Robust moment closure method for the chemical master equation

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Naghnaeian, Mohammad, and Domitilla Del Vecchio. “Robust Moment Closure Method for the Chemical Master Equation.” 2017 IEEE Conference on Control Technology and Applications (CCTA) (August 2017), Mauna Lani, HI, USA, Institute of Electrical and Electronics Engineers (IEEE), 2017.</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/CCTA.2017.8062585">http://dx.doi.org/10.1109/CCTA.2017.8062585</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers (IEEE)</td>
</tr>
<tr>
<td>Version</td>
<td>Original manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Mar 11 20:35:22 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/119165">http://hdl.handle.net/1721.1/119165</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
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 \( l_\infty \) 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
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 $l_\infty$ 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. 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 \(\mathbb{Z}_{\geq 0}\) (\(\mathbb{R}_{\geq 0}\)). 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 \(x_{i_1}x_{i_2}...x_{i_k}\), 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

\[\Psi_1(X) = [x_1, x_2]^T, \quad \Psi_2(X) = [x_1^2, x_1x_2, x_2^2]^T, \quad \Psi_3(X) = [x_1^3, x_1^2x_2, x_1x_2^2, x_2^3]^T.\]

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

\[\bar{\Psi}_n(X) = [\Psi_1^T(X), \Psi_2^T(X), ..., \Psi_n^T(X)]^T,\]

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.

\[\Psi_i(X) = c_i [\Psi_1^T(X), \Psi_2^T(X), ..., \Psi_n^T(X)]^T.\]

The \(l_\infty\) 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 \(P \in \mathbb{R}_{\geq 0}^n\) is called a probability vector if \(\|P\|_{l_1} = 1\). The set of all probability vectors with dimension \(p\) is denoted by \(\mathcal{P}\). 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 \(\mathcal{R}[M]\) and \(\mathcal{C}[M]\), we mean the \(i^{th}\) row and \(j^{th}\) column of \(M\), respectively. That is,

\[\mathcal{R}[M]_i = [m_{i1}, m_{i2}, ..., m_{im}], \quad \mathcal{C}[M]_j = [m_{1j}, m_{2j}, ..., m_{mj}]^T,\]

for \(i = 1, 2, ..., m\) and \(j = 1, 2, ..., n\). The \(l_\infty\), \(l_1\), and \(l_1\) to \(l_\infty\) induced norms of \(M\) are defined as \(\|M\|_{l_1-l_\infty} = \max_j \sum_{i=1}^m |m_{ij}|\), \(\|M\|_{l_\infty-l_1} = \max_i \sum_{j=1}^n |m_{ij}|\), 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

\[\mathcal{N}(M) = \{x: Mx = 0\}, \quad \mathcal{N}^\perp(M) = \{y: y^Tx = 0, \forall x \in \mathcal{N}(M)\}.\]

Also, define \(\mathcal{N}(M)\) and \(\mathcal{N}^\perp(M)\) 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

\[\sup_{P \in \mathcal{P}} \|MP\| = \|M\|_{l_1-l_\infty}.\]

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 \(\gamma_j\) defined as

\[Pr(S(t) + dt) = Pr(S(t) + \gamma_j)Pr(S(t)) = a_j(t, S(t)) dt + O(dt^2),\]

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

\[\frac{d}{dt} Pr(S(t) = k) = \sum_{j=1}^J \left\{ -a_j(t, k) Pr(S(t) = k) + a_j(t, k - \gamma_j) Pr(S(t) = k - \gamma_j) \right\}.\]

This equation is referred to as the Chemical Master Equation [13][14]. 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 \(S\) [15][14]. That is, for \(j = 1, 2, ..., J\),

\[a_j(t, S) = \sum_{l=0}^I \theta^l_j(t) \Psi_l(S),\]

for some \(l \in \mathbb{Z}_{\geq 0}\), where \(\theta^l_j(t)\)’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 [16] 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 [15], [14], and [17] 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
\[
p := \prod_{i=1}^{q} (1 + U_i).
\]

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 \( \mu_i \). Recall that \( \mu_i := E[\Psi_i(S)] = \sum_{S \in \mathbb{Z}_0^d} \Psi_i(k) P_r(S = k) \), where \( \Psi_i(S) \) is a vector composed of the entries of the form \( s_1^{k_1} s_2^{k_2} \ldots s_q^{k_q} \) with \( k_1 + k_2 + \ldots + k_q = i \). 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 \( i = 1, 2, \ldots \),
\[
\frac{d}{dt}\mu_i (t) = \beta_{i,0}(t) + \sum_{n=1}^{i+1-1} \beta_{i,n} \mu_n(t),
\]
with initial condition \( \mu_i (0) = \Psi_i (S(0)) \), for some properly defined matrices \( \beta_{i,n} (\cdot) \) 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 [2] 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, is a high dimensional system 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, \( \mu_{n+1}, \mu_{n+2}, \ldots, \mu_{p+1} \), by a possibly nonlinear function of the lower-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 (3). 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 [17]. 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
\[
\frac{d}{dt} E_n(t) = A(t) E_n(t) + b(t) \mu_{n+1} + r(t); \quad \text{with } E_n(0) \text{ given}
\]
where \( E_n := [\mu_1^T, \mu_2^T, \ldots, \mu_n^T]^T \) 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
\[
\mu_{n+1}(t) = H_n P(t),
\]
\[
E_n(t) = V_n P(t),
\]
where \( P \in \mathbb{R}_{\geq 0}^n \) is the vector composed of entries \( P_r(S = k) \) with \( k \in \mathbb{Z}_0^d \) and \( k \leq U \). For example, for a one dimensional random variable \( S \),
\[
H_n = \begin{bmatrix}
0 & 1 & 2 & \ldots & U
0 & 1^2 & 2^2 & \ldots & U^2
\vdots & \vdots & \vdots & \ddots & \vdots
0 & 1^n & 2^n & \ldots & U^n
\end{bmatrix},
\]
\[
V_n = \begin{bmatrix}
0 & 1 & 2 & \ldots & U
0 & 1^2 & 2^2 & \ldots & U^2
\vdots & \vdots & \vdots & \ddots & \vdots
0 & 1^n & 2^n & \ldots & U^n
\end{bmatrix}.
\]

**Remark 5:** Notice that if \( l > 2 \) in (5), the moments equation takes the form
\[
\frac{d}{dt} E_n(t) = A(t) E_n(t) + b(t) [\mu_{n+1}^T, \ldots, \mu_{n+l-1}^T]^T + r(t).
\]
In this case, we modify the definition of \( H_n \) given in (7).

We define \( H_n^l \) as matrix such that
\[
[\mu_{n+1}^T, \mu_{n+2}^T, \ldots, \mu_{n+l-1}^T]^T = H_n^l P.
\]

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 \( \mu_{n+1} \) by \( \phi(E_n) \), where \( \phi(\cdot) \) is some (possibly nonlinear) function of the first \( n \) moments. In this case, the closed system of moments is given by
\[
\frac{d}{dt} \nu = A \nu + b \phi(\nu) + r; \quad \text{with } \nu(0) = E_n(0),
\]
which is analogous to (5) with \( \mu_{n+1} \) replaced by \( \phi(E_n) \). The function \( \phi \) is the MCF and should be chosen such that the error between \( \mu_{n+1} \) and \( \phi(E_n) \) is minimized. This error is clearly a function of the probability vector and \( \phi \). More precisely, define
\[
\rho_{NL}(P, \phi) = \| \mu_{n+1} - \phi(E_n) \|,
\]
where the norm \( \| \cdot \| \) is taken to be the \( l_\infty \) norm. Above, we have made the dependency of \( \rho_{NL} \) on \( \phi(\cdot) \) and the probability vector \( P \) explicit; and the subscript \( NL \) in \( \rho_{NL} \) refers to the fact that \( \phi \) can be a nonlinear function, in general. Further, since the probability vector is not known, in the Robust Moment Closure (RMC) technique, \( \phi(\cdot) \) is chosen such that the worst-case error is minimized. This amounts to the following min-max problem:

\[
\rho^0_{NL} = \inf_{P \in \mathbb{P}} \sup_{\phi \in \Phi} \rho(P, \phi)
\]

(10)

\[
= \inf_{P \in \mathbb{P}} \sup_{\phi \in \Phi} \| H_n P - \phi \circ V_n P \|,
\]

(11)

where \( P \) is restricted to the set of probability vectors, \( P \geq 0 \) and \( \| P \|_1 = 1 \). To solve this optimization problem, notice that any \( P \in \mathbb{P} \) can be uniquely written as

\[
P = N(V_n) x + N^\perp(V_n) y,
\]

(12)

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

\[
D : = \{ y \in \mathbb{R}^r | \exists x \in \mathbb{R}^{p-r} : N(V_n) x + N^\perp(V_n) y \in \mathbb{P} \},
\]

\[
\Omega_y : = \{ x \in \mathbb{R}^{p-r} | N(V_n) x + N^\perp(V_n) y \in \mathbb{P} \}.
\]

We solve (10) in the next theorem.

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

\[
\rho^0_{NL} = \frac{1}{2} \max_{i} \max_{y \in D} \max_{x \in \Omega_y} \mathcal{R}[H]_{i} x - \min_{x \in \Omega_y} \mathcal{R}[H]_{i} x,
\]

(15)

where \( \mathcal{H} = H_n N(V_n) \).

The above theorem characterizes the optimal error when the MFC 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

\[
\rho^0_{NL} = \frac{1}{2} \min_{\tilde{\eta}_i}
\]

subject to

\[
-\tilde{\eta}_i \leq \tilde{\eta}_i,
\]

\[
-\mathcal{R}[H]_{i} x_1 + \mathcal{R}[H]_{i} x_2 \leq \eta_i,
\]

\[
N(V_n) x_j + N^\perp(V_n) y \geq 0,
\]

\[
1^T N(V_n) x_j + 1^T N^\perp(V_n) y \geq 1,
\]

for all \( i = 1, 2, \ldots, m \) and \( j = 1, 2 \), where \( m \) is the number of rows of \( \mathcal{H} \). However, computing the MFC itself is a harder problem. In fact, the optimal moment closure function is parameterized by \( y \in \mathcal{D} \) and is constructed in the proof of the above theorem. It is given by (13) and (12), 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, \( \phi_{\text{optimal}}(\cdot) \) 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

\[
\phi_{\text{Affine}}(V_n P) = K V_n P + K_0,
\]

(16)

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

\[
\rho_{\text{affine}}(K, K_0) = \sup_{P \in \mathbb{P}} \| H_n P - (K V_n P + K_0) \|,
\]

\[
\rho^0_{\text{affine}} = \inf_{K, K_0} \sup_{P \in \mathbb{P}} \| H_n P - (K V_n P + K_0) \|.
\]

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

\[
\rho_{\text{affine}}(K, K_0) = \| H_n - (K V_n + K_0 1^T) \|_{1-l_\infty},
\]

where we used \( K_0 = K_0 1^T P \) 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:

\[
\rho^0_{\text{affine}} = \inf_{\phi_{\text{affine}} \in \mathbb{P}} \sup_{P \in \mathbb{P}} \rho(P, \phi) = \min_{K, K_0} \gamma
\]

subject to

\[
-\gamma 1^T \leq \mathcal{R}[H_n - (K V_n + K_0 1^T)]_{i} \leq \gamma 1^T,
\]

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

\[
\rho^0_{\text{affine}} = \inf_{\phi_{\text{affine}} \in \mathbb{P}} \sup_{P \in \mathbb{P}} \rho(P, \phi) = \max_{i \in \{1, 2, \ldots, m\}} \max_{f} \left[ \mathcal{R}[H_n]_{i} f \right],
\]

(17)

subject to

\[
\left[ \begin{array}{c}
V_n^T \\
1
\end{array} \right] f = 0,
\]

(18)

\[
\| f \|_{l_1} \leq 1.
\]

(19)

The above theorem provides the machinery to find the optimal affine moment closure function. In general, one expects that \( \rho^0_{\text{affine}} \geq \rho^0_{NL} \), 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

\[
\rho^0_{NL} = \rho^0_{\text{affine}}.
\]
B. Error Quantification

We derive the error bounds between the closed system of moments (6) 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

\[
\dot{e} = AE_n + b\mu_{n+1} - A\nu_n - bK\nu_n - bK_0
\]

\[
= (A + bK) e + b [\mu_{n+1} - (KE_n + K_0)]; \quad e(0) = 0.
\]

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

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

(21)

where \( c_i \) is defined in (23); \( \Phi(t, \tau) \) is the state transition matrix associated with the pair \((A + bK, b)\) and

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

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

\[
\text{ess sup}_{t \in [0,T]} \| e_i(t) \| \leq \left[ \int_0^T \| c_i \Phi(t, \tau) b \| d\tau \right] \times \rho_{\text{affine}}(K, K_0),
\]

(22)

V. CONCLUSION

In this paper, we studied the moment closure problem for the CME. We developed the Robust Moment Closure technique in which we use 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 \( l_\infty \) 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.

APPENDIX

Proof of Theorem 6

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

\[
\rho_{NL}^0 = \inf_{\phi_1, \phi_2, ..., \phi_m, \phi_{i+1} \in [1, 2, ..., m]} \max_{P \in \mathbb{P}} \sup_{t \in [0,T]} |\mathcal{R}[H_n]_t P - \phi_i (V_n P)|
\]

(23)

where \( \phi_i(\cdot) \) is the \( i^{th} \) entry of vector \( \phi(\cdot) \) and \(|\cdot|\) is the absolute value function. Hence, \( \rho_{NL}^0 = \max \{ \bar{\eta}_1, \bar{\eta}_2, ..., \bar{\eta}_m \} \),

where \( \bar{\eta}_i = \inf_{\phi_i, P \in \mathbb{P}} |\mathcal{R}[H_n]_t P - \phi_i (V_n P)| \), for \( i = 1, 2, ..., m \).

(24)

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

\[
\bar{\eta}_i = \inf_{\phi_i, y \in \mathcal{D}} \max_{x \in \Omega_y} |\mathcal{R}[H_n]_t x - \phi_i (V_n N^\perp (V_n) y)|
\]

where we used (12), and \( \mathcal{D} \) and \( \Omega_y \) are defined in (13), (14).

Since, \( \phi_i(\cdot) \) is not restricted to any class of functions, one can define

\[
f_i(y) = \phi_i (V_n N^\perp (V_n) y) - \mathcal{R}[H_n]_t N^\perp (V_n) y,
\]

(27)

and perform the optimization over \( f_i(y) \). Therefore, (25) reduces to

\[
\bar{\eta}_i = \inf_{\phi_i, y \in \mathcal{D}} \max_{x \in \Omega_y} |\mathcal{R}[H]_t x - f_i(y)|
\]

(28)

where \( H = H_n N^\perp (V_n) \). From (28), we have

\[
\bar{\eta}_i \geq \max_{y \in \mathcal{D}} \min_{x \in \Omega_y} |\mathcal{R}[H]_t x - f_i(y)|
\]

(29)

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

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

(30)

In this case,

\[
\min_{f_i(y) \in \mathcal{D}} \max_{x \in \Omega_y} |\mathcal{R}[H]_t x - f_i(y)|
\]

\[
= \frac{1}{2} \left[ \max_{x_1 \in \Omega_y} \mathcal{R}[H]_t x_1 + \min_{x_2 \in \Omega_y} \mathcal{R}[H]_t x_2 \right]
\]

\[
= \frac{1}{2} \left[ \min_{x_1 \in \Omega_y} -\mathcal{R}[H]_t x_1 + \max_{x_2 \in \Omega_y} \mathcal{R}[H]_t x_2 \right]
\]

Therefore, the lower bound in (29) is given by

\[
\bar{\eta}_i \geq \frac{1}{2} \max_{y \in \mathcal{D}} \min_{x \in \Omega_y} |\mathcal{R}[H]_t x - \mathcal{R}[H]_t x|.
\]

(31)

Furthermore, the lower bound, in (29), is achievable. In fact, the lower bound is attainable for any \( f_i(\cdot) \), in (28), with the property that it coincides with \( f_i^\text{optimal}(\cdot) \) on the set \( \mathcal{D} \). From (27), an optimal \( f_i^\text{optimal}(\cdot) \) is the one whose value at \( V_n N^\perp (V_n) y \), for \( y \in \mathcal{D} \), is given by

\[
\phi_i^\text{optimal}(V_n N^\perp (V_n) y) = f_i^\text{optimal}(y) + \mathcal{R}[H_n]_t N^\perp (V_n) y,
\]

(32)

and arbitrary otherwise. That is (31) is a tight inequality and hence taking the \( \max \) over \( i \) completes the proof.

Proof of Theorem 6

Notice that for affine moment closure functions we have

\[
\rho_{\text{affine}}^0 = \min_{K, K_0} \max_{P \in \mathbb{P}} \| H_n P - (KV_n P + K_0) \|
\]

\[
= \min_{K, K_0} \max_{P \in \mathbb{P}} \left\| H_n - (KV_n + K_0 T) \right\| P.
\]

(33)

By Lemma 1, one obtains

\[
\rho_{\text{affine}}^0 = \min_{K, K_0} \max_{i, j} \left\| H_n - (KV_n + K_0 T) \right\|_{ij},
\]

(34)
where $[H_n - (KV_n + K_0T)]_{ij}$ denotes the entry on row $i$ and column $j$ of the matrix $H_n - (KV_n + K_0T) \in \mathbb{R}^{m \times p}$. Then,

$$
\max_{i,j} \left\| [H_n - (KV_n + K_0T)]_{ij} \right\| = \min \gamma,
$$

subject to

$$-\gamma 1^T \leq R [H_n - (KV_n + K_0T)]_i \leq \gamma 1^T,
$$

for $i = 1, 2, \ldots, m$. This completes the proof of the first part. For the second part, note that (34) can be written as

$$
\rho_{affine}^\star = \min_i \max_{\gamma \geq 0} \max_{K,K_0} \gamma + (R [H_n - (KV_n + K_0T)]_i - \gamma 1^T \xi_i)
$$

where $\xi_i$’s and $\xi_i$’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,

$$
\rho_{affine}^\star = \max_i \min_{\gamma \geq 0} \max_{K,K_0} \gamma + (R [H_n - (KV_n + K_0T)]_i - \gamma 1^T \xi_i)
$$

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

$$
\mathcal{G}_i (\xi, \xi) = \min_{\gamma \geq 0} \max_{K,K_0} \left[ 1 - 1^T \sum_{i=1}^m (\xi_i + \xi_i) \right]
$$

$$+ [R [H_n]_i - R [K]_i \cdot V_n - R [K_0]_i 1^T] (\xi_i - \xi_i).
$$

One can easily verify that

$$
\mathcal{G}_i (\xi, \xi) = R [H_n]_i (\xi_i - \xi_i),
$$

if

$$
V_n (\xi_i - \xi_i) = 0, \text{ for } i = 1, 2, \ldots, m,
$$

$$
1^T (\xi_i - \xi_i) = 0, \text{ for } i = 1, 2, \ldots, m,
$$

and otherwise $\mathcal{G}_i (\xi, \xi) = -\infty$. The proof is complete by defining the new variables $f_i = \zeta_i - \xi_i$. More precisely, for any set of $\zeta_i$’s and $\xi_i$’s that satisfies (37)-(39), 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 $\zeta_i = f_i^+ + \xi_i$ and $\xi_i = f_i^- - f_i^+$ where $f_i = f_i^+ - f_i^-$ is the positive decomposition of $f_i$.

**Proof of Theorem 5**

We will show that $\rho_{NL}^\star \geq \rho_{affine}^\star$, where $\rho_{NL}^\star$ and $\rho_{affine}^\star$ are given in (13) and (14), respectively. To this end, let $i^*$ and $f^*$ be the maximizers of (17). That is, $V_n f^* = 0, 1^T f^* = 0, \|f^*\| = 1$.

Furthermore, from (40) we have

$$
V_n (f^*)^+ - V_n (f^*)^- = V_n N^\perp (V_n) (\beta_1 - \beta_2) = 0,
$$

$$
1^T N (V_n) (\alpha_1 - \alpha_2) + 1^T N^\perp (V_n) (\beta_1 - \beta_2) = 0,
$$

$$
1^T N (V_n) (\alpha_1 + \alpha_2) + 1^T N^\perp (V_n) (\beta_1 + \beta_2) = 1.
$$

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

$$
21^T N (V_n) (\alpha_1 + 21^T N^\perp (V_n) (\beta) = 1,
$$

$$
21^T N (V_n) (\alpha_2 + 21^T N^\perp (V_n) (\beta) = 1.
$$

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

$$
\rho_{NL}^\star = \frac{1}{2} \max_{i \in D} \max_{x \in \Omega_y} \max_{x \in \Omega_y} \|x\|^2,
$$

which is the same as (22).

**REFERENCES**


