loss of generality we assume $l = 2$.

Suppose that we are interested in closing the system of moments for the first n moments. To this end, we approximate μ_{n+1} by $\phi(E_n)$, where $\phi(.)$ is some (possibly nonlinear) function of the first n moments. In this case, the closed system of moments is given by

$$\frac{d}{dt}\nu = A\nu + b\phi(v) + r; \text{ with } \nu(0) = E_n(0), \tag{9}$$

which is analogous to (6) with μ_{n+1} replaced by $\phi(E_n)$. The function ϕ is the MCF and should be chosen such that the error between μ_{n+1} and $\phi(E_n)$ is minimized. This error is clearly a function of the probability vector and ϕ. More precisely, define

$$\rho_{NL}(P,\phi) = \left\|\mu_{n+1} - \phi(E_n)\right\|,$$

where the norm $\|.\|$ is taken to be the l_∞ norm. Above, we have made the dependency of ρ_{NL} on $\phi(.)$ and the probability vector P explicit; and the subscript NL in ρ_{NL} refers to the fact that ϕ can be a nonlinear function, in general. Further, since the probability vector is not known, in the Robust Moment Closure (RMC) technique, $\phi(.)$ is chosen such that the worst-case error is minimized. This amounts to the following min-max problem:

$$\begin{aligned}\rho^o_{NL} &= \inf_{\phi}\sup_{P\in\mathbb{P}}\rho(P,\phi) &(10)\\ &= \inf_{\phi}\sup_{P\in\mathbb{P}}\left\|H_nP - \phi\circ V_nP\right\|, &(11)\end{aligned}$$

where P is restricted to the set of probability vectors, $P\geq 0$ and $\|P\|_{l_1}=1$. To solve this optimization problem, notice that any $P\in\mathbb{P}$ can be uniquely written as

$$P = N(V_n)\,x + N^{\perp}(V_n)\,y, \tag{12}$$

for some $x\in\boldsymbol{R}^{p-r}$ and $y\in\boldsymbol{R}^r$, where r is the rank of V_n. Define, $\mathcal{D}$ to be the set of y's such that (12) holds for some $x\in\boldsymbol{R}^{p-r}$. Also, given $y\in\mathcal{D}$, let Ω_y be the set of x's such that (12) holds. Those are,

$$\begin{aligned}\mathcal{D} &: = \left\{y\in\boldsymbol{R}^r | \exists x\in\boldsymbol{R}^{p-r} : N(V_n)\,x + N^{\perp}(V_n)\,y\in\mathbb{P}\right\}, &(13)\\ \Omega_y &: = \left\{x\in\boldsymbol{R}^{p-r} | N(V_n)\,x + N^{\perp}(V_n)\,y\in\mathbb{P}\right\}. &(14)\end{aligned}$$

We solve (10) in the next theorem.

Theorem 6: The optimal value in (10) is given by

$$\rho^o_{NL} = \frac{1}{2}\max_i\max_{y\in\mathcal{D}}\left[\max_{x\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x - \min_{x\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x\right], \tag{15}$$

where $\bar{H} = H_nN(V_n)$.

The above theorem characterizes the optimal error when the MCF is not restricted to any particular class. Furthermore, one can cast (15) as a linear program and hence compute it in a tractable way. In fact, the optimal cost in (15) can be rewritten as

$$\rho^o_{NL} = \frac{1}{2}\min\bar{\eta},$$

subject to

$$\begin{aligned}-\eta_i &\leq \bar{\eta},\\ -\mathcal{R}\left[\bar{H}\right]_i x_1 + \mathcal{R}\left[\bar{H}\right]_i x_2 &\leq \eta_i,\\ N(V_n)\,x_j + N^{\perp}(V_n)\,y &\geq 0,\\ \mathbf{1}^TN(V_n)\,x_j + \mathbf{1}^TN^{\perp}(V_n)\,y &= 1,\end{aligned}$$

for all $i=1,2,...,m$ and $j=1,2$, where m is the number of rows of $\bar{H}$. However, computing the MCF itself is a harder problem. In fact, the optimal moment closure function is parametrized by $y\in\mathcal{D}$ and is constructed in the proof of the above theorem. It is given by (33) and (35) , and can be computed via LP for a given value of $y\in\mathcal{D}$. However, as the LPs do not have a closed form, $\phi^{optimal}(.)$ does not have a closed form either and this makes the use of this MCF challenging from the computational point of view. Therefore, we focus on the affine moment closure functions next and show that designing the optimal affine MCF is in fact a LP and hence tractable. Moreover, we compare the performance of the affine MCF (defined in (16)) with (15) and show that nonlinear MCF cannot outperform affine ones.

A. Affine versus Nonlinear Moment Closure Functions

In this section, we consider affine moment closure functions of the form

$$\phi_{Affine}(V_nP) = KV_nP + K_0, \tag{16}$$

where K and K_0 are matrices with appropriate dimensions. For compactness, we adopt the following notation:

$$\begin{aligned}\rho_{affine}(K,K_0) &= \sup_{P\in\mathbb{P}}\left\|H_nP - (KV_nP+K_0)\right\|,\\ \rho^o_{affine} &= \inf_{K,K_0}\sup_{P\in\mathbb{P}}\left\|H_nP - (KV_nP+K_0)\right\|.\end{aligned}$$

We note that given K and K_0, one can use Lemma 1 to compute $\rho_{affine}(K,K_0)$ as

$$\rho_{affine}(K,K_0) = \left\|H_n - \left(KV_n + K_0\mathbf{1}^T\right)\right\|_{l_1-l_\infty},$$

where we used $K_0 = K_0\mathbf{1}^TP$ for $P\in\mathbb{P}$.

The next theorem provides a LP for computing the optimal affine moment closure function.

Theorem 7: The optimal affine moment closure function, in the form (16), can be found from the following LP:

$$\rho^o_{affine} = \inf_{\phi\text{ affine}}\sup_{P\in\mathbb{P}}\rho(P,\phi) = \min_{K,K_0}\gamma$$

subject to

$$-\gamma\mathbf{1}^T \leq \mathcal{R}\left[H_n - \left(KV_n + K_0\mathbf{1}^T\right)\right]_i \leq \gamma\mathbf{1}^T,$$

for $i=1,2,...,m$, where m is the number of rows in H_n. Furthermore, the optimal cost is given by

$$\rho^o_{affine} = \inf_{\phi\text{ affine}}\sup_{P\in\mathbb{P}}\rho(P,\phi) = \max_{i\in\{1,2,...,m\}}\max_f\left[\mathcal{R}\left[H\right]_i f\right], \tag{17}$$

subject to

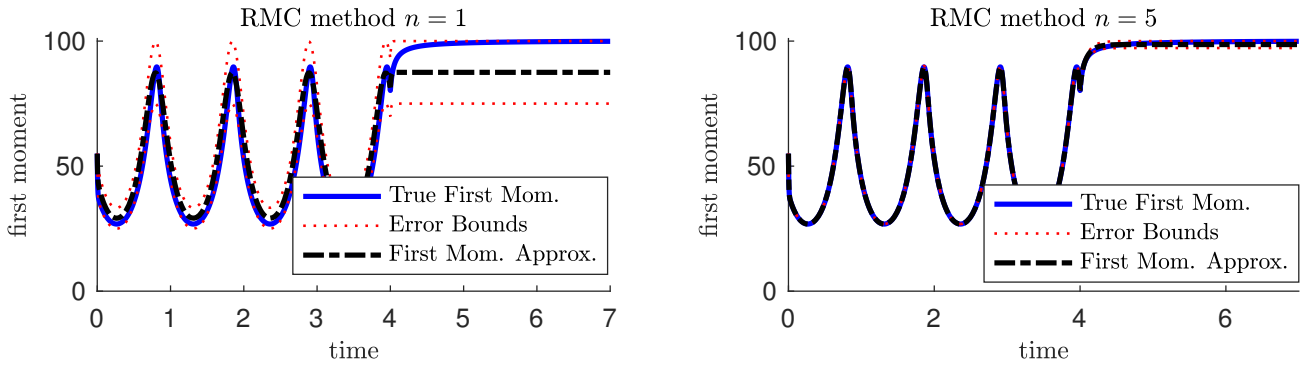
$$\begin{aligned}\begin{bmatrix}V_n\\ \mathbf{1}^T\end{bmatrix} f &= 0, &(18)\\ \|f\|_{l_1} &\leq 1. &(19)\end{aligned}$$

The above theorem provides the machinery to find the optimal affine moment closure function. In general, one expects that $\rho^o_{affine}\geq\rho^o_{NL}$, as affine functions form a proper subset of all functions. However, in what follows, we will show that no moment closure function can outperform affine ones.

Theorem 8: The following equality holds

$$\rho^o_{NL} = \rho^o_{affine}. \tag{20}$$

In light of this theorem, we use the affine MCF in our RMC scheme. Next, we quantify the error between the true and the approximate system of moments.



(a) RMC method $n = 1$

(b) RMC method $n = 5$

Fig. 1: The robust moment closure method is applied on the Birth-Death process to close the system of moments for $n = 1$ and $n = 5$.

B. Error Quantification

We derive the error bounds between the closed system of moments (9) and the true system (5) by studying the error dynamics. Let the error be given by e, where $e = E_n - \nu_n$. Then, the error dynamics is given by

$$\begin{aligned} \dot{e} &= AE_n + b\mu_{n+1} - A\nu_n - bK\nu_n - bK_0 \\ &= (A + bK)\, e + b\left[\mu_{n+1} - (KE_n + K_0)\right]; e(0) = 0. \end{aligned}$$

Furthermore, for $i = 1, 2, ..., n$, e_i, which is the error in the i^{th} moment between the true and the closed system, can be written as

$$e_i(t) = \int_0^t c_i \Phi(t,\tau)\left[\mu_{n+1}(\tau) - (KE_n(\tau) + K_0)\right] d\tau, \tag{21}$$

where c_i is defined in (2); $\Phi(t,\tau)$ is the state transition matrix associated with the pair $(A + bK, b)$ and

$$\frac{d}{dt}\Phi(t,t_0) = (A(t) + b(t)K)\,\Phi(t,t_0); \text{ with } \Phi(t_0,t_0) = I.$$

This error is quantified in the next theorem.

Theorem 9: Given K and K_0, the error in the i^{th} moment due to the RMC is given by

$$ess \sup_{t\in[0,T]} \|e_i(t)\| \leq \left[\int_0^T \|c_i \Phi(t,\tau)\, b\|\, dt\right] \times \rho_{affine}(K, K_0). \tag{22}$$

V. Illustrative Examples

In this section, we show the utility of our approach with two example. The first example is an academic birth-death process with nonlinear rates and, in the second example, we model an enzymatic reaction.

A. Birth-Death Process

In this example, we consider an academic Birth-Death process for species S with nonlinear propensity functions as given in the following. The birth process

$$\emptyset \xrightarrow{(1)} Y,$$

is characterized by the propensity function $a_1(S) = (N - S)^2$ and the stoichiometry coefficient $\gamma_1 = +1$. And, the death process

$$Y \xrightarrow{(2)} \emptyset,$$

is characterized by the propensity function $a_2(t, S) = u(t)\, S(t)$ and the stoichiometry coefficient $\gamma_2 = -1$, where $u(t)$ is the exogenous input to the system. In this example, we let $N = 100$ and

$$u(t) = \begin{cases} N(1 + \sin(6t)), & \text{if } 0 \leq t \leq 4 \\ 0, & \text{otherwise} \end{cases}.$$

The chemical master equation for this system is given by

$$\begin{aligned} &\frac{d}{dt}\Pr(S(t) = i) = (N - i + 1)^2 \Pr(S(t) = i - 1) \\ &+ u(t)(i + 1)\Pr(S(t) = i + 1) \\ &- \left[u(t)\, i + (N - i)^2\right]\Pr(S(t) = i). \end{aligned}$$

It can be easily verified that the induced system of moments is not closed. In fact, the first moment can be written as

$$\dot{\mu}_1 = -(2N + u(t))\,\mu_1 + N^2 + \mu_2.$$

We apply the RMC method on this system. To this end, we approximate μ_2 by an affine function of μ_1. More precisely, we solve the optimization problem

$$\rho := \inf_{K, K_0} \sup_{P\in\mathbb{P}} \|\mu_2 - K\mu_1 - K_0\|,$$

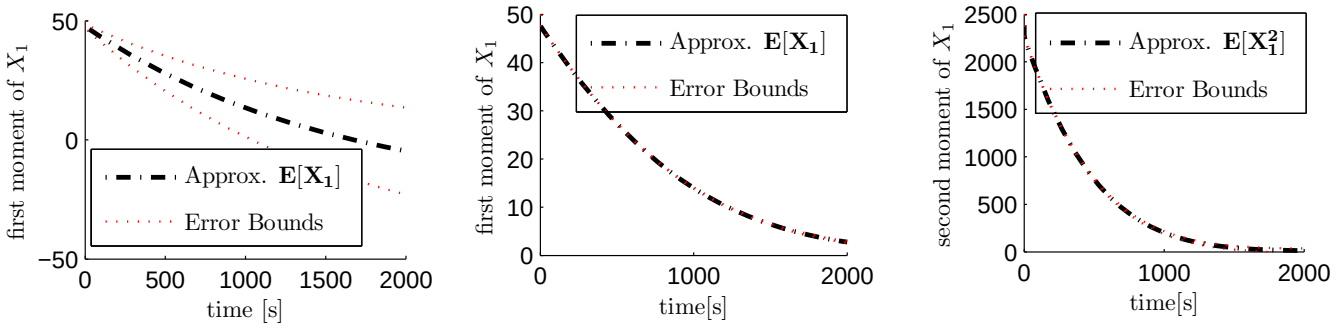
to find K and K_0 by converting it to a LP as given in Theorem 7. For this example, we converted the LP into its canonical form [12] and used the MATLAB linear programming toolbox to find the optimal solution to be

$$K = 100, K_0 = -1250, \rho = 1250.$$

Hence, the approximate and the error systems are respectively given by

$$\dot{\nu} = -(2N + u(t) + 100)\,\mu_1 + \left(N^2 - 1250\right), \tag{23}$$

$$\dot{e} = -(2N + u(t) + 100)\, e + r, \tag{24}$$



(a) RMC method $n = 1$ (b) RMC method $n = 3$ (c) RMC method $n = 3$

Fig. 2: Low-order moments of X_1 approximated via applying the RMC method for $n = 1$ and $n = 3$. For this simulation $G = 0.02[s^{-1}]$, $k_{off} = 1[s^{-1}]$, $\frac{k_{on}}{\Omega} = 0.02[(molecules)^{-2}s^{-1}]$, $\Omega = 0.5[\mu m^3]$, $S_T = 50[molecus]$, $E_T = 5[molecules]$.

with $\|r\| \leq \rho = 1250$. Figure 1a depicts the trajectories of the approximate system (23), the error bounds induced by the error dynamics (24), and the true first moment from simulating the CME for $n = 1$. The simulations conform with our theoretical results in that the true first moment remains between the error bounds that we derived. Furthermore, while theoretically we do not have the proof for it, the simulation results, Figure 1b, support the fact that as n is increased, the error in RMC method is decreased. This is a consistent behavior with the derivative matching [11] and cumulant truncation [10] moment closure methods.

B. Application Example: Enzymatic Reaction

In this example, we consider the standard form of an enzymatic reaction [4]:

$$S + E \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} C, C \xrightarrow{G} E + S^*. \tag{25}$$

Above, S, E, C, and S^* are respectively the substrate, the enzyme, the binding complex, and the reaction product S^*. We assume that the total number of substrate and enzyme molecules is conserved. That is, for some positive integers S_T and E_T, $S+C+S^* = S_T$ and $E+C = E_T$. We apply the RMC method for the first moments. To this end let X_1 and X_2 be the amount of free substrate and enzyme, respectively; i.e., $X_1 = S$ and $X_2 = E$. Then, for the reactions in (25), we have the following propensity functions and stoichiometry coefficients (see standard ref e.g. [4]):

$$\begin{aligned} a_1 &= \frac{k_{on}}{\Omega} X_1 X_2, \gamma_1 = [-1, -1]^T, \\ a_2 &= k_{off}\left(E_T - X_2\right), \gamma_2 = [+1, +1]^T, \\ a_3 &= G\left(E_T - X_2\right), \gamma_3 = [0, +1]^T, \end{aligned}$$

where Ω is the volume. Note that the first moments of X_1 and X_2 are given by

$$\frac{d}{dt}\begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{bmatrix} = \begin{bmatrix} 0 & -k_{off} \\ 0 & -k_{off} - G \end{bmatrix}\begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{bmatrix} + \begin{bmatrix} -\frac{k_{on}}{\Omega} \\ -\frac{k_{on}}{\Omega} \end{bmatrix}\mathbb{E}[X_1 X_2] + \begin{bmatrix} k_{off}E_T \\ k_{off}E_T + GE_T \end{bmatrix}.$$

Figure 2a shows the approximated first moment of X_1 together with its error bounds. As it can be seen, the error in this approximation is substantial. However, if one applies the RMC technique for larger n, e.g. $n = 3$, one can possibly find a better approximate system. This might be due to the fact that as we increase n, the dimension of the approximate system increases as well. It is reasonable to expect that the increase in the dimension of the approximate system yields a more accurate estimation. This is supported by our simulations as shown in Figure 2. Furthermore, having approximated the low-order moments, we can approximate various dispersion indices. A dispersion index measures the extent to which the probability distribution is dispersed. In this example, we compute the Variance-to-Mean Ratio (VMR), which is defined by $VMR := \frac{\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2}{\mathbb{E}[X_1]}$ [13]. We use the approximated first and second moments to compute the VMR as it is shown in Figure 3. The graph starts from zero since we have a deterministic initial condition. As time increases, the approximate VMR may not be accurate since its denominator ($\mathbb{E}[X_1]$) approaches to zero and small error in it yields a large error in VMR. This might be the reason that the graph rises to the right.

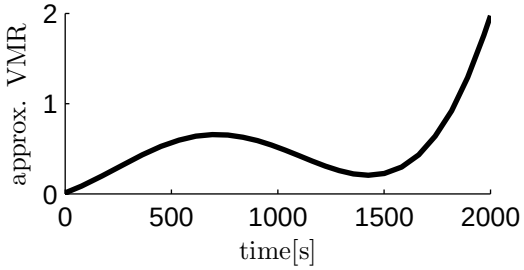


Fig. 3: approximated VMR for X_1.

VI. Conclusion

In this paper, we studied the moment closure problem for the CME. We developed the Robust Moment Closure technique in which we used the affine moment closure functions to approximate the higher order moments in terms of the

lower order ones. We showed that, in the absence of a priori information on the probability distribution, the affine MCFs are optimal and, furthermore, they can be found via LP. Consequently, utilizing the affine moment closure functions, we derived a system of finite dimension that approximates the low-order moments. Moreover, we quantified the error in this approximation in terms of the l_∞ induced norm of some linear system. Our results allow for the explicit simulation and analytical computation of approximate moments, which can be effectively used to characterize the noise properties of the biochemical network under study as illustrated by two examples.

APPENDIX

Proof on Theorem 6

First, notice that, the definition of l_∞norm, (10) can be rewritten as

$$\rho_{NL}^{o}=\inf_{\phi_1,\phi_2,...,\phi_m}\max_{i\in\{1,2,...,m\}}\sup_{P\in\mathbb{P}^p}\left|\mathcal{R}\left[H_n\right]_i P-\phi_i\left(V_nP\right)\right|, \tag{26}$$

where $\phi_i(.)$ is the i^{th} entry of vector $\phi(.)$ and $|.|$ is the absolute value function. Hence, $\rho_{NL}^{o}=\max\{\bar{\eta}_1,\bar{\eta}_2,...,\bar{\eta}_m\}$, where

$$\bar{\eta}_i=\inf_{\phi_i}\sup_{P\in\mathbb{P}^p}\left|\mathcal{R}\left[H_n\right]_i P-\phi_i\left(V_nP\right)\right|,\ \text{for } i=1,2,...,m. \tag{27}$$

Given $i\in\{1,2,...,m\}$, (20) is equivalent to

$$\bar{\eta}_i=\inf_{\phi_i}\max_{y\in\mathcal{D}}\max_{x\in\Omega_y} \tag{28}$$

$$\left|\mathcal{R}\left[H_n\right]_i\left(N\left(V_n\right)x+N^{\perp}\left(V_n\right)y\right)-\phi_i\left(V_nN^{\perp}\left(V_n\right)y\right)\right|, \tag{29}$$

where we used (12), and $\mathcal{D}$ and Ω_y are defined in (13)-(14). Since, $\phi_i(.)$ is not restricted to any class of functions, one can define

$$f_i(y)=\phi_i\left(V_nN^{\perp}\left(V_n\right)y\right)-\mathcal{R}\left[H_n\right]_i N^{\perp}\left(V_n\right)y, \tag{30}$$

and perform the optimization over $f_i(y)$. Therefore, (28) reduces to

$$\bar{\eta}_i=\inf_{f_i}\max_{y\in\mathcal{D}}\max_{x\in\Omega_y}\left|\mathcal{R}\left[\bar{H}\right]_i x-f_i(y)\right|, \tag{31}$$

where $\bar{H}=H_nN\left(V_n\right)$. From (31), we have

$$\bar{\eta}_i\geq\max_{y\in\mathcal{D}}\min_{f_i(y)}\max_{x\in\Omega_y}\left|\mathcal{R}\left[\bar{H}\right]_i x-f_i(y)\right|. \tag{32}$$

Now, given $y\in\mathcal{D}$, the optimal value of $f_i(y)$ to minimize $\max_{x\in\Omega_y}\left|\mathcal{R}\left[\bar{H}\right]_i x-f_i(y)\right|$ is the algebraic average between the largest and smallest values of $\mathcal{R}\left[\bar{H}\right]_i x$ where $x\in\Omega_y$. That is,

$$f_i^{optimal}(y)=\frac{1}{2}\left[\max_{x_1\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x_1+\min_{x_2\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x_2\right]. \tag{33}$$

In this case,

$$\begin{aligned}&\min_{f_i(y)}\max_{x\in\Omega_y}\left|\mathcal{R}\left[\bar{H}\right]_i x-f_i(y)\right|\\ =\ &\frac{1}{2}\left[\max_{x_1\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x_1-\min_{x_2\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x_2\right]\\ =\ &-\frac{1}{2}\left[\min_{x_1\in\Omega_y}-\mathcal{R}\left[\bar{H}\right]_i x_1+\min_{x_2\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x_2\right]\end{aligned}$$

Therefore, the lower bound in (32) is given by

$$\bar{\eta}_i\geq\frac{1}{2}\max_{y\in\mathcal{D}}\left[\max_{x\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x-\min_{x\in\Omega_y}\mathcal{R}\left[\bar{H}\right]_i x\right]. \tag{34}$$

Furthermore, the lower bound, in (32), is achievable. In fact, the lower bound is attainable for any $f_i(.)$, in (31), with the property that it coincides with $f_i^{optimal}(.)$ on the set $\mathcal{D}$. From (30), an optimal $\phi_i^{optimal}(.)$ is the one whose value at $V_nN^{\perp}\left(V_n\right)y$, for $y\in\mathcal{D}$, is given by

$$\phi_i^{optimal}\left(V_nN^{\perp}\left(V_n\right)y\right)=f_i^{optimal}(y)+\mathcal{R}\left[H_n\right]_i N^{\perp}\left(V_n\right)y, \tag{35}$$

and arbitrary otherwise. That is (34) is a tight inequality and hence taking the $\max$ over i completes the proof.

Proof of Theorem 7

Notice that for affine moment closure functions we have

$$\begin{aligned}\rho_{affine}^{o}&=\min_{K,K_0}\max_{P\in\mathbb{P}}\left\|H_nP-\left(KV_nP+K_0\right)\right\|\\ &=\min_{K,K_0}\max_{P\in\mathbb{P}}\left\|\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]P\right\|.\end{aligned} \tag{36}$$

By Lemma 1, one obtains

$$\rho_{affine}^{o}=\min_{K,K_0}\max_{i,j}\left|\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]_{ij}\right|, \tag{37}$$

where $\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]_{ij}$ denotes the entry on row i and column j of the matrix $H_n-\left(KV_n+K_0\mathbf{1}^T\right)\in\mathbb{R}^{m\times p}$. Then,

$$\max_{i,j}\left\|\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]_{ij}\right\|=\min\gamma,$$

subject to

$$-\gamma\mathbf{1}^T\leq\mathcal{R}\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]_i\leq\gamma\mathbf{1}^T,$$

for $i=1,2,...,m$. This completes the proof of the first part. For the second part, note that (37) can be written as

$$\begin{aligned}\rho_{affine}^{o}=\min_{\substack{\gamma\geq0\\K,K_0}}\max_{i}\max_{\zeta_i\in\mathbb{R}^p_{\geq0},\xi_i\in\mathbb{R}^p_{\geq0}}\ &\gamma+\left(\mathcal{R}\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]_i-\gamma\mathbf{1}^T\right)\zeta_i\\ &-\left(\mathcal{R}\left[H_n-\left(KV_n+K_0\mathbf{1}^T\right)\right]_i+\gamma\mathbf{1}^T\right)\xi_i,\end{aligned}$$

where ζ_i's and ξ_i's are the Lagrange multipliers. Due to the convexity of the objective function and constraints (zero duality gap), one can change the order of the $\min$ and $\max$. Hence,

$$\rho_{affine}^{o}=\max_{i}\max_{\zeta_i\in\mathbb{R}^p_{\geq0},\xi_i\in\mathbb{R}^p_{\geq0}}\mathcal{G}(\zeta,\xi), \tag{38}$$

where $\mathcal{G}(\zeta,\xi)$ is the so-called Lagrangian and is given by

$$\begin{aligned}\mathcal{G}_i(\zeta,\xi)=\min_{\substack{\gamma\geq0\\K,K_0}}&\left[1-\mathbf{1}^T\sum_{i=1}^{m}\left(\zeta_i+\xi_i\right)\right]\gamma\\ &+\left[\mathcal{R}\left[H_n\right]_i-\mathcal{R}\left[K\right]_i V_n-\mathcal{R}\left[K_0\right]_i\mathbf{1}^T\right]\left(\zeta_i-\xi_i\right).\end{aligned}$$

One can easily verify that

$$\mathcal{G}_i(\zeta,\xi)=\mathcal{R}\left[H_n\right]_i\left(\zeta_i-\xi_i\right), \tag{39}$$

if

$$\begin{aligned} V_n(\zeta_i - \xi_i) &= 0, \text{ for } i \in \{1,2,...,m\}, \quad (40)\\ \mathbf{1}^T(\zeta_i - \xi_i) &= 0, \text{ for } i \in \{1,2,...,m\}, \quad (41)\\ \mathbf{1}^T(\zeta_i + \xi_i) &\leq 1, \quad (42)\end{aligned}$$

and otherwise $\mathcal{G}_i(\zeta,\xi) = -\infty$. The proof is complete by defining the new variables $f_i = \zeta_i - \xi_i$. More precisely, for any set of ζ_i's and ξ_i's that satisfies (40)-(42), one can define $f_i = \zeta_i - \xi_i$ satisfying (18)-(19). And, conversely, for any set of f_i's that satisfies (18) and (19), define $\zeta_i = f_i^+$ and $\xi_i = f_i^-$ where $f_i = f_i^+ - f_i^-$ is the positive decomposition of f_i.

Proof of Theorem 8

We will show that $\rho^o_{NL} \geq \rho^o_{affine}$, where ρ^o_{NL} and ρ^o_{affine} are given in (15) and (17), respectively. To this end, let i^* and f^* be the maximizers of (17). That is,

$$V_n f^* = 0, \mathbf{1}^T f^* = 0, \|f^*\| = 1, \quad (43)$$

and $\rho^o_{affine} = \mathcal{R}[H]_{i^*} f^*$. Now, let $f^* = (f^*)^+ - (f^*)^-$ be the positive decomposition of f^*. The nonnegative vectors $(f^*)^+$ and $(f^*)^-$ can be written as unique summations of elements from $\mathcal{N}(V_n)$ and $\mathcal{N}^\perp(V_n)$. More precisely, there exist $\alpha_1, \alpha_2 \in \mathbb{R}^{q-r}$ and $\beta_1, \beta_2 \in \mathbb{R}^r$ such that

$$\begin{aligned} (f^*)^+ &= N(V_n)\alpha_1 + N^\perp(V_n)\beta_1, \quad (44)\\ (f^*)^- &= N(V_n)\alpha_2 + N^\perp(V_n)\beta_2. \quad (45)\end{aligned}$$

Furthermore, from (43) we have

$$\begin{aligned} &V_n(f^*)^+ - V_n(f^*)^- = V_n N^\perp(V_n)(\beta_1 - \beta_2) = 0,\\ &\mathbf{1}^T N(V_n)(\alpha_1 - \alpha_2) + \mathbf{1}^T N^\perp(V_n)(\beta_1 - \beta_2) = 0,\\ &\mathbf{1}^T N(V_n)(\alpha_1 + \alpha_2) + \mathbf{1}^T N^\perp(V_n)(\beta_1 + \beta_2) = 1.\end{aligned}$$

From above expressions, since $V_n N^\perp(V_n)$ is full column rank, $\beta_1 = \beta_2 = \beta$, $\mathbf{1}^T N(V_n)\alpha_1 = \mathbf{1}^T N(V_n)\alpha_2$, and

$$\begin{aligned} 2\mathbf{1}^T N(V_n)\alpha_1 + 2\mathbf{1}^T N^\perp(V_n)(\beta) &= 1,\\ 2\mathbf{1}^T N(V_n)\alpha_2 + 2\mathbf{1}^T N^\perp(V_n)(\beta) &= 1.\end{aligned}$$

Let $x_1 = 2\alpha_1$, $x_2 = 2\alpha_2$, and $y = 2\beta$. Then, it is easy to verify that $y \in \mathcal{D}$ and $x_1, x_2 \in \Omega_y$, where $\mathcal{D}$ and Ω_y are defined in (13)-(14). Also, from (15), we have

$$\begin{aligned} \rho^o_{NL} &= \frac{1}{2}\max_i \max_{y\in\mathcal{D}} \left[\max_{x\in\Omega_y} \mathcal{R}\left[\bar{H}\right]_i x + \max_{x\in\Omega_y} -\mathcal{R}\left[\bar{H}\right]_i x\right]\\ &\geq \frac{1}{2}\left[\mathcal{R}\left[\bar{H}\right]_{i^*} x_1 - \mathcal{R}\left[\bar{H}\right]_{i^*} x_2\right]\\ &= \left[\mathcal{R}[H]_{i^*} N(V_n)\alpha_1 - \mathcal{R}[H]_{i^*} N(V_n)\alpha_2\right]\\ &= \mathcal{R}[H]_{i^*}\left[(f^*)^+ - (f^*)^-\right] = \mathcal{R}[H]_{i^*} f^* = \rho^o_{affine}.\end{aligned}$$

The proof is complete.

Proof of Theorem 9

The proof is based on the direct calculation; from (21), one obtains

$$\begin{aligned} &ess\sup_{t\in[0,T]} \|e_i(t)\| = \sup_{t\in[0,T]}\\ &\left\|\int_0^t c_i \Phi(t,\tau)\, b\left[\mu_{n+1}(\tau) - (KE_n(\tau) + K_0)\right] d\tau\right\|\\ \leq &\sup_{t\in[0,T]} \int_0^t \|c_i\Phi(t,\tau)\, b\|\, \bar{\rho}_L(K,K_0)\, d\tau,\end{aligned}$$

which is the same is (22).

REFERENCES

[1] I. Oppenheim, K. Shuler, and G. Weiss, "Stochastic and deterministic formulation of chemical rate equations," *The Journal of Chemical Physics*, vol. 50, no. 1, pp. 460–466, 1969.

[2] T. G. Kurtz, "The relationship between stochastic and deterministic models for chemical reactions," *The Journal of Chemical Physics*, vol. 57, no. 7, pp. 2976–2978, 1972.

[3] C. V. Rao, D. M. Wolf, and A. P. Arkin, "Control, exploitation and tolerance of intracellular noise," *Nature*, vol. 420, no. 6912, pp. 231–237, 2002.

[4] D. Del Vecchio and R. M. Murray, *Biomolecular feedback systems*. Princeton University Press, 2015.

[5] I. Matheson, D. Walls, and C. Gardiner, "Stochastic models of firstorder nonequilibrium phase transitions in chemical reactions," *Journal of Statistical Physics*, vol. 12, no. 1, pp. 21–34, 1975.

[6] V. Sotiropoulos and Y. N. Kaznessis, "Analytical derivation of moment equations in stochastic chemical kinetics," *Chemical engineering science*, vol. 66, no. 3, pp. 268–277, 2011.

[7] P. Whittle, "On the use of the normal approximation in the treatment of stochastic processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 268–281, 1957.

[8] A. Singh and J. P. Hespanha, "Lognormal moment closures for biochemical reactions," in *Decision and Control, 2006 45th IEEE Conference on*. IEEE, 2006, pp. 2063–2068.

[9] I. Krishnarajah, A. Cook, G. Marion, and G. Gibson, "Novel moment closure approximations in stochastic epidemics," *Bulletin of mathematical biology*, vol. 67, no. 4, pp. 855–873, 2005.

[10] J. H. Matis and T. R. Kiffe, "On interacting bee/mite populations: a stochastic model with analysis using cumulant truncation," *Environmental and Ecological Statistics*, vol. 9, no. 3, pp. 237–258, 2002.

[11] A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 414–418, 2011.

[12] D. G. Luenberger, *Introduction to linear and nonlinear programming*. Addison-Wesley Reading, MA, 1973, vol. 28.

[13] R. Peck, C. Olsen, and J. L. Devore, *Introduction to statistics and data analysis*. Cengage Learning, 2015.

[14] N. G. Van Kampen and W. P. Reinhardt, "Stochastic processes in physics and chemistry," 1983.

[15] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A: Statistical Mechanics and its Applications*, vol. 188, no. 1-3, pp. 404–425, 1992.

[16] ——, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of computational physics*, vol. 22, no. 4, pp. 403–434, 1976.

[17] A. Gupta, C. Briat, and M. Khammash, "A scalable computational framework for establishing long-term behavior of stochastic reaction networks," *PLoS Comput Biol*, vol. 10, no. 6, p. e1003669, 2014.

[18] D. T. Gillespie, L. R. Petzold, and E. Seitaridou, "Validity conditions for stochastic chemical kinetics in diffusion-limited systems," *The Journal of chemical physics*, vol. 140, no. 5, p. 02B604_1, 2014.