Reduced Linear Noise Approximation for Biochemical Reaction Networks with Time-Scale Separation: The Stochastic tQSSA

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>Herath, Narmada and Domitilla Del Vecchio. “Reduced Linear Noise Approximation for Biochemical Reaction Networks with Time-Scale Separation: The Stochastic tQSSA.” The Journal of Chemical Physics 148, 9 (March 2018): 094108 © 2018 Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1063/1.5012752">http://dx.doi.org/10.1063/1.5012752</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Institute of Physics (AIP)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Wed Apr 10 01:28:56 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/118988">http://hdl.handle.net/1721.1/118988</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution 4.0 International License</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a></td>
</tr>
</tbody>
</table>
Reduced linear noise approximation for biochemical reaction networks with time-scale separation: The stochastic tQSSA+
Narmada Herath, and Domitilla Del Vecchio

Citation: J. Chem. Phys. 148, 094108 (2018); doi: 10.1063/1.5012752
View online: https://doi.org/10.1063/1.5012752
View Table of Contents: http://aip.scitation.org/toc/jcp/148/9
Published by the American Institute of Physics

Articles you may be interested in

Bounds on stochastic chemical kinetic systems at steady state
The Journal of Chemical Physics 148, 084106 (2018); 10.1063/1.5009950

Relaxation rates of gene expression kinetics reveal the feedback signs of autoregulatory gene networks
The Journal of Chemical Physics 148, 095102 (2018); 10.1063/1.5009749

The chemical Langevin equation
The Journal of Chemical Physics 113, 297 (2000); 10.1063/1.481811

A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks
The Journal of Chemical Physics 148, 014105 (2018); 10.1063/1.5003892

Incorporating extrinsic noise into the stochastic simulation of biochemical reactions: A comparison of approaches
The Journal of Chemical Physics 148, 064111 (2018); 10.1063/1.5016338

Remarks on the chemical Fokker-Planck and Langevin equations: Nonphysical currents at equilibrium
The Journal of Chemical Physics 148, 064114 (2018); 10.1063/1.5016158
Reduced linear noise approximation for biochemical reaction networks with time-scale separation: The stochastic tQSSA

Narmada Herath1,a) and Domitilla Del Vecchio2,b)

1Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
2Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

(Received 7 November 2017; accepted 13 February 2018; published online 6 March 2018)

Biochemical reaction networks often involve reactions that take place on different time scales, giving rise to “slow” and “fast” system variables. This property is widely used in the analysis of systems to obtain dynamical models with reduced dimensions. In this paper, we consider stochastic dynamics of biochemical reaction networks modeled using the Linear Noise Approximation (LNA). Under time-scale separation conditions, we obtain a reduced-order LNA that approximates both the slow and fast variables in the system. We mathematically prove that the first and second moments of this reduced-order model converge to those of the full system as the time-scale separation becomes large. These mathematical results, in particular, provide a rigorous justification to the accuracy of LNA models derived using the stochastic total quasi-steady state approximation (tQSSA). Since, in contrast to the stochastic tQSSA, our reduced-order model also provides approximations for the fast variable stochastic properties, we term our method the “stochastic tQSSA+”. Finally, we demonstrate the application of our approach on two biochemical network motifs found in gene-regulatory and signal transduction networks. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/1.5012752

I. INTRODUCTION

Many biochemical processes involve reactions that occur on different time scales. For example, in bacterial cells, the binding of transcription factor to DNA takes place on the time scale of seconds, while protein production and dilution are on the order of hours. Such a separation in time scales allows the system variables to be separated into slow and fast groups, and this property can be exploited to reduce the complexity of dynamical models. In particular, for deterministic systems, the quasi-steady state approximation gives a reduced-order model for the slow variables, assuming that the fast variables rapidly reach a steady state. In the mathematical literature, a system of ordinary differential equations (ODEs) with multiple time scales is represented as slow and fast subsystems by using a small parameter $\epsilon$ to capture the separation in time scales. The mathematical treatment of such systems is given by two main methods: singular perturbation and averaging. The singular perturbation approach, formalized by Tikhonov’s theorem, involves setting $\epsilon = 0$ in the system dynamics to obtain an algebraic equation that approximates the fast variables, which is in turn used to derive an approximation for the slow variable dynamics. In the averaging method, a reduced-order model for the slow variables is obtained by the elimination of the fast dynamics via integration of the system functions.

As opposed to deterministic models, employing time-scale separation for model order reduction remains an ongoing area of research for stochastic models of biological systems. Yet, obtaining reduced descriptions of stochastic dynamics is even more critical than those for deterministic dynamics in order to increase the speed of simulation and aid analytical studies. Furthermore, accurate reduced-order models are important for precise parameter estimations. The most prominent model used to capture the stochasticity in biological systems is the Chemical Master Equation (CME), which considers the species counts as a set of discrete random variables and describes the time evolution of their probability distributions using a set of ordinary differential equations. There have been several studies that obtain reduced-order representations of the CME under time-scale separation conditions. Among these, a common approach used to reduce the complexity in the simulations of the CME is to approximate the fast variables by their deterministic quasi-steady state expressions. However, validity of this method still remains under investigation.

The theoretical analysis of the CME is challenging due to the large system size and the lack of appropriate analytical tools. Therefore, several approximations to the chemical master equation have been developed under the assumptions that the system’s volume and the molecular counts are sufficiently large. One such approximation is the Fokker-Planck equation (FPE) where partial differential equations are used to describe the evolution of the probability distribution of the species counts. Another approximation, equivalent to the FPE, is

9nherath@mit.edu
10ddv@mit.edu
The chemical Langevin equation (CLE) where the dynamics of species counts are described by stochastic differential equations.\textsuperscript{28}

There have been several recent studies that consider the problem of model order reduction for biochemical reaction networks modeled by the CLE. The work by Contou-Carrère \textit{et al.} obtains a reduced-order system for the slow variables of the CLE by adiabatic elimination of the fast variable dynamics.\textsuperscript{29} They provide a numerical analysis on the error between the full and reduced-order systems; however, an analytical error quantification is not provided. Furthermore, their work does not provide an approximation for the fast variables of the full system. In our previous work,\textsuperscript{30–32} we considered CLE models of biochemical reaction networks with linear reaction rates and obtained a reduced-order system for both slow and fast variables, following a similar approach to the deterministic singular perturbation technique. It was mathematically demonstrated that the moments of the reduced system are within an $O(\epsilon)$-neighborhood of the moments of the full system.

In addition to the above methods, the mathematical literature also offers model order reduction techniques for multiscale stochastic differential equations via averaging methods.\textsuperscript{5,33} Recently, these methods have been applied in the analysis of systems modeled by the chemical Langevin equation.\textsuperscript{34} However, averaging methods require integration of system functions, which may be challenging for systems that are of high dimension or are nonlinear. Moreover, the averaging methods also do not provide an approximation to the fast variable dynamics. In the case of biochemical reaction networks, it is typically important to approximate the fast variable dynamics, because many species are mixed—that is, their concentrations are given by the combination of slow and fast variable concentrations. Therefore, we often require both slow and fast components in the reduced-order model to correctly approximate such species dynamics.

The Linear Noise Approximation (LNA) is another approximate model for the CME, where stochasticity is represented as random fluctuations around a deterministic trajectory using stochastic differential equations or partial differential equations.\textsuperscript{10,35} Recently, model order reduction methods for the LNA have been developed using projection operators\textsuperscript{27,36} or singular perturbation analysis.\textsuperscript{37} However, in these studies, the error between the full and reduced-order models is not analytically quantified. The work by Sootla and Anderson\textsuperscript{38} provides an error quantification for model order reduction of LNA developed by Thomas \textit{et al.}\textsuperscript{27,36} but to do so they assume Lipschitz continuity of system functions, which are not Lipschitz-continuous for general systems. Furthermore, the above studies only provide an approximation for the slow variable dynamics and do not approximate the fast variables.

In this work, we consider biochemical reaction networks modeled using the LNA. We consider systems with separation of time scales where the dynamics can be represented in the singular perturbation form with $\epsilon$ as the singular perturbation parameter. We then obtain a reduced-order model that approximates both slow and fast variables. We mathematically demonstrate that first and second moments of the reduced system variables are within an $O(\epsilon)$-neighborhood of the first and second moments of the full system variables. These results, in turn, provide a rigorous justification for the accuracy of LNA models derived using the stochastic tQSSA in comparison to the standard quasi-steady state approximation (QSSA). Furthermore, in contrast to the stochastic tQSSA, our reduced system also provides approximations for the fast variables’ stochastic properties. Hence, we term our method the stochastic tQSSA*. The application of our approach is then demonstrated on two biochemical network motifs found in gene-regulatory networks and signal transduction cascades. Through these examples, we illustrate the practical applications of the reduced-order models and the necessity of both slow and fast variable approximations for analysis. In particular, using the reduced-order model for the gene-regulatory network motif, we further investigate the parameter conditions under which the standard QSSA provides accurate results in the stochastic setting.

This paper is organized as follows. In Sec. II, we present the LNA model considered in this paper. In Sec. III, we introduce the reduced-order system and present our results on the error quantification between the full and reduced dynamics. In Sec. V, we illustrate the application of the results with two examples of biochemical network motifs.

**Notation:** $\mathbb{E}[-]$ denotes the expected value of a random variable. $\|\cdot\|$ denotes the Euclidean norm for vectors and $\|\cdot\|_F$ denotes the Frobenius norm for matrices.

**II. THE LINEAR NOISE APPROXIMATION WITH TIME-SCALE SEPARATION**

**A. Linear noise approximation**

Consider a biochemical reaction network with $n$ species $Y_1, \ldots, Y_n$ in a volume $\Omega$, interacting through $m$ reactions of the form

$$p_{i1}Y_1 + \cdots + p_{im}Y_m \xrightarrow{k_i} r_{i1}Y_1 + \cdots + r_{im}Y_m,$$

where $k_i$ denotes the rate constant of reaction $i$ and $p_{ij} - r_{ij}$ is the change in the number of molecules of $Y_j$ due to the reaction $i$. Let $y(t) = [y_1, \ldots, y_n]^T$ be the state of the system at a given time $t$ where each component $y_i$ represents the molecular count for each species as a discrete random variable. Then, the Chemical Master Equation for this system is of the form

$$\frac{dP(y,t)}{dt} = \sum_{i=1}^{m} \left[ a_i(y - q_i, t)P(y - q_i, t) - a_i(y, t)P(y, t) \right],$$

where $a_i(y, t)$ is the microscopic reaction rate proportional to $k_i$ with $a_i(y, t)dt$ being the probability that the reaction $i$ will take place in an infinitesimal time step $dt$. The variable $q_i = r_i - p_i$ is the stoichiometry vector where $p_i = [p_{i1}, \ldots, p_{in}]^T$ and $r_i = [r_{i1}, \ldots, r_{in}]^T$ for $i = 1, \ldots, m$.\textsuperscript{39}

The Linear Noise Approximation (LNA) is an approximation to the CME, where the molecular counts are represented by continuous variables under the assumption that the system volume and the molecular counts are sufficiently large. As shown in the work of van Kampen,\textsuperscript{10} the LNA is derived by taking $y = \Omega v + \sqrt{\Omega} \xi$ in the CME, where $\Omega$ is the system volume, $v$ is a vector of deterministic variables, and $\xi$ is a vector...
of random variables that represents the stochastic fluctuations. Then, performing a Taylor series expansion about the deterministic variable \( \Omega \) and equating the terms of order \( \Omega^{1/2} \) and \( \Omega^1 \), it is shown that \( v \) gives the macroscopic concentrations and the elements of \( \xi \) are Gaussian random variables with the dynamics

\[
\dot{v} = f(v, t), \\
\dot{\xi} = A(v, t)\xi + \sigma(v, t)\Gamma,
\]

in which \( \Gamma \) is an \( m \)-dimensional white noise process, 
\( f(v, t) = \sum_{i=1}^m q_i \bar{a}_i(v, t), \ A(v, t) = \frac{\partial f(v, t)}{\partial \nu}, \) and \( \sigma(v, t) = [q_1 \sqrt{a_1(T^{-1}[x^T, z]^T)^T}, \ldots, q_m \sqrt{a_m(T^{-1}[x^T, z]^T)^T}]^T \) represents the reaction rates of slow reactions and \( \bar{a}_i(v, t) \) is the macroscopic reaction rate which can be approximated by \( \bar{a}_i(v, t) = \frac{1}{\Omega^0} a_i(\Omega v, t) \) as \( \Omega \to \infty \) and \( y \to \infty \) such that the concentration \( v = y/\Omega \) remains constant.\(^{40}\)

### B. System model with time-scale separation

In this work, we consider biochemical reaction networks in which the chemical reactions take place on two well-separated time scales. For the system (2) and (3), let \( m_s \) be the number of slow reactions and \( m_f \) be the number of fast reactions where \( m_s + m_f = m \). Then, by using a small parameter \( \epsilon \), the reaction rate vector can be arranged in the form \( \bar{a}(v, t) = [\bar{a}_s(v, t)/(1/\epsilon)\bar{a}_f(v, t)]^T \), where \( \bar{a}_s(v, t) \in \mathbb{R}^{m_s} \) represents the reaction rates of slow reactions and \( (1/\epsilon)\bar{a}_f(v, t) \in \mathbb{R}^{m_f} \) represents the reaction rates of fast reactions. The corresponding stoichiometry vectors \( q_i \) can be written in the form

\[
q = [q_1, \ldots, q_{m_s}, q_{m_s+1}, \ldots, q_{m_s+m_f}], \text{ where } q_i \text{ for } i = 1, \ldots, m_s \text{ represent the change in the molecular counts given by the slow reactions, and } q_i \text{ for } i = m_s + 1, \ldots, m_s + m_f \text{ represent the change in the molecular counts given by the fast reactions. Because chemical species often take part in both slow and fast reactions, the above separation in reaction rates does not necessarily correspond to a partitioning of the system’s species into fast and slow. Often, a coordinate change is necessary to identify the slow and fast variables in the system and write it in the standard singular perturbation form.}^{29,41}\]

Therefore, here we consider systems in which the species can be partitioned into \( n_s \) slow variables and \( n_f \) fast variables with \( n_s + n_f = n \), according to the following claim:

**Claim 1.** Assume that there is an invertible matrix \( T = [T_s^T, T_z^T]^T \) with \( T_s \in \mathbb{R}^{n_s \times n_s} \) and \( T_z \in \mathbb{R}^{n_f \times n_f} \) such that the change of variables \( x = T_s v, \ z = T_z v \), takes the system (2) into the singular perturbation form

\[
\dot{x} = f_s(x, z, t), \\
\epsilon \dot{z} = f_z(x, z, t, \epsilon).
\]

Then, the change of variables \( \dot{\psi}_s = T_s \dot{\xi}, \ \psi_z = T_z \dot{\xi} \) transforms the system (3) into the singular perturbation form

\[
\dot{\psi}_s = S_s(x, z, t)\psi_s + S_z(x, z, t)\psi_z + \sigma_s(x, z, t)\Gamma_x, \\
\epsilon \dot{\psi}_z = F_s(x, z, t, \epsilon)\psi_s + F_z(x, z, t, \epsilon)\psi_z + \sigma_z(x, z, t, \epsilon)\Gamma_z,
\]

where \( \Gamma_x \) is an \( m_s \)-dimensional white noise process, \( \Gamma_z \) is an \( m_f \)-dimensional white noise process and

\[
S_s(x, z, t) = \frac{\partial f_s(x, z, t)}{\partial x}, \\
S_z(x, z, t, \epsilon) = \frac{\partial f_z(x, z, t, \epsilon)}{\partial z}, \\
F_s(x, z, t, \epsilon) = \frac{\partial f_s(x, z, t, \epsilon)}{\partial x}, \\
F_z(x, z, t, \epsilon) = \frac{\partial f_z(x, z, t, \epsilon)}{\partial z}, \epsilon \dot{z} = f_z(x, z, t, \epsilon).
\]

Note that the noise \( \Gamma_x \) on the slow variables is \( m_s \)-dimensional, while the noise \( \Gamma_z \) is \( (m_s + m_f) \)-dimensional. This is due to the fact that the fast reactions do not affect the slow variable dynamics in the above form, i.e., \( \epsilon \) does not appear in the slow variable dynamics (4), but the fast dynamics can be dependent on both slow and fast reactions (see Appendix A for details).

**Proof.** See Appendix A. \( \square \)

There are several studies in the literature that have investigated the existence of such coordinate transformations that allow the separation of slow and fast variables for deterministic systems\(^{12}\) and also for chemical Langevin equation models.\(^{29}\)

Following the results of Claim 1, we consider biochemical reaction networks where the linear noise approximation model can be written in the standard singular perturbation form...
where \( x \in \mathbb{R}^n \), \( \psi_x \in \mathbb{R}^n \) are the slow variables and \( z \in \mathbb{R}^m \), \( \psi_z \in \mathbb{R}^m \) are the fast variables. \( \Gamma \) is an \( m \times n \) dimensional white noise process. Then, \( \Gamma = [\Gamma_x^T, \Gamma_z^T]^T \), where \( \Gamma_z \) is an \( m \times n \) dimensional white noise process.

We refer to the system (8)–(11) as the full system and we first make the following assumptions for \( x \in \mathbb{R}^n \) and \( z \in \mathbb{R}^m \).

**Assumption 1.** The functions \( f_x(x, z, t, \epsilon) \) are twice continuously differentiable. The Jacobian \( \frac{\partial f_x(x, z, t, \epsilon)}{\partial x} \) has continuous first and second partial derivatives with respect to its arguments.

**Assumption 2.** The matrix-valued functions \( \sigma_x(x, z, t) \) and \( \sigma_z(x, z, t) \) are continuously differentiable. Furthermore, let \( \Lambda(x, z, t, \epsilon) = \sigma_x(x, z, t) \sigma_z(x, z, t)^T \), where we have that \( \sigma_x(x, z, t) \) is continuous with respect to its arguments. Also, the initial condition \( z_0 \) is in the region of attraction of the equilibrium point \( z = \gamma_1(x_0, 0) \) for the system \( \frac{dx}{dt} = f_x(x, z, t, 0) \).

**Assumption 3.** There exists an isolated real root \( \gamma_1(x, t) \) for the equation \( f_z(x, z, t, 0) = 0 \), for which, the matrix \( \frac{\partial f_z(x, z, t, 0)}{\partial x} \) is Hurwitz (i.e., eigenvalues of the matrix have strictly negative real parts), uniformly in \( x \) and \( t \). Furthermore, we have that the first partial derivative of \( \gamma_1(x, t) \) is continuous with respect to its arguments. Also, the initial condition \( z_0 \) is in the region of attraction of the equilibrium point \( z = \gamma_1(x_0, 0) \) for the system \( \frac{dx}{dt} = f_x(x, z, t, 0) \).

**Assumption 4.** The system \( \dot{x} = f_x(x, \gamma_1(x, t), t) \) has a unique bounded solution for \( t \in [0, t_1] \).

In Sec. III, we present the reduced-order model that we define to approximate the slow and fast variables when \( \epsilon \) is small in the full system (8)–(11) and quantify the error between the moment dynamics of the full and the reduced systems.

### III. RESULTS

To define the reduced system, we follow a similar approach to the singular perturbation theory \(^3\) by setting \( \epsilon = 0 \) in the full system (8)–(11). This yields

\[
\dot{x} = f_x(x, z, t, 0),
\]

\[
F_z(x, z, t, 0)\psi_x + F_x(x, z, t, 0)\psi_z = 0.
\]

Let \( \zeta = \gamma_1(x, t) \) be an isolated root of Eq. (12). Then, it follows that the unique solution to Eq. (13) is

\[
\psi_z = -F_z(x, \gamma_1(x, t), t, 0)^{-1}F_x(x, \gamma_1(x, t), t, 0)\psi_x.
\]

Note that the invertibility of matrix \( F_z \) follows from Assumption 3 where it is assumed that \( F_z(x, \gamma_1(x, t), t, 0) = \frac{\partial f_z(x, z, t, 0)}{\partial x} |_{z=(\gamma_1(x, t))} \) is Hurwitz. Let

\[
\gamma_2(x, t) = -F_x(x, \gamma_1(x, t), t, 0)^{-1}F_z(x, \gamma_1(x, t), t, 0)\psi_z.
\]

Substituting \( z = \gamma_1(x, t) \) and \( \psi_z = \gamma_2(x, t)\psi_z \) in Eqs. (8) and (10), we obtain the following candidate approximation for the slow variable dynamics:

\[
\dot{x} = f_x(\bar{x}, \gamma_1(\bar{x}, t), t),
\]

\[
\psi_x = S(\bar{x}, t)\psi_x + \sigma_x(\bar{x}, \gamma_1(\bar{x}, t), t)\Gamma_x,
\]

\[
\psi_z(0) = \psi_{z0},
\]

\[
\psi_z(0) = \psi_{z0}.
\]

Next, to introduce a candidate approximation for the fast variable dynamics we define

\[
\bar{x} = \gamma_1(x, t),
\]

\[
\psi_x = \gamma_2(\bar{x}, \gamma_1(\bar{x}, t), t)\psi_x + g(x, t)N(0, 1),
\]

where \( N(0, 1) \) is a vector of standard normal random variables and \( g(x, t) : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^m \) is a function that satisfies the Lyapunov equation

\[
F_z(x, \gamma_1(x, t), t, 0)g(x, t)^Tg(x, t)^T = -\Lambda(x, \gamma_1(x, t), t, 0),
\]

where \( \Lambda(x, z, t, \epsilon) \) is defined in Assumption 2.

We refer to Eqs. (14)–(17) as the reduced system. Next, we show that the reduced system (14)–(17) is a good approximation of the full system (8)–(11) as \( \epsilon \) tends to zero. In particular, we demonstrate that the error in the first and second moments between the full and the reduced system variables are \( O(\epsilon) \). To this end, denote by \( \mathbb{E} \) the expected value of a random variable. Then, we have the following theorem:

**Theorem 1.** Consider the full system (8)–(11) and the reduced system in (14)–(17). Under Assumptions 1–4, there exist \( \epsilon^* \geq 0 \) such that for \( 0 < \epsilon < \epsilon^*, t \in [0, t_1] \) we have

\[
\|x(t) - \bar{x}(t)\| = O(\epsilon),
\]

\[
\|\mathbb{E}[\psi_x(t)] - \mathbb{E}[\bar{\psi}_x(t)]\| = O(\epsilon),
\]

\[
\|\mathbb{E}[\psi_x(t)] - \mathbb{E}[\bar{\psi}_x(t)]\|^2_F = O(\epsilon),
\]

and for any \( 0 < t_0 < t_1 \) there exists \( \epsilon^{**} \leq \epsilon^* \) such that for \( \epsilon < \epsilon^{**} \) and \( t \in [t_0, t_1] \) we have

\[
\|z(t) - \bar{z}(t)\| = O(\epsilon),
\]

\[
\|\mathbb{E}[\psi_z(t)] - \mathbb{E}[\bar{\psi}_z(t)]\| = O(\epsilon),
\]

\[
\|\mathbb{E}[\psi_z(t)] - \mathbb{E}[\bar{\psi}_z(t)]\|^2_F = O(\epsilon),
\]

\[
\|\mathbb{E}[\psi_z(t)] - \mathbb{E}[\bar{\psi}_z(t)]\|^2_F = O(\epsilon).
\]

The proof of this theorem is presented in Appendix B together with several intermediate results that are used in the proof. We provide an outline of the proof here. First, we derive the moment dynamics of the full system (8)–(11) and show
that they are also in the standard singular perturbation form (Lemma 1, Appendix B). Then, we derive the moment dynamics of the reduced system (14)–(17) in Lemma 2. Next, we show that setting \( \varepsilon = 0 \) in the moment dynamics of the full system yields the moment dynamics of the reduced system (Lemmas 3 and 4, Appendix B). As the moment dynamics are deterministic, we then apply Tikhonov’s theorem\(^4\) using the stability conditions provided by Assumption 3 to prove Theorem 1.

Theorem 1 shows that the reduced system (14)–(17) provides a good approximation to the slow and fast variables of the full system (8)–(11) in terms of the first and second moments of the stochastic fluctuations \( \psi_x \) and \( \psi_z \). Thus, we have that, as \( \varepsilon \) tends to zero, the mean, the variance, and the covariance of the stochastic fluctuations \( \psi_x \) and \( \psi_z \) are well approximated by those of the reduced system variables \( \bar{\psi}_x \) and \( \bar{\psi}_z \). Therefore, in the case where the full system in the singular perturbation \( \varepsilon \) setting is in the standard singular perturbation form, we then apply Tikhonov’s theorem\(^3\) and (4, Appendix B). As the moment dynamics are deterministic, we then apply Tikhonov’s theorem required for the deterministic systems.\(^4\) However, the fast variable approximation \( \bar{\psi}_z \) requires an additional term \( g(\bar{x}, t) \). This additional term is required to capture the noise properties of the fast variable, which are not captured by \( \psi_z(\bar{x}, t) \) alone, since setting \( \varepsilon = 0 \) in Eq. (11) eliminates the diffusion term \( \sigma_z(x, z, t, \varepsilon) \), which contributes to the second moment of the fast variable. To illustrate this point, we provide a simple example. Consider the system

\[
\begin{align*}
\frac{d\bar{\psi}_x}{dt} &= a_1\bar{\psi}_x + a_2\bar{\psi}_z, \\
\varepsilon \frac{d\bar{\psi}_z}{dt} &= a_3\bar{\psi}_z + \sqrt{\varepsilon}v_1\Gamma,
\end{align*}
\]

in which referring to the system (10) and (11) we have \( \sigma_x = 0 \) and \( \sigma_z = v_1 \). Setting \( \varepsilon = 0 \) in this system gives

\[
\frac{d\bar{\psi}_x}{dt} = a_1\bar{\psi}_x, \quad \bar{\psi}_z = 0.
\]

Computing the steady-state second moments of the system (26) and (27) and system (28) obtained by setting \( \varepsilon = 0 \), we have

\[
\begin{align*}
\mathbb{E}[\bar{\psi}_x^2] &= \frac{a_1^2 v_1^2}{2a_3 (1 + a_1\varepsilon)}, \\
\mathbb{E}[\bar{\psi}_z^2] &= 0, \\
\mathbb{E}[v_1^2] &= \frac{v_1^2}{2a_3}, \\
\mathbb{E}[\bar{\psi}_z^2] &= 0.
\end{align*}
\]

From here, it can be seen that as \( \varepsilon \) tends to zero the second moment of \( \bar{\psi}_x \) tends to the moment of the approximation \( \bar{\psi}_z \); however, the moment of \( \bar{\psi}_z \) does not converge to the moment of \( \bar{\psi}_z \). Thus, the reduced system obtained by setting \( \varepsilon = 0 \) provides a good approximation for the moments of the slow variable, but not for those of the fast variable.

As it can be seen from Eqs. (27) and (28), the reason for the poor approximation of the fast variable is due to the elimination of the noise term \( v_1 \) when \( \varepsilon \) is set to zero in (26) and (27).

This can be further analyzed by representing the full system in the fast time scale \( \tau = t\varepsilon \),

\[
\begin{align*}
\frac{d\bar{\psi}_x}{d\tau} &= \varepsilon (a_1\bar{\psi}_x + a_2\bar{\psi}_z), \\
\frac{d\bar{\psi}_z}{d\tau} &= a_3\bar{\psi}_z + v_1\Gamma,
\end{align*}
\]

where \( \Gamma = \sqrt{\varepsilon} \Gamma \) is a white noise process in the fast time scale. Here, we observe that setting \( \varepsilon = 0 \) does not eliminate the noise term of the fast variable dynamics, and thus, this noise term should be taken into account in the fast variable approximation. The term \( g(\bar{x}, t) \) in the reduced fast system captures this fast variable noise. This can be seen by computing the values of \( \gamma_2(\bar{x}, t) \) and \( g(\bar{x}, t)^2 \) for the system (26) and (27). From (27) and (13), we have that \( \gamma_2(\bar{x}, t) = 0 \). Writing the Lyapunov equation (18) for the system (27), we have

\[
a_3 g(\bar{x}, t)^2 + g(\bar{x}, t)^2 a_3 = \frac{v_1^2}{2}.
\]

which yields \( g(\bar{x}, t)^2 = \frac{v_1^2}{2a_3} \). This gives the steady-state moment \( \mathbb{E}[\bar{\psi}_z^2] = \frac{v_1^2}{2a_3} \) for the reduced fast system, which is

![FIG. 1. Schematic diagram illustrating the model reduction approach given by Theorem 1. The \( O(\varepsilon) \)-closeness of the moments of the original species concentrations and the moments obtained via the reduced system is proved in Appendix C.](image)
the same as \( \mathbb{E}[\psi^2] \). The dynamics of the slow variable \( \psi_z \) vary at a much slower rate than the dynamics of \( \psi_z \), and therefore, the noise of the fast variable \( \psi_z \) can be neglected in the slow variable approximation as it is essentially “filtered out.” \(^{30,32}\)

**Remark 1.** The stochastic fluctuations \( \psi_z \) and \( \dot{\psi}_z \) are multivariate Gaussian random variables and thus their probability distributions are fully characterized by the mean and the covariance. \(^{10}\) From Theorem 1, we have that the first and second moments of \( \psi_z(t) \) converge to the moments of the vector \( \dot{\psi}_z(t) \) and the first and second moments of \( \psi_z(t) \) converge to the moments of \( \dot{\psi}_z(t) \), as \( \epsilon \) tends to zero. Thus, we further have that the vectors \( \psi_z(t) \) and \( \dot{\psi}_z(t) \) converge in distribution to the vectors \( \dot{\psi}_z(t) \) and \( \dot{\psi}_z(t) \), respectively.

Next, we investigate how the reduced system derived in this section relates to the commonly used total quasi-steady state approximation in stochastic analysis.

### IV. VALIDITY OF LNA MODELS DERIVED USING STOCHASTIC tQSSA

The quasi-steady state approximation is widely used in the deterministic setting to obtain reduced-order models of biochemical reaction networks under time-scale separation conditions. Recently, the QSSA has also been extended to stochastic systems, where the fast variables are approximated by their deterministic quasi-steady state expressions in stochastic simulations. This method is termed stochastic QSSA. \(^{20,43}\)

Several studies in the deterministic setting have shown that the validity of the standard QSSA is limited to certain parameter conditions. Thus, the total quasi-steady state approximation (tQSSA) has been introduced as an alternate approximation with increased accuracy. \(^{44,45}\) The tQSSA, first proposed for enzymatic reactions, considers a coordinate change to identify the slow variables in the system. This has then been adapted in the stochastic setting where stochastic simulations of the CME are performed using the deterministic quasi-steady state expressions of the fast variables obtained using the tQSSA. \(^{20,43,46}\)

There are several studies that investigate the validity of these stochastic quasi-steady state approximations. Particularly, the work by Kim *et al.* demonstrates, via two-dimensional models modeled by the CME and LNA, that the tQSSA provides a better approximation to the original system in comparison to the standard QSSA. \(^{20,22}\) Here, we use the results derived in Theorem 1 to provide a rigorous mathematical justification for the accuracy of the stochastic tQSSA for general reaction networks modeled through the LNA. The inaccuracy of the standard QSSA in the deterministic setting results from treating the species that consists of both slow and fast dynamics as purely slow variables. By contrast, the tQSSA involves defining the slow variables of the system via a coordinate change, in which the slow variable dynamics do not depend on the fast reactions. This corresponds to Claim 1, where a coordinate transformation is used to identify the slow and fast variables in the system. Thus, the deterministic counterpart of the reduced LNA given by Eq. (14) is equivalent to the reduced system obtained by using the tQSSA. We next derive the dynamics of the stochastic fluctuations under the stochastic tQSSA. To this end, let \( \hat{x} \) and \( \hat{\psi}_x \) represent the variables in the LNA model obtained using the stochastic tQSSA.

Next, deriving the corresponding dynamics for the stochastic fluctuation using the definition of the LNA in Sec. II, we have

\[
\dot{\psi}_x = \frac{\partial f_x(\hat{x}, \gamma_1(\hat{x}, t), t)}{\partial \hat{x}} \hat{\psi}_x + \sigma_x(\hat{x}, \gamma_1(\hat{x}, t), t).
\]

Using the chain rule, we can write this system as

\[
\dot{\psi}_x = \left( \frac{\partial f_x(\hat{x}, \gamma_1(\hat{x}, t), t)}{\partial \hat{x}} + \frac{\partial f_x(\hat{x}, \gamma_1(\hat{x}, t), t)}{\partial \gamma_1(\hat{x}, t)} \frac{\partial \gamma_1(\hat{x}, t)}{\partial \hat{x}} \right) \psi_x + \sigma_x(\hat{x}, \gamma_1(\hat{x}, t), t).
\]

Then, by the implicit function theorem, \(^{47}\) we have that

\[
\frac{\partial \gamma_1(\hat{x}, t)}{\partial \hat{x}} = -\frac{\frac{\partial f_x(\hat{x}, \gamma_1(\hat{x}, t), t)}{\partial \gamma_1(\hat{x}, t)}}{\frac{\partial f_x(\hat{x}, \gamma_1(\hat{x}, t), t)}{\partial \hat{x}}}
\]

where we have from Sec. III that \(-F_x(\hat{x}, \gamma_1(\hat{x}, t), t, 0)^{-1} F_x(\hat{x}, \gamma_1(\hat{x}, t), t, 0) = \gamma_2(\hat{x}, t)\).

Substituting \(\frac{\partial \gamma_1(\hat{x}, t)}{\partial \hat{x}} = \gamma_2(\hat{x}, t)\) in (32), we obtain

\[
\dot{\psi}_x = (S_x(\hat{x}, \gamma_1(\hat{x}, t), t) + S_x(\hat{x}, \gamma_1(\hat{x}, t), t) \gamma_2(\hat{x}, t)) \psi_x + \sigma_x(\hat{x}, \gamma_1(\hat{x}, t), t),
\]

which is equivalent to the dynamics of \(\hat{\psi}_x\) in the reduced system (8) and (11) derived through singular perturbation. Thus, we have that the LNA model obtained under stochastic tQSSA is equivalent to the slow variable approximations (14) and (15) in our reduced system. This is illustrated in the schematic in Fig. 2. Then, from Theorem 1 we have that the moments of the LNA model obtained using the stochastic tQSSA provide a good approximation for the dynamics of the slow variables of the full LNA model, under all parameter conditions. Furthermore, this indicates that when the molecular counts are sufficiently large, the stochastic tQSSA yields a valid approximation for the CME under time-scale separation, similar to the observations in previous studies. \(^{20,22}\)

Therefore, our results provide a rigorous justification for the validity of stochastic tQSSA models, which has been absent in previous work. However, unlike the tQSSA, the reduced model derived in (14)–(17) also provides approximations for the fast variable stochastic properties. Therefore, we term our method the stochastic tQSSA*. These fast variables’ stochastic properties can be used to analyze the reduced dynamics of the original species concentrations. Similar considerations in the literature suggest using the prefactor QSSA \(^{48}\) method to study...
FIG. 2. Schematic diagram illustrating the relationship between the singular perturbation approach and tQSSA approach.

the dynamics of the fast variables in the reduced setting. The prefactor QSSA method, first proposed for deterministic models of gene-regulatory networks, involves first using the coordinate transformation to identify the slow and fast variables to perform the model reduction and then transforming the reduced system variables back into the original form. This is equivalent to the singular perturbation approach in deterministic setting and thus produces accurate results in deterministic models. However, Kim et al. has shown that although this provides a good approximation in deterministic models, using the reactions rates obtained through the prefactor QSSA in stochastic simulations does not produce accurate results.

By contrast, our results provide a method to obtain accurate approximations of the original system variables.

In Sec. V, we demonstrate how the results developed in this section can be applied to biochemical reaction networks with two time scales. Furthermore, we use an application example to investigate the parameter conditions under which the standard quasi-steady state provides a good approximation of the original system dynamics.

V. APPLICATION EXAMPLES

In this section, we illustrate the application of our results to two biochemical network motifs that exhibit time-scale separation. First, we consider a gene-regulatory network motif. We derive the reduced LNA and validate it through numerical simulations. We then use the reduced system to investigate the validity of quasi-steady state approximations that are often performed in the stochastic setting. As a second example, we consider a cellular signaling motif.

A. Example I: Gene-regulatory network motif

We consider the gene-regulatory network motif shown in Fig. 3, in which two genes are regulating each other in a negative feedback loop. Gene-regulatory networks describe the interactions between genes and the proteins that regulate the expressions of these genes. The regulatory proteins, known as transcription factors, can activate or repress the expression of a gene by binding to its promoter site and by either aiding or interfering with the transcription process by RNA polymerase. Through these activation and repression processes, the cell regulates the levels of proteins in the cell, which are responsible for a vast majority of cellular functions.

In deterministic models of gene-regulatory networks, activation and repression processes are often modeled by Hill functions. These are derived using the QSSA where it is assumed that the binding reactions between the DNA promoter sites and transcription factors equilibrate faster than transcription and translation. Similarly, it has become common practice to use the deterministic quasi-steady state expressions (e.g., Michaelis-Menten/Hill functions) in the simulation of stochastic biochemical reaction networks as a convenient way to eliminate the fast dynamics and reduce simulation time. However, the validity of such approximations in the stochastic setting remains under investigation. In Sec. IV, we demonstrated that the reduced LNA model obtained using the stochastic tQSSA is equivalent to the slow variable dynamics in the reduced system (14)–(17) and thus provides a good approximation for the full system variables under all parameter conditions. In this example, we derive the reduced LNA for the system in Fig. 3 and use it to investigate the validity of reduced-order LNA models obtained through the stochastic standard QSSA.

Consider the gene-regulatory network motif in Fig. 3. We have protein M activating the production of protein G, which in turn represses the production of protein M. The chemical reactions are given by

![Chemical reactions diagram]

Protein G represses the production of M by binding to the promoter P1 leading to the inactive complex C1. The production of protein M is given by the second reaction, in which we have lumped together transcription and translation. Protein M decays at rate δ1, which accounts for degradation and dilution.

Similarly, the chemical reactions that describe the expression and decay of protein G can be written as

![Chemical reactions diagram]
where $M$ activates the production of protein $G$ by first binding to promoter $P_2$ to produce the transcriptionally active complex $C_2$. We assume that the total concentration of promoters is conserved, giving $p_{r1} = p_1 + c_1$, $p_{r2} = p_2 + c_2$ where the lower-case letters represent macroscopic concentrations. Then, the macroscopic reaction rate equations can be written as

\[
\begin{align*}
\frac{dM}{dt} &= \beta_1 p_{r1} - c_1 - \delta_1 m - k_{on1}m(p_{r2} - c_2) + k_{off2}c_2, \\
\frac{dc_1}{dt} &= k_{on1}g(p_{r1} - c_1), \\
\frac{dc_2}{dt} &= k_{on2}m(p_{r2} - c_2) - k_{off2}c_2, \\
\frac{dg}{dt} &= \beta_2 c_2 - \delta_2 g - k_{on1}g(p_{r1} - c_1) + k_{off1}c_1.
\end{align*}
\]

(34)

Since binding and unbinding reactions are much faster than production and decay, we can define the small parameter $\epsilon = \delta_1/k_{off1}$ where $\epsilon \ll 1$. Let $k_{d1} = k_{off1}k_{on1}$ and $k_{d2} = k_{off2}k_{on2}$ with $a = k_{off2}/k_{off1}$. Then, considering the species vector $v = [m, g, c_1, c_2]$, we can partition the reaction rate vector into slow and fast groups as $\bar{d}(v, t) = [\bar{d}_s(v, t), \bar{d}_f(v, t)]$, where $\bar{d}_s(v, t) = [\beta_1 p_{r1} - c_1, \delta_1 m, \beta_2 c_2, \delta_2 g]$ and $\bar{d}_f(v, t) = [\frac{\delta_1}{k_{d1}}g(p_{r1} - c_1), \delta_1 c_1, \frac{\delta_2}{k_{d2}}g(p_{r2} - c_2), \delta_2 c_2]$, and the corresponding stoichiometry matrix is given by

\[
q = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 1 \\
0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & -1 & 1 & 1
\end{bmatrix}.
\]

The variables $m$ and $g$ are not fast nor slow because $M$ and $G$ participate in both slow and fast reactions. Thus, we consider the coordinate change $w = m + c_2$, $u = g + c_1$, which takes the system (34) into singular perturbation form

\[
\begin{align*}
\frac{dw}{dt} &= \beta_1 p_{r1} - c_1 - \delta_1 (w - c_2), \\
\frac{du}{dt} &= \beta_2 c_2 - \delta_2 (u - c_1), \\
\frac{dc_1}{dt} &= \frac{\delta_1}{k_{d1}}(u - c_1)(p_{r1} - c_1) - \delta_1 c_1, \\
\frac{dc_2}{dt} &= \frac{\delta_2}{k_{d2}}(w - c_2)(p_{r2} - c_2) - a\delta_1 c_2,
\end{align*}
\]

(35)

where the slow variables are given by $w$ and $u$ and the fast variables are given by $c_1$ and $c_2$. Considering Claim 1, the change of coordinates $w = m + c_2$, $u = g + c_1$ corresponds to $T_x = [1 0 0 1, 0 1 1 0]^T$, $T_{zv} = [0 0 0 1, 0 0 0 1]^T$, where $x = [w, u]^T$ and $z = [c_1, c_2]^T$. Then, as described in Sec. II, the dynamics of the stochastic fluctuations of the LNA can be written using (35) as

\[
\begin{align*}
\frac{dw}{dt} &= \delta_1 p_{r1} - \beta_1 p_{r1} - \delta_1 w + \sqrt{\beta_1(p_{r1} - c_1)} \Gamma_1 \\
\frac{du}{dt} &= -\delta_2 w + \beta_2 c_2 + \sqrt{\beta_2 c_2} \Gamma_2, \\
\frac{dc_1}{dt} &= \frac{\delta_1}{k_{d1}}(p_{r1} - c_1) - \delta_1 c_1 - \sqrt{\delta_2(u - c_1)} \Gamma_3, \\
\frac{dc_2}{dt} &= \frac{\delta_2}{k_{d2}}(w - c_2) + \sqrt{\delta_1 c_1} \Gamma_4 + \sqrt{\epsilon \delta_2(u - c_1)} \Gamma_5, \\
\frac{dg}{dt} &= \frac{a\delta_1}{k_{d2}}(p_{r2} - w - c_2) + \epsilon \delta_1 \Gamma_6 + \sqrt{\epsilon \delta_2(w - c_2)} \Gamma_7.
\end{align*}
\]

(36)

We note that the system functions are given by polynomials of the system variables, and thus Assumptions 1 and 2 are satisfied. Considering Assumption 3, we have that $f_z$ yields the unique solution $[c_1, c_2]^T = [(\frac{\delta_1}{k_{d1}}(u - c_1)(p_{r1} - c_1) - \delta_1 c_1, \frac{\delta_2}{k_{d2}}(w - c_2)(p_{r2} - c_2) - a\delta_1 c_2]^T$ is Hurwitz for all parameter values, and setting $\epsilon = 0$ in the fast variable dynamics $f_z$ yields the solution $[c_1, c_2]^T = [(\frac{\delta_1}{k_{d1}}(u - c_1)(p_{r1} - c_1) - \delta_1 c_1, \frac{\delta_2}{k_{d2}}(w - c_2)(p_{r2} - c_2) - a\delta_1 c_2]^T$ which is feasible under conditions $0 \leq c_1 \leq p_{r1}$ and $0 \leq c_2 \leq p_{r2}$. Next, in order to derive the reduced system, we set $\epsilon = 0$ in the system (35) and (36) which yields

\[
\begin{align*}
c_1 &= \frac{(u + p_{r1} + k_{d1})}{2} - \sqrt{(u + p_{r1} + k_{d1})^2 - 4up_{r1}}, \\
c_2 &= \frac{(w + p_{r2} + k_{d2})}{2} - \sqrt{(w + p_{r2} + k_{d2})^2 - 4wp_{r2}}, \\
\psi_{c1} &= \frac{(p_{r1} - u - 2c_1)\psi_u}{((p_{r1} - u - 2c_1) + k_{d1})^2}, \\
\psi_{c2} &= \frac{(p_{r2} - c_2)\psi_w}{((p_{r2} + w - 2c_2) + k_{d2})^2}.
\end{align*}
\]

(37)

Next, in order to determine the additional terms in the fast variable approximations for the stochastic variables $\psi_{c1}$ and $\psi_{c2}$, we write the Lyapunov equation (18) as
which yields
\begin{equation}
g(w) = \begin{bmatrix}
\sqrt{(\bar{u} - \bar{c}_1)(\bar{p}_{T1} - \bar{c}_1) + k_d\bar{c}_1} \\
2\sqrt{(\bar{u} + \bar{p}_{T1} + k_d)^2 - 4\bar{u}\bar{p}_{T1}} \\
0 \\
\sqrt{(\bar{w} - \bar{c}_2)(\bar{p}_{T2} - \bar{c}_2) + k_d\bar{c}_2} \\
2\sqrt{(\bar{w} + \bar{p}_{T2} + k_d)^2 - 4\bar{w}\bar{p}_{T2}}
\end{bmatrix}.
\end{equation}

Then, substituting the terms (37) in the system (35) and (36) and using \( g(w) \) and a vector of normal random variables \([N_1(0, 1), N_2(0, 1)]^T\) according to (17), we obtain the reduced system
\begin{align}
\frac{d\bar{w}}{dt} &= \beta_1(p_{T1} - \bar{c}_1) - \delta_1(\bar{w} - \bar{c}_2), \\
\frac{d\bar{u}}{dt} &= \beta_2\bar{c}_2 - \delta_2(\bar{u} - \bar{c}_1), \\
\frac{d\bar{g}_w}{dt} &= -\delta_1\bar{g}_w - \frac{\beta_1(p_{T1} - \bar{c}_1)}{p_{T1} + \bar{u} - 2\bar{c}_1 + k_d1} + \delta_1\frac{(p_{T2} - \bar{c}_2)\bar{g}_w}{p_{T2} + \bar{w} - 2\bar{c}_2 + k_d2} + \sqrt{\beta_1(p_{T1} - \bar{c}_1)}\Gamma_1 - \sqrt{\delta_1(\bar{w} - \bar{c}_2)}\Gamma_2, \\
\frac{d\bar{g}_u}{dt} &= -\delta_2\bar{g}_u + \frac{\beta_2(p_{T1} - \bar{c}_1)}{p_{T1} + \bar{u} - 2\bar{c}_1 + k_d1} + \frac{\beta_2(p_{T2} - \bar{c}_2)}{p_{T2} + \bar{w} - 2\bar{c}_2 + k_d2} + \sqrt{\beta_2\bar{c}_2}\Gamma_3 - \sqrt{\delta_2(\bar{u} - \bar{c}_1)}\Gamma_4, \\
\bar{c}_1 &= \frac{(\bar{c}_1 + p_{T1} + k_d1)}{2} - \frac{\sqrt{(\bar{u} + p_{T1} + k_d1)^2 - 4\bar{u}p_{T1}}}{2}, \\
\bar{c}_2 &= \frac{(\bar{w} + p_{T2} + k_d2)}{2} - \frac{\sqrt{(\bar{w} + p_{T2} + k_d2)^2 - 4\bar{w}p_{T2}}}{2}, \\
\bar{\psi}_{c1} &= \frac{\beta_1(p_{T1} - \bar{c}_1)\bar{u}_w}{(p_{T1} + \bar{u} - 2\bar{c}_1 + k_d1)} + \frac{\sqrt{(\bar{u} - \bar{c}_1)(p_{T1} - \bar{c}_1) + k_d1\bar{c}_1}}{2\sqrt{(\bar{u} + p_{T1} + k_d1)^2 - 4\bar{u}p_{T1}}} N_1(0, 1), \\
\bar{\psi}_{c2} &= \frac{\beta_2(p_{T2} - \bar{c}_2)\bar{u}_w}{(p_{T2} + \bar{w} - 2\bar{c}_2 + k_d2)} + \frac{\sqrt{(\bar{w} - \bar{c}_2)(p_{T2} - \bar{c}_2) + k_d2\bar{c}_2}}{2\sqrt{(\bar{w} + p_{T2} + k_d2)^2 - 4\bar{w}p_{T2}}} N_2(0, 1).
\end{align}

Next \( c_1 = \frac{g_{p_{T1}}}{\bar{g} + k_d1}, \quad c_2 = \frac{mp_{T2}}{m + k_d2} \).

Then, substituting these in system (34), we obtain the reduced deterministic dynamics,
\begin{align}
\frac{d\bar{m}}{dt} &= \beta_1 \frac{p_{T1}k_d1}{\bar{g} + k_d1} - \delta_1\bar{m}, \\
\frac{d\bar{g}}{dt} &= \beta_2 \frac{p_{T2}k_d2}{\bar{m} + k_d2} - \delta_2\bar{g}.
\end{align}

Next, we derive the dynamics for the corresponding stochastic fluctuations as explained in Sec. II, which yields
\begin{align}
\frac{d\bar{\psi}_m}{dt} &= -\delta_1\bar{\psi}_m - \beta_1 \frac{p_{T1}k_d1}{(\bar{g} + k_d1)^2}\bar{\psi}_g + \sqrt{\beta_1} \frac{p_{T1}k_d1}{\bar{g} + k_d1} \Gamma_1 \\
&\quad + \sqrt{\delta_1\bar{m}}\Gamma_2, \\
\frac{d\bar{\psi}_g}{dt} &= -\delta_2\bar{\psi}_g + \beta_2 \frac{p_{T2}k_d2}{(\bar{m} + k_d2)^2}\bar{\psi}_m + \sqrt{\beta_2} \frac{\bar{m}p_{T2}}{\bar{m} + k_d2} \Gamma_3 \\
&\quad + \sqrt{\delta_2\bar{g}}\Gamma_4.
\end{align}

Thus, the Michaelis-Menten function based LNA model is given by the system (39)–(42). The dynamics for the second moments of \( \bar{\psi}_m \) and \( \bar{\psi}_g \) are given by
where we have defined \( R_1(\bar{g}) = \frac{p_{T1} k_{d1}}{\bar{g} + k_{d1}} \) and \( R_2(\bar{m}) = \frac{p_{T2} k_{d2}}{\bar{m} + k_{d2} + 1} \).

Next, we use the reduced-order model developed in our work to precisely identify the conditions under which the above Michaelis-Menten function based LNA model provides a good approximation for the statistical properties of the original system. To this end, we use the reduced system (38) to derive the moment dynamics (see Appendix D for details) as

\[
\begin{align*}
\frac{dE[\bar{\psi}_m^2]}{dt} &= -2\delta_1 E[\bar{\psi}_m^2] - 2R_1(\bar{g})\beta_1 E[\bar{\psi}_m \bar{\psi}_g] + \beta_1 \frac{p_{T1} k_{d1}}{\bar{g} + k_{d1}} + \delta_1 \bar{m}, \\
\frac{dE[\bar{\psi}_g^2]}{dt} &= -2\delta_2 E[\bar{\psi}_g^2] + 2R_2(\bar{m})\beta_2 E[\bar{\psi}_m \bar{\psi}_g] + \beta_2 \frac{\bar{m} p_{T1}}{\bar{m} + k_{d1}} + \delta_2 \bar{g}, \\
\frac{dE[\bar{\psi}_m \bar{\psi}_g]}{dt} &= (-\delta_1 - \delta_2) E[\bar{\psi}_m \bar{\psi}_g] - R_1(\bar{g})\beta_1 E[\bar{\psi}_g^2] + R_2(\bar{m})\beta_2 E[\bar{\psi}_m^2], \\
\end{align*}
\]

\[
\begin{align*}
\frac{d\bar{m}}{dt} &= \frac{1}{1 + \frac{R_1(\bar{g})}{R_2(\bar{m})}} \left( \beta_1 \frac{p_{T1} k_{d1}}{\bar{g} + k_{d1}} - \delta_1 \bar{m} \right), \\
\frac{d\bar{g}}{dt} &= \frac{1}{1 + \frac{R_1(\bar{g})}{R_2(\bar{m})}} \left( \beta_2 \frac{p_{T2}}{\bar{m} + k_{d2}} - \delta_2 \bar{g} \right), \\
\frac{dE[\bar{\psi}_m^2]}{dt} &= \frac{1}{1 + \frac{R_1(\bar{g})}{R_2(\bar{m})}} \left( -2\delta_1 + F_2(\bar{m}) \right) E[\bar{\psi}_m^2] \\
&\quad + 2R_1(\bar{g})\beta_1 E[\bar{\psi}_m \bar{\psi}_g] + \left( -G_2(\bar{m}) + 1 \right) \delta_1 \bar{m}, \\
\frac{dE[\bar{\psi}_g^2]}{dt} &= \frac{1}{1 + \frac{R_1(\bar{g})}{R_2(\bar{m})}} \left( -2\delta_2 + F_1(\bar{g}) \right) E[\bar{\psi}_g^2] \\
&\quad + 2R_2(\bar{m})\beta_2 E[\bar{\psi}_m \bar{\psi}_g] + \left( -G_1(\bar{g}) + 1 \right) \delta_2 \bar{g},
\end{align*}
\]
\[
\frac{d\mathbb{E}[\tilde{v}_m\tilde{v}_S]}{dt} = \left( -\frac{\delta_1}{1 + R_2(\tilde{m})} - \frac{\delta_2}{1 + R_1(\tilde{g})} + H_2(\tilde{m}) \right)
+ \frac{H_1(\tilde{g})}{1 + R_1(\tilde{g})} \mathbb{E}[\tilde{v}_m\tilde{v}_S] - \frac{R_1(\tilde{g})\beta_1\mathbb{E}[\tilde{v}_S^2]}{1 + R_2(\tilde{m})} + R_2(\tilde{m})\beta_2 \mathbb{E}[\tilde{v}_m^2] - I(\tilde{m}, \tilde{g}),
\]

where
\[
F_i(y) = 4 - \frac{R_i(y)}{(y + k_d)^2} \frac{dy}{dt},
\]
\[
G_i(y) = \frac{2yR_i(y)(1 + 2R_i(y))}{(y + k_d)(1 + R_i(y))^2},
\]
\[
H_i(y) = \frac{2R_i(y)}{(y + k_d)(1 + R_i(y))} \frac{dy}{dt},
\]
\[
I(\tilde{m}, \tilde{g}) = \frac{1}{1 + R_2(\tilde{m})} \frac{1}{1 + R_1(\tilde{g})} (\beta_1 g R_1(\tilde{g})^2 + \beta_2 \tilde{v}_m R_2(\tilde{m})^2).
\]

The boxed terms in (44)–(48) highlight the extra terms compared to the Michaelis-Menten function based moment dynamics (39), (40), and (43). From this, it follows that the Michaelis-Menten function based model of (39)–(42) becomes a good approximation of the full LNA model only when the boxed terms are sufficiently small. In order to analyze the conditions under which the boxed terms become small, we first note from Eqs. (44) and (45) that
\[
\tilde{m}(t) \leq \max[\tilde{m}(0), \beta_1 p_{T_1}/\delta_1] := b_{\tilde{m}}
\]
and
\[
\tilde{g}(t) \leq \max[\tilde{g}(0), \beta_2 p_{T_2}/\delta_2] := b_{\tilde{g}}
\]
for all \(t\). Thus, we have that
\[
\frac{d\tilde{m}}{dt} \leq \beta_1 p_{T_1} + \delta_1 \max[\tilde{m}(0), \beta_1 p_{T_1}/\delta_1] := b_{\tilde{d}m}
\]
and
\[
\frac{d\tilde{g}}{dt} \leq \beta_2 p_{T_2} + \delta_2 \max[\tilde{g}(0), \beta_2 p_{T_2}/\delta_2] := b_{\tilde{d}g}.
\]

Then, considering the terms \(F_1(\tilde{g})\) and \(F_2(\tilde{m})\), we have that
\[
\left| F_1(\tilde{g}) \right| \leq 2 \left| \frac{dR_1(\tilde{g})}{d\tilde{g}} \right| (\beta_2 p_{T_2} + \delta_2 \max[\tilde{g}(0), \beta_2 p_{T_2}/\delta_2]),
\]
\[
\left| F_2(\tilde{m}) \right| \leq 2 \left| \frac{dR_2(\tilde{m})}{d\tilde{m}} \right| (\beta_1 p_{T_1} + \delta_1 \max[\tilde{m}(0), \beta_1 p_{T_1}/\delta_1]).
\]

From this, we have that for given bounds \(b_{\tilde{d}m}\) and \(b_{\tilde{d}g}\), \(|F_i(y)|\) terms become small when the terms \(\left| \frac{dR_i(\tilde{y})}{d\tilde{y}} \right|\) are sufficiently small. Similarly, from the expressions of \(R_i(y), G_i(y), H_i(y),\) and \(I(\tilde{m}, \tilde{g})\), we have that for given bounds \(b_{\tilde{m}}, b_{\tilde{g}}, b_{\tilde{d}m}\) and \(b_{\tilde{d}g}\), the terms \(|H_i(y)|\) become small when the terms \(\left| \frac{dR_i(y)}{dy} \right|\) are sufficiently small and the terms \(G_i(y)\), \(I(\tilde{m}, \tilde{g})\) become small when \(R_i(y)\) terms are sufficiently small. Thus, in this case, we can conclude that when \(R_i(y)\) and \(\left| \frac{dR_i(y)}{dy} \right|\) are sufficiently small, the Michaelis-Menten function based moment dynamics (39), (40), and (43) become close to the moment dynamics (44)–(48), which is in turn a good approximation of the moments of the original system when \(\epsilon\) is small. Thus, it follows that the Michaelis-Menten function based models becomes a good approximation of the original system when \(\epsilon\) becomes small and when \(R_i(y)\) and \(\left| \frac{dR_i(y)}{dy} \right|\) are also small. We illustrate this via simulations in Fig. 5, which shows that the moments of the Michaelis-Menten function based model become close to those of the singular perturbation-based model when the dissociation constants \(k_{d1}\) and \(k_{d2}\) are large [making \(R_i(y)\) and \(\left| \frac{dR_i(y)}{dy} \right|\) small].

The terms \(R_1(\tilde{g})\) and \(R_2(\tilde{m})\) in the deterministic dynamics (44) and (45) have been previously studied in the context of modularity in gene regulatory networks. In particular, Del Vecchio et al.\textsuperscript{49,51} showed that the terms of the form \(R_1(\tilde{g})\) and \(R_2(\tilde{m})\) physically arise due to the “loading” that promoter binding sites apply to their transcription factor regulators. These effects, termed retroactivity, cause a change in the dynamics of transcription factors upon binding to the DNA promoter sites and are not captured by Michaelis-Menten function based models alone. Thus, the results obtained in this section indicate that, similar to deterministic systems, retroactivity effects also impact the stochastic dynamics of the system.

As previously described in Sec. IV, the use of the deterministic QSSA in stochastic models has been previously studied by several researchers in the context of CME and LNA. In particular, the study by Kim et al.\textsuperscript{48} considers two dimensional systems (with one slow variable and one fast variable) modeled by the CME and points to two main sources of inaccuracy for the stochastic models obtained through the standard QSSA.\textsuperscript{20} The first source of inaccuracy arises from treating species that contain both slow and fast variables as purely slow variables (such as the free proteins M and G in this example). This is also a source of inaccuracy in the deterministic QSSA as noted by several studies.\textsuperscript{48,51} The second source of inaccuracy is disregarding the noise of the fast variable when taking the QSSA. In line with this, Kim et al.\textsuperscript{48} analyzed the steady state distribution and the variance of the fast variable and determine that the stochastic QSSA becomes accurate when the sensitivity of the quasi-steady state expression to changes of the slow variable is small. This sensitivity term is equivalent to the retroactivity terms \(R_1(\tilde{g})\) and \(R_2(\tilde{m})\) in our model. Thus, our results recapitulate the findings of Kim’s steady state analysis on systems with two species, but, being based on general and rigorous derivation of the moment dynamics, our results are applicable to systems with more than two species and can also be used to study transient dynamics, as we have shown in this example.

Additionally, a similar study has been performed by Thomas et al.,\textsuperscript{27} where a reduced-order approximation for the LNA is proposed. They analyze several enzyme and gene network motifs and determine that the discrepancy between the quasi-steady state LNA models and the reduced-order LNA models proposed in their work arises mainly due to disregarding the noise of the fast variables. However, the reduced-order LNA developed in their work also regards some species concentrations that are the combination of fast and slow variables as purely slow variables. Therefore, their transient dynamics may not be accurately approximated by their reduction. In contrast to this, here, we identify conditions for
FIG. 5. Moments of the Michaelis-Menten function based LNA model and the singular perturbation based LNA model. The parameter values are $\beta_1 = 1 \text{ hr}^{-1}$, $\beta_2 = 2 \text{ hr}^{-1}$, $\delta_1 = 1 \text{ hr}^{-1}$, $\delta_2 = 3 \text{ hr}^{-1}$, $p_{T1} = 10 \text{ nM}$, and $p_{T2} = 10 \text{ nM}$.

which both the transient dynamics and the steady state are accurate.

The analysis in this section, being based on mathematical derivations as opposed to simulations, therefore provides a general criterion to determine when the quasi-steady state LNA of gene-regulatory networks is close to the original system. The fast variable approximations can also be used to analyze how the noise of the transcription factor-DNA complexes change with retroactivity effects, which can give insights into the interplay between modularity and noise in gene-regulatory networks.

B. Example II: Signal transduction network motif

As another example, we consider the dynamics of species involved in signal transduction via phosphorylation cycles. Phosphorylation cycles are network motifs that form signal transduction cascades in cells. They play a central role in cell physiology for transmitting signals from outside the cell to initiate cellular responses by activating target gene expression. Furthermore, engineered phosphorylation systems have also been used in synthetic biology applications to transmit signals while buffering from loading effects that were discussed in the previous example.\(^{41,51,53,54}\) Here, we consider the system shown in Fig. 6. Protein M is phosphorylated by kinase Z and produces protein $M^*$, which is dephosphorylated by phosphatase Y. The phosphorylated protein $M^*$ acts as a transcription factor that activates protein G.

Phosphorylation and dephosphorylation processes can be modeled by the one-step reactions\(^ {49}\)

$$M + Z \xrightarrow{k_1} M^* + Z, \quad M^* + Y \xrightarrow{k_2} M + Y,$$

where $k_1$ and $k_2$ are the rate constants. The binding between protein $M^*$ and promoter P produces a complex C,
where \( k_{\text{on}} \) and \( k_{\text{off}} \) are the binding and unbinding rate constants. Then, the production and decay of protein G is given by

\[
C \rightarrow C + G, \quad \delta \rightarrow \phi,
\]

where \( \beta \) is the production rate constant and \( \delta \) is the decay rate constant. The total concentration of protein M and promoter P is conserved, giving \( M_T = m + m^* + c \) and \( p_T = p + c \), where we use lower-case letters to denote the corresponding macroscopic concentrations. Then, the macroscopic reaction rate equations for this system can be written as

\[
\begin{align*}
\frac{dm^*}{dt} &= k_1Z(t)(M_T - m^* - c) - k_2Ym^* - k_{\text{on}}m^*(p_T - c) + k_{\text{off}}c, \\
\frac{dc}{dt} &= k_{\text{on}}m^*(p_T - c) - k_{\text{off}}c, \\
\frac{dg}{dt} &= \beta c - \delta g,
\end{align*}
\]

where we consider the concentration of the kinase given by \( Z(t) \) to be a deterministic input to the system.

As binding and unbinding reactions are much faster than phosphorylation and dephosphorylation reactions, we have that \( k_{\text{off}} \gg k_2Y \). Thus, we can define the small parameter \( \epsilon = k_2Y/k_{\text{off}} \). Let \( k_d = k_{\text{off}}/k_{\text{on}} \) be the dissociation constant for the binding reaction between M* and P. Then, as described in Sec. II, we can consider the species vector \( v = [m^*, g, c]^T \) and partition the reaction rate vector into slow and fast groups in the form \( \hat{a}(y, t) = [\hat{a}_s(v, t), (1/\epsilon)\hat{a}_f(y, t)] \), where \( \hat{a}_s(v, t) = [k_1Z(t)(M_T - m^* - c), k_2Ym^*, \beta c, \delta g] \) and \( \hat{a}_f(v, t) = [k_2Ym^*(p_T - c), k_2Yc] \), with the corresponding stoichiometry matrix given by

\[
q = \begin{bmatrix}
1 & -1 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0
\end{bmatrix}.
\]

This yields the system dynamics

\[
\begin{align*}
\frac{dm^*}{dt} &= k_1Z(t)(M_T - m^* - c) - k_2Ym^* - \frac{k_2Y}{\epsilon}(p_T - c) + \frac{k_2Y}{\epsilon}c, \\
\frac{dg}{dt} &= \beta c - \delta g, \\
\frac{dc}{dt} &= \frac{k_2Y}{\epsilon}(m^*(p_T - c) - \frac{k_2Y}{\epsilon}c).
\end{align*}
\]

Although, the singular perturbation parameter appears in the system of equations (49), we note that the slow and fast dynamics are not well separated and the system is not in the standard singular perturbation form given in (8) and (9). Thus, we consider the change of coordinates \( w = m^* + c \), which yields

\[
\begin{align*}
\frac{dw}{dt} &= k_1Z(t)(M_T - w) - k_2Y(w - c), \\
\frac{dg}{dt} &= \beta c - \delta g, \\
\frac{dc}{dt} &= \frac{k_2Y}{k_d}(w - c)(p_T - c) - k_2Yc,
\end{align*}
\]

where we have that the slow variables are \( w \) and \( g \), and the fast variable is given by \( c \). Referring to Claim 1, we have that the coordinate change \( w = m^* + c \) corresponds to \( T_x = [1 \ 0 \ 1 \ 0 \ 1 \ 0]^T \), \( T_z = [0 \ 0 \ 1 \ 0 \ 1 \ 0]^T \). Therefore, from (50) we can write the following equations for the dynamics of the stochastic fluctuations as explained in Sec. II:

\[
\frac{dy_{w}}{dt} = (-k_1Z(t) - k_2Y)\psi_{w} + k_2Y\psi_{c} + \sqrt{k_2Y}Z(t)(M_T - w + c)\Gamma_1 - \sqrt{k_2Y}(w - c)\Gamma_2,
\]

\[
\frac{dy_{c}}{\epsilon dt} = \beta \psi_{c} - \delta \psi_{g} + \sqrt{\beta c} \Gamma_3 - \sqrt{\delta g} \Gamma_4,
\]

\[
\frac{dy_{g}}{\epsilon dt} = k_2Y(p_T - c) - \frac{k_2Yp_T}{k_d} \psi_{w} + \left( -\frac{k_2Yp_T}{k_d} + \frac{k_2Yv}{k_d} + \frac{k_2Y}{k_d}2c - k_2Y \right) \psi_{c} + \sqrt{\epsilon} \frac{k_2Y(w - c)(p_T - c)\Gamma_5 - \sqrt{\epsilon k_2Yc}\Gamma_6}.
\]

Next, we derive the reduced-order dynamics of the system (50) and (51) using Theorem 1. From (50), it follows that the system functions are polynomials of the state variables. Therefore, Assumptions 1 and 2 are satisfied. Solving \( f_c = \frac{k_2Y}{k_d}(w - c)(p_T - c) - k_2Yc = 0 \) yields the unique solution \( y_1(w) = \frac{1}{2}(w + p_T + k_d) - \frac{1}{2}\sqrt{(w + p_T + k_d)^2 - 4wp_T} \), feasible under the physical constraints \( 0 \leq c \leq p_T \). We note that \( \frac{dy_{c}}{\epsilon dt} \) is negative for all parameter values and thus Assumption 3 is satisfied. Next, to determine the fast variable approximation for \( \psi_{c} \) in the form of Eq. (17), we write

\[
\begin{align*}
-\frac{k_2Yp_T}{k_d} + \frac{k_2Yv}{k_d} + \frac{k_2Y}{k_d}2c - k_2Y \cdot g(w)g(w)^T \\
+ g(w)g(w)^T \left( -\frac{k_2Yp_T}{k_d} + \frac{k_2Yv}{k_d} + \frac{k_2Y}{k_d}2c - k_2Y \right)
\end{align*}
\]

which yields

\[
g(w) = \sqrt{\frac{k_2Y}{k_d}(w - c)(p_T - c) + k_2Yc} \\
\sqrt{\frac{k_2Y}{k_d}(w + p_T + k_d)^2 - 4wp_T}.
\]
Then, the reduced system is given by

\[
\frac{d\bar{w}}{dt} = k_1 Z(t)(M_T - \bar{w}) - k_2 Y(\bar{w} - c),
\]

\[
\frac{d\bar{g}}{dt} = \beta \bar{c} - \delta \bar{g},
\]

\[
\frac{d\bar{\psi}_w}{dt} = (-k_1 Z(t) - k_2 Y)\bar{\psi}_w + \frac{k_2 Y(p_T - \bar{c})}{(p_T + \bar{w} - 2\bar{c} + k_d)} \bar{\psi}_w + \sqrt{k_1 Z(t)(M_T - \bar{w} + \bar{c})} \Gamma_1 - \sqrt{k_2 Y(\bar{w} - \bar{c})} \Gamma_2,
\]

\[
\frac{d\bar{\psi}_c}{dt} = \bar{\psi}_c = \frac{(p_T - \bar{c})}{(p_T + \bar{w} - 2\bar{c} + k_d)} \bar{\psi}_w + \frac{k_2 Y}{k_d} \frac{1}{2} \frac{(\bar{w} - \bar{c})(p_T - \bar{c}) + k_2 Y \bar{c}}{2k_y^{\bar{c}} \sqrt{(\bar{w} + p_T + k_d)^2 - 4p_T \bar{w}}} N(0, 1).
\]

FIG. 7. Sample moments of the full and reduced systems obtained by numerically simulating the systems (50)–(52). The simulations were performed using the Euler-Maruyama method for the stochastic differential equations and the moments are computed using the average of 300,000 simulation runs. The parameter values are \(Z(t) = 10 + 8.5 \sin (6t) \text{ nM}, k_1 = 0.06 \text{ nM h}^{-1}, k_2 = 0.6 \text{ nM h}^{-1}, k_d = 50 \text{ nM}, M_T = 100 \text{ nM}, Y = 10 \text{ nM}, p_T = 50 \text{ nM}, \delta = 6 \text{ hr}^{-1}, \beta = 10 \text{ hr}^{-1}, w(0) = 0, c(0) = 0, \bar{g}(0) = 0, \bar{\psi}_w(0) = 0, \bar{\psi}_c(0) = 0, \) and \(\psi_{\epsilon}(0) = 0.\)
Figure 7 illustrates simulation results for the second moments of $\psi_w$, $\psi_s$, and $\psi_c$.

As $\epsilon$ tends to zero, the moments of the full system tend to the moments of the reduced system in accordance to Theorem 1.

The reduced system derived in this section can be used to mathematically analyze the noise properties of signals transmitted through phosphorylation cycles. A common measure of noise is the coefficient of variation, which is defined as the ratio of standard deviation to the mean of a random variable. The moment dynamics of the reduced system can therefore be used in the computation of such noise measures. Finally, note that since the species that carry the signal $M^+$ is a mixed (fast and slow) species, both fast and slow variable approximations are necessary to analyze noise propagation in these signal transduction systems.

VI. CONCLUSION

In this work, we addressed the problem of model order reduction for biochemical reaction networks with time-scale separation, where the system dynamics are modeled with the LNA. After transforming the system into a standard singular perturbation form, we developed a reduced-order model that approximates the slow and fast dynamics of the full system when the time-scale separation is large. In particular, we showed that the error between the moment dynamics of the full system and the reduced systems is $O(\epsilon)$, where $\epsilon$ is the singular perturbation parameter that captures the time-scale separation. This error quantification only requires stability of the fast variables boundary layer dynamics, which is a simple condition to check and the same condition required in deterministic singular perturbation. Different from existing work, we have also presented an approximation of the fast variables. Furthermore, we show that the slow variable dynamics in our reduced-order model are equivalent to the reduced model obtained using the stochastic tQSSA approximation. Thus, our results also provide a rigorous justification for the validity of the LNA models obtained using the stochastic tQSSA. When the molecular counts are sufficiently large, this further provides a justification for the validity of the stochastic tQSSA in the CME. Since our reduction, different from the stochastic tQSSA, also provides an approximation of the fast variable stochastic properties, we have termed it the stochastic tQSSA*.

We have detailed the application of our work to two examples of biochemical reaction networks: a gene-regulatory network motif and a signal transduction module. For these examples, we derived the reduced-order LNA and verified the analytical results through numerical simulations. Through these examples, we highlighted the necessity of both slow and fast variable approximations for practical applications of the model reduction framework. Finally, for the gene-regulatory network motif, we identified conditions under which commonly used stochastic quasi-steady state models provide a good approximation. Our results can be used to substantially reduce the dimensionality of stochastic models of biochemical reaction networks, thus aiding analytical quantification of noise and simulation time. Furthermore, our results would also be useful in parameter estimations where accurate reduced-order models are required to obtain accurate and precise parameter estimations.

ACKNOWLEDGMENTS

This work was funded by AFOSR Grant No. FA9550-14-1-0060.

APPENDIX A: PROOF OF CLAIM 1

Applying the coordinate transformation $x = T_x v$, $z = T_z v$ to Eq. (2), with $\tilde{a}(v, t) = [\tilde{a}_q(v, t), (1/\epsilon)\tilde{d}_q(v, t)]^T$ and $q = [q_1, \ldots, q_m, q_{m+1}, \ldots, q_{m+m_f}]$, with $v = T^{-1}[x^T, z^T]^T$, we have

$$\dot{x} = T_x f(T^{-1}[x^T, z^T]^T, t) = T_x \sum_{i=1}^{m_f} q_i \tilde{a}_i(T^{-1}[x^T, z^T]^T, t)$$

$$+ T_x \sum_{i=m_f+1}^{m_f+m_s} q_i(1/\epsilon)\tilde{d}_i(T^{-1}[x^T, z^T]^T, t)$$

$$= f_s(x, z, t),$$

(A1)

$$\dot{z} = T_z f(T^{-1}[x^T, z^T]^T, t) = T_z \sum_{i=1}^{m_s} \tilde{a}_i(T^{-1}[x^T, z^T]^T, t)$$

$$+ T_z \sum_{i=m_f+1}^{m_f+m_s} q_i(1/\epsilon)\tilde{d}_i(T^{-1}[x^T, z^T]^T, t)$$

$$= \frac{1}{\epsilon} f_s(x, z, t).$$

(A2)

Thus, from Eq. (A1), it follows that $T_x q_i = 0$ for $i = m_s + 1, \ldots, m_f + m_s$. Thus, the fast reactions do not appear in the slow dynamics; however, the slow reactions can appear in the fast dynamics.

Applying the coordinate transformation $\psi_x = T_x \xi$, $\psi_z = T_z \xi$, to Eq. (3), we have that

$$\dot{\psi}_x = T_x[A(v, t)\xi] + T_x \sigma(v, t)\Gamma,$$

$$\dot{\psi}_z = T_z[a(v, t)\xi] + T_z \sigma(v, t)\Gamma.$$

Since $A(v, t) = \frac{\partial f(v, t)}{\partial v}$ and $v = T^{-1}[x^T, z^T]^T$, using the chain rule we can write

$$\dot{\psi}_x = T_x \left[ \frac{\partial f(T^{-1}[x^T, z^T]^T, t)}{\partial x} \frac{\partial x}{\partial \psi_x} + \frac{\partial f(T^{-1}[x^T, z^T]^T, t)}{\partial z} \frac{\partial z}{\partial \psi_x} \right] \xi + T_x \left[ q_1 \sqrt{\tilde{a}_1(T^{-1}[x^T, z^T]^T, t)}, \ldots, q_m \sqrt{\tilde{a}_m(T^{-1}[x^T, z^T]^T, t)} \right] \Gamma,$$

$$\dot{\psi}_z = T_z \left[ \frac{\partial f(T^{-1}[x^T, z^T]^T, t)}{\partial x} \frac{\partial x}{\partial \psi_z} + \frac{\partial f(T^{-1}[x^T, z^T]^T, t)}{\partial z} \frac{\partial z}{\partial \psi_z} \right] \xi + T_z \left[ q_1 \sqrt{\tilde{a}_1(T^{-1}[x^T, z^T]^T, t)}, \ldots, q_m \sqrt{\tilde{a}_m(T^{-1}[x^T, z^T]^T, t)} \right] \Gamma.$$

Using the linearity of the differentiation operator and the transformation $x = T_x v$, $z = T_z v$, we obtain
where $\Gamma = [\Gamma^T_1, \Gamma^T_f T]$. From (A1) we have that $T_e q_i = 0$ for $i = m + 1, \ldots, m + m_f$. Then, multiplying (A4) by $\epsilon$ and taking $\Gamma_\epsilon = [\Gamma^T_1, \Gamma^T_f T]$, we can write the system (A3) and (A4) in the form of the system (6) and (7).

**APPENDIX B: PROOF OF THEOREM 1**

In this section, we present the Proof of Theorem 1. We first derive a set of intermediate results, given in Claims 2–4 and Lemma 1, which will be used in the proof. In Claim 2, we derive the moment dynamics of the full system and show that these moment dynamics are also in the singular perturbation form. In Claim 3, we derive the moment dynamics of the reduced system. Next, in Claim 4, we show that setting $\epsilon = 0$ in the moment dynamics of the full system yields the moment dynamics of the reduced system. Then, these results are summarized in Lemma 1, which is then used to prove Theorem 1.

**Lemma 1.** The first and second moment dynamics for the variables $\psi_x$ and $\psi_z$ of the full system (8)–(11) can be expressed in the singular perturbation form

\[
\begin{align*}
\dot{\psi}_x &= \frac{\partial E[\psi_x]}{\partial x}, \\
\dot{\psi}_z &= \frac{\partial E[\psi_x, \psi_z^T]}{\partial x} + \frac{\partial E[\psi_x, \psi_z^T]}{\partial z} + \epsilon \frac{\partial E[\psi_x]}{\partial t}, \\
\epsilon \frac{dE[\psi_x]}{dt} &= F_x(x, z, t, e)E[\psi_x] + F_z(x, z, t, e)E[\psi_z], \\
\epsilon \frac{dE[\psi_x, \psi_z^T]}{dt} &= F_x(x, z, t, e)E[\psi_x, \psi_z^T] + F_z(x, z, t, e)E[\psi_x, \psi_z^T] + F_z(x, z, t, e)E[\psi_x, \psi_z^T].
\end{align*}
\]

where $x$ and $z$ are the solutions of Eqs. (8) and (9), and the initial conditions are given by $E[\psi_x(0)] = \psi_{x0}$, $E[\psi_x, \psi_z^T(0)] = \psi_{x0} \psi_{z0}^T$, $E[\psi_x(0)] = \psi_{z0}$, and $E[\psi_x, \psi_z^T(0)] = \psi_{x0} \psi_{z0}^T, E[\psi_x, \psi_z^T(0)] = \psi_{z0} \psi_{z0}^T$.

**Proof.** We express Eqs. (10) and (11) in the form

\[
\begin{align*}
\dot{\psi}_x &= S_x(x, z, t)\psi_x + S_z(x, z, t)\psi_z + [\sigma_x(x, z, t) 0] \Gamma_\epsilon, \\
\epsilon \frac{dE[\psi_x]}{dt} &= F_x(x, z, t, e)\psi_x + F_z(x, z, t, e)\psi_z + \sigma_x(x, z, t, e)\Gamma_\epsilon,
\end{align*}
\]
where \( \sigma_z(x, z, t) \) is deterministic, we use the linearity of the expectation operator to derive the dynamics for the first moments as

\[
\frac{dE[\psi_x]}{dt} = S(x, z, t)E[\psi_x] + S_z(x, z, t)E[\psi_z],
\]

where \( \psi \) to derive the dynamics for the first moments as

\[
\frac{dE[\psi_x]}{dt} = \frac{1}{\epsilon} S(x, z, t, \epsilon)E[\psi_x] + \frac{1}{\epsilon} S_z(x, z, t, \epsilon)E[\psi_z].
\]

Similarly, Proposition III.1 in the work of Bence et al. can be used to write the second moment dynamics as

\[
\frac{dE[\psi_x^2]}{dt} = \frac{1}{\epsilon} (S(x, z, t, \epsilon)E[\psi_x^2] + \frac{1}{\epsilon} S_z(x, z, t, \epsilon)E[\psi_z^2])
\]

Then, summing the corresponding entries of the matrices in Eq. (B8) and using the linearity of the expectation operator, Eqs. (B6)–(B8) can be written in the form (B1)–(B5). We have that \( E[\psi_x^2] = (E[\psi_x^2])^T \), and thus, we do not consider the dynamics of the variable \( E[\psi_x^2] \) in Eqs. (B1)–(B5). Furthermore, the initial conditions \( \psi_{x0} \) and \( \psi_{z0} \) are deterministic, which yields \( E[\psi_x(0)] = \psi_{x0}, E[\psi_x^2(0)] = \psi_{x0}^2, \) and \( E[\psi_z(0)] = \psi_{z0}, E[\psi_z^2(0)] = \psi_{z0}^2 \).

Next, we derive the moment dynamics of the reduced system (14)–(17).

\[ \text{Lemma 2. The first and second moment dynamics for the variable } \bar{\psi}_x \text{ of the reduced system (14) and (15) can be written in the form} \]

\[
\frac{dE[\bar{\psi}_x]}{dt} = \frac{1}{\epsilon} S(x, z, t, \epsilon)E[\bar{\psi}_x] + \frac{1}{\epsilon} S_z(x, z, t, \epsilon)E[\bar{\psi}_z],
\]

and the first and second moments for the variable \( \bar{\psi}_x \) of the reduced system (14) and (15) can be written in the form

\[ E[\bar{\psi}_x] = \gamma_2(\bar{x}, t)E[\psi_x], \quad E[\bar{\psi}_x^2] = \gamma_2^2(\bar{x}, t)E[\psi_x^2] \]

We have that the dynamics of \( \bar{x} \) in (14) are deterministic. Therefore, using the linearity of the expectation operator, the moment dynamics of \( \bar{\psi}_x \) in (15) can be written in the form (B9) and (B10).

In order to derive the moment of the variable \( \bar{\psi}_z \), we take the expected value of Eq. (16), which yields \( E[\bar{\psi}_z] = E[g(\bar{x}, t)\bar{\psi}_x] \) as the elements of the vector \( N(0, 1) \) are normal random variables with zero mean. Since \( \bar{x} \) is deterministic, we have that \( E[\bar{\psi}_z] = \gamma_2(\bar{x}, t)E[\psi_x] \). Calculating the second moment of \( \bar{\psi}_z \), we obtain

\[ E[\bar{\psi}_z^2] = E[(\gamma_2(\bar{x}, t)\psi_x + g(\bar{x}, t)N(0, 1))(\gamma_2(\bar{x}, t)\psi_x + g(\bar{x}, t)N(0, 1))^T]. \]

Expanding further and using the fact that \( N(0, 1) \) is independent of \( \psi_x \), we have

\[ E[\bar{\psi}_z^2] = \gamma_2^2(\bar{x}, t)E[\psi_x^2] \gamma_2(\bar{x}, t)^T + g(\bar{x}, t)E[N(0, 1)] \gamma_2(\bar{x}, t)^T \]

\[ + g(\bar{x}, t)E[N(0, 1)]^T \gamma_2(\bar{x}, t) \]

Since \( N(0, 1) \) is a vector of standard normal random variables, we have that \( E[N(0, 1)] = 0 \) and \( E[N(0, 1)]^T = I_{d \times d} \) where \( I_{d \times d} \) is an \( d \times d \) identity matrix. Thus, we have that \( E[\bar{\psi}_z^2] = \gamma_2^2(\bar{x}, t)E[\psi_x^2] \gamma_2(\bar{x}, t)^T + g(\bar{x}, t)g(\bar{x}, t)^T \).

Next, we derive the set of reduced-order moments obtained by setting \( \epsilon = 0 \) in the moment dynamics of the full system (B1)–(B5).

\[ \text{Lemma 3. Setting } \epsilon = 0 \text{ in the system of moment dynamics (B1)–(B5) and the dynamics of } x \text{ and } z \text{ given by (8) and (9)} \]

\[ \text{yields the moment dynamics of the reduced system (B9)–(B12) where the dynamics of } \bar{x} \text{ and } \bar{z} \text{ are given by (14) and (16), respectively.} \]

\[ \text{Proof. Setting } \epsilon = 0, \text{ we have} \]
0 = f_{c}(x, z, t, 0), \quad (B13)
0 = F_{c}(x, z, t, 0)\mathbb{E}[\psi_{c}] + F_{c}(x, z, t, 0)\mathbb{E}[\psi_{c}], \quad (B14)
0 = F_{c}(x, z, t, 0)\mathbb{E}[\psi_{c}^{T}T] + F_{c}(x, z, t, 0)\mathbb{E}[\psi_{c}^{T}T], \quad (B15)
0 = F_{c}(x, z, t, 0)\mathbb{E}[\psi_{c}^{T}T] + F_{c}(x, z, t, 0)\mathbb{E}[\psi_{c}^{T}T] + \mathbb{E}[\psi_{c}^{T}T]F_{c}(x, z, t, 0)^{T} + \mathbb{E}[\psi_{c}^{T}T]F_{c}(x, z, t, 0)^{T} + \Lambda(x, z, t, 0). \quad (B16)

Under Assumption 3, there exists an isolated real root \( z = \gamma_{1}(x, t) \) for Eq. (B13). Thus, the corresponding unique solutions to Eqs. (B14) and (B15) are given by

\[
\begin{align*}
\mathbb{E}[\psi_{c}] &= -F_{c}(x, \gamma_{1}(x, t), t, 0)^{-1}(F_{c}(x, \gamma_{1}(x, t), t, 0)\mathbb{E}[\psi_{c}]) \\
&= \gamma_{2}(x, t)\mathbb{E}[\psi_{c}], \quad (B17) \\
\mathbb{E}[\psi_{c}^{T}T] &= -F_{c}(x, \gamma_{1}(x, t), t, 0)^{-1}(F_{c}(x, \gamma_{1}(x, t), t, 0)\mathbb{E}[\psi_{c}^{T}T]) \\
&= \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]. \quad (B18)
\end{align*}
\]

Substituting \( z = \gamma_{1}(x, t) \) and Eqs. (B17) and (B18) in (8) and the moment Eqs. (B1)–(B5) yields the dynamics of \( \hat{x} \) given by (14) and the moment dynamics of the variable \( \hat{\psi}_{c} \) reduced system given by (B9) and (B10).

Furthermore, we have that Eq. (B17) is equivalent to the first moment of the variable \( \hat{\psi}_{c} \) of the reduced system given by

\[
F_{c}(x, \gamma_{1}(x, t), t, 0)\gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]\gamma_{2}(x, t)^{T} + F_{c}(x, \gamma_{1}(x, t), t, 0)g(x, t)g(x, t)^{T} + \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]F_{c}(x, \gamma_{1}(x, t), t, 0)^{T} + \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]F_{c}(x, \gamma_{1}(x, t), t, 0)^{T} + \Lambda(x, \gamma_{1}(x, t), t, 0).
\]

Then, canceling the common terms on both sides, we obtain

\[
F_{c}(x, \gamma_{1}(x, t), t, 0)g(x, t)g(x, t)^{T} + \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]F_{c}(x, \gamma_{1}(x, t), t, 0)^{T} + \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]F_{c}(x, \gamma_{1}(x, t), t, 0)^{T} + \Lambda(x, \gamma_{1}(x, t), t, 0)
\]

which satisfies Eq. (18) in the definition of the reduced fast system.

We then have the following result.

Lemma 4. Consider the full system in (8)–(11), the reduced system in (14) and (15), and the moment dynamics for the full and reduced systems in (B1)–(B5) and (B9)–(B12), respectively. We have that, under Assumptions 1–3, the commutative diagram in Fig. 8 holds.

Proof. Proof follows from Lemmas 1–3.

Although Lemma 1 shows that the setting \( \epsilon = 0 \) in the moment dynamics of the full system yields the moment dynamics of the reduced system, it does not guarantee that the trajectories of the moments become close to each other as \( \epsilon \) decreases. Therefore, we next use Tikhonov’s theorem to prove that the moments of the reduced system are within an \( O(\epsilon) \)-neighborhood of the moments of the full system.

Proof of Theorem 1. It can be seen from the commutative diagram in Lemma 1 that setting \( \epsilon = 0 \) in the moment equations of the full system together with the dynamics of \( x \) and \( z \) yields the moment equations of the reduced system. Thus, as the moment dynamics are deterministic, we can apply Tikhonov’s theorem to the moments of the full system together with the dynamics of \( x \) and \( z \) to prove the results given in Theorem 1. Towards this end, we first prove that the assumptions of Tikhonov’s theorem are satisfied. We first consider the boundary layer dynamics of the moment dynamics (B1)–(B5), where we define the boundary layer variables as

\[
\begin{align*} 
&b_{1} = z - \gamma_{1}(x, t), \\
&b_{2} = \mathbb{E}[\psi_{c}] - \gamma_{2}(x, t)\mathbb{E}[\psi_{c}], \\
&b_{3} = \mathbb{E}[\psi_{c}^{T}T] - \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T], \\
&b_{4} = \mathbb{E}[\psi_{c}^{T}T] - (\gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]g(x, t)^{T} + g(x, t)g(x, t)^{T}).
\end{align*}
\]

Define \( \gamma_{3}(x, \mathbb{E}[\psi_{c}^{T}T], t) = \gamma_{2}(x, t)\mathbb{E}[\psi_{c}^{T}T]g(x, t)^{T} + g(x, t)g(x, t)^{T} \). Then, we have that the dynamics of the boundary layer variables are given by
Then, substituting for the moment dynamics from Eqs. (9) and (B3)–(B5), we obtain
\[
\frac{db_1}{d\tau} = f_z(x, z, t, \epsilon) - \epsilon \frac{\partial \gamma_1(x, t)}{\partial t} - \epsilon \frac{\partial \gamma_1(x, t)}{\partial x} \frac{dx}{dt},
\]
\[
\frac{db_2}{d\tau} = F_z(x, z, t, \epsilon) + F_z(x, z, t) E[\psi_z] - \epsilon E[\psi_z] \frac{\partial \gamma_2(x, t)}{\partial t} - \epsilon E[\psi_z] \frac{\partial \gamma_2(x, t)}{\partial x} \frac{dx}{dt}
- \epsilon \gamma_2(x, t) \frac{dE[\psi_z]}{dt}.
\]
\[
\frac{db_3}{d\tau} = \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial t} + \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial x} \frac{dx}{dt} - \epsilon \gamma_3(x, E[\psi_z]) \frac{dE[\psi_z]}{dt}
- \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial t} - \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial x} \frac{dx}{dt}.
\]
Then, substituting for the moment dynamics from Eqs. (9) and (B3)–(B5), we obtain
\[
\frac{db_1}{d\tau} = f_z(x, z, t, \epsilon) - \epsilon \frac{\partial \gamma_1(x, t)}{\partial t} - \epsilon \frac{\partial \gamma_1(x, t)}{\partial x} \frac{dx}{dt},
\]
\[
\frac{db_2}{d\tau} = F_z(x, z, t, \epsilon) + F_z(x, z, t) E[\psi_z] - \epsilon E[\psi_z] \frac{\partial \gamma_2(x, t)}{\partial t} - \epsilon E[\psi_z] \frac{\partial \gamma_2(x, t)}{\partial x} \frac{dx}{dt}
- \epsilon \gamma_2(x, t) \frac{dE[\psi_z]}{dt}.
\]
\[
\frac{db_3}{d\tau} = \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial t} + \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial x} \frac{dx}{dt} - \epsilon \gamma_3(x, E[\psi_z]) \frac{dE[\psi_z]}{dt}
- \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial t} - \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial x} \frac{dx}{dt}.
\]
Taking \( \tau = t/\epsilon \) to be the time variable in the fast time scale, we have that
\[
\frac{db_1}{d\tau} = \frac{dz}{d\tau} - \frac{\partial \gamma_1(x, t)}{\partial t} - \frac{\partial \gamma_1(x, t)}{\partial x} \frac{dx}{dt},
\]
\[
\frac{db_2}{d\tau} = \frac{dE[\psi_z]}{d\tau} - \frac{\partial \gamma_2(x, t)}{\partial t} - \frac{\partial \gamma_2(x, t)}{\partial x} \frac{dx}{dt}
- \gamma_2(x, t) \frac{dE[\psi_z]}{dt},
\]
\[
\frac{db_3}{d\tau} = \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial t} + \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial x} \frac{dx}{dt} - \epsilon \gamma_3(x, E[\psi_z]) \frac{dE[\psi_z]}{dt}
- \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial t} - \epsilon \frac{\partial \gamma_3(x, E[\psi_z])}{\partial x} \frac{dx}{dt}.
\]
Next, in order to prove that the origin is an exponentially stable equilibrium point of the boundary layer system, we linearize the system (B28)--(B31) about the origin. Towards this end, we first represent the matrix variable $b_1$ and $b_3$ in vector form. Let $A$ be an $m \times n$ matrix and let $\text{vec}(A) = [a_1, \ldots, a_m, a_1, \ldots, a_n]^T$, where $a_0$ are the elements of the matrix $A$. Then, considering the dynamics of $\text{vec}(b_3)$ and $\text{vec}(b_4)$, we use the Kronecker product denoted by $\otimes$ to obtain

\[
\begin{align*}
\frac{db_1}{d\tau} &= \bar{F}_z(x, b_1 + \gamma_1(x, t), t, 0)\mathbb{E}[\psi_x]
+ \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)(b_2 + 2\gamma_2(x, t)\mathbb{E}[\psi_x])^T, \\
\frac{db_2}{d\tau} &= \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)\mathbb{E}[\psi_x]
+ \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)(b_3 + 2\gamma_2(x, t)\mathbb{E}[\psi_x])^T \\
\end{align*}
\]

where

\[
\begin{align*}
g_2(b_1, x, \psi_x, t) &= \bar{F}_z(x, b_1 + \gamma_1(x, t), t, 0)\mathbb{E}[\psi_x^T]
+ \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)\gamma_2(x, t)\mathbb{E}[\psi_x^T], \\
g_3(b_1, x, \psi_x, t) &= \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)\mathbb{E}[\psi_x^T] \gamma_2(x, t)^T \\
+ \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)\gamma_3(x, t)\mathbb{E}[\psi_x^T],
\end{align*}
\]

and linearizing about $\dot{e} = 0$, we obtain the dynamics of $\dot{e} = e - \dot{e}$ in the form

\[
\begin{align*}
\frac{d\hat{e}}{d\tau} &= \begin{bmatrix} J_{11} & 0 & 0 & 0 \\
J_{21} & J_{22} & 0 & 0 \\
J_{31} & J_{32} & 0 & 0 \\
J_{41} & J_{42} & J_{43} & J_{44} \end{bmatrix} \hat{e},
\end{align*}
\]

where the diagonal entries are given by $J_{11} = \frac{\partial f(x, b_1 + \gamma_1(x, t), 0)}{\partial b_1}|_{b_1 = 0}$, $J_{22} = \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)|_{b_1 = 0}$, $J_{33} = (I \otimes \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0))|_{b_1 = 0}$, and $J_{44} = (F_x(x, b_1 + \gamma_1(x, t), t, 0) \otimes \bar{F}_x(x, b_1 + \gamma_1(x, t), t, 0)|_{b_1 = 0}$ where $\otimes$ denotes the Kronecker sum and the lower diagonal entries $J_{21}$, $J_{31}$, $J_{41}$, $J_{43}$ are appropriate functions.

Since the eigenvalues of a block triangular matrix are given by the union of eigenvalues of the diagonal blocks, we consider the eigenvalues of the diagonal entries. We have that $F_x(x, z, t, \epsilon) = \frac{\partial f(x, z, t, 0)}{\partial x} |_{z=\gamma_1(x, t)}$ is Hurwitz from Assumption 3. Furthermore, we have that $F_x(x, z, t, \epsilon) = \frac{\partial f(x, z, t, 0)}{\partial x} |_{z=\gamma_1(x, t)}$ is Hurwitz under Assumption 3. Thus, the diagonal term $J_{22}$ is Hurwitz. Considering the eigenvalues of $J_{33}$, we have that any eigenvalue of a Kronecker product of two matrices is given by the product of the eigenvalues of the individual matrices. Thus, the eigenvalues of $J_{33}$ are given by the eigenvalues of $F_x(x, b_1 + \gamma_1(x, t), t, 0)|_{b_1 = 0}$, which is Hurwitz under Assumption 3. Next, we consider the diagonal term $J_{44}$. From Theorem 13.16 of the work of Laub, we have that any eigenvalue of $J_{44}$ is given by the sum of two eigenvalues of $F_x(x, b_1 + \gamma_1(x, t), t, 0)$. Since the matrix $F_x$ is Hurwitz under Assumption 3, we have that all eigenvalues of $J_{44}$ have negative real parts. Therefore, we have that the eigenvalues of the linearized system (B32) have negative real parts and thus the origin is an exponentially stable equilibrium point uniformly for $x$ and $t$.

Next, we prove that the initial conditions of the system

\[
\begin{align*}
\frac{db_1}{d\tau} &= f_z(x_0, b_1 + \gamma_1(x_0, 0), 0, 0), \\
\frac{db_2}{d\tau} &= f_x(x_0, b_1 + \gamma_1(x_0, 0), 0, 0)\psi_{x_0}
+ F_z(x_0, b_1 + \gamma_1(x_0, 0), 0, 0)(b_2 + 2\gamma_2(x_0, 0)\psi_{x_0}),
\end{align*}
\]

are in the region of attraction of the equilibrium point at the origin. From Assumption 3, we have that the initial condition $z_0$ is in the region of attraction of the equilibrium point $z = \gamma_1(x_0, 0)$ of system $\frac{dz}{d\tau} = f(x_0, z, 0, 0)$. Thus, it follows that $z_0 = \gamma_1(x_0)$ is in the region of attraction of equilibrium point of $b_1 = 0$ for the system (B33). Furthermore, we have that the trajectory $b_1(\tau)$ that corresponds to the initial condition $z_0 = \gamma_1(x_0, 0)$ is
bounded and \( \lim_{\tau \to +\infty} b_1(\tau) = 0 \). Using this fact and the linearity of the system (B34)–(B36), we next prove that any trajectory of the system, starting with any initial condition, converges to zero as \( \tau \to \infty \). Towards this end, we first define the vector \( r = [\hat{b}_2, \text{vec}(b_3), \text{vec}(b_4)] \) and write the system (B34)–(B36) in the form

\[
\frac{d r}{d \tau} = \begin{bmatrix}
H_{11} & 0 & 0 \\
0 & H_{32} & 0 \\
0 & H_{32} & H_{33}
\end{bmatrix} r + \begin{bmatrix}
C_{11}(\tau) & 0 & 0 \\
0 & C_{22}(\tau) & 0 \\
0 & C_{32}(\tau) & C_{33}(\tau)
\end{bmatrix} r + \begin{bmatrix}
0 \\
D_{2}(\tau) \\
D_{3}(\tau)
\end{bmatrix},
\]

where

\[
\begin{align*}
H_{11} &= F_{1}(x_0, \gamma_1(x_0, 0), 0, 0), \\
H_{32} &= I \otimes F_{2}(x_0, \gamma_1(x_0, 0), 0, 0), \\
H_{33} &= I \otimes F_{2}(x_0, \gamma_1(x_0, 0), 0, 0) + F_{2}(x_0, \gamma_1(x_0, 0), 0, 0) \otimes I, \\
C_{11}(\tau) &= F_{1}(x_0, b_1(\tau) - \gamma_1(x_0, 0, 0)) - F_{1}(x_0, \gamma_1(x_0, 0, 0)), \\
C_{22}(\tau) &= I \otimes F_{2}(x_0, b_1(\tau) + \gamma_1(x_0, 0, 0)), \\
C_{33}(\tau) &= I \otimes F_{2}(x_0, b_1(\tau) + \gamma_1(x_0, 0, 0)) + F_{2}(x_0, b_1(\tau) + \gamma_1(x_0, 0, 0)) \otimes I - I \otimes F_{2}(x_0, \gamma_1(x_0, 0, 0)), \\
D_{2}(\tau) &= \text{vec}(g_2(b_1(\tau), x_0, \psi_{x_0})), \\
D_{3}(\tau) &= \text{vec}(g_3(b_1(\tau), x_0, \psi_{x_0})).
\end{align*}
\]

Then, we apply Lemmas 9.4–9.6 from the work of Khalil to show that \( \tau \) tends to zero as \( \tau \to \infty \) for any initial condition \( r(0) \). From Assumption 3, we have that the matrix \( H \) is Hurwitz, and thus the system \( \frac{d r}{d \tau} = H r \) is globally exponentially stable. Therefore, there exists a Lyapunov function \( V(r, \tau) = r^T P(\tau) r \) that satisfies (9.3)–(9.5) from the work of Khalil. Furthermore, we have that \( ||C(\tau) h + D(\tau)|| \leq ||C(\tau)|| h || \) and \( ||D(\tau)|| \). We note that \( ||D(\tau)|| \) is bounded as the functions \( g_2(b_1(\tau), x_0, \psi_{x_0}) \) and \( g_3(b_1(\tau), x_0, \psi_{x_0}) \) are continuous in \( b_1(\tau) \) from Assumptions 1 and 2, and \( b_1(\tau) \) is bounded due to the asymptotic stability of the equilibrium point \( \bar{b}_1 = 0 \). Furthermore, since \( \lim_{\tau \to \infty} b_1(\tau) = 0 \), it follows from the definitions of the functions \( g_2 \) and \( g_3 \) that \( ||C(\tau)|| \to 0 \) and \( ||D(\tau)|| \to 0 \). Therefore, from Lemmas 9.5, 9.6 and 9.4 in the work of Khalil, we have that \( \lim_{\tau \to \infty} \frac{d r}{d \tau} = 0 \) for all \( r(0) \in \mathbb{R}^3 \). Thus, it follows that the region of attraction for the system (B34)–(B36) is given by \( R_{b1} \times \mathbb{R}^3 \) where \( R_{b1} \) is the region of attraction of the equilibrium point \( x_0 - \gamma_1(x_0, 0) \).

Next, we consider the remaining assumptions of Tikhonov’s theorem. We have that functions \( f_1, f_2, S_1, S_2, F_1, F_2, \sigma_1, \sigma_2, \sigma_1^T, \sigma_2^T \), and \( \sigma_1^T, \sigma_2^T \) and their first partial derivatives are continuously differentiable from Assumptions 1 and 2. Under Assumption 1, we also have that \( \frac{\partial f_i(x, t, 0)}{\partial c_i} \), \( \frac{\partial f_i(x, t, 0)}{\partial \psi_{x_i}} \), and \( \frac{\partial f_i(x, t, 0)}{\partial \psi_{x_i}^T} \) have continuous first partial derivatives with respect to their arguments. From Assumptions 1 and 3, the first partial derivatives of \( \gamma_1(x, t), \gamma_2(x, t) \), \( \gamma_1(x, t) \), and \( \gamma_1(x, t) \) with respect to their arguments are also continuous. Under Assumption 4, there exists a unique, bounded solution for the reduced system (14) for \( t \in [0, t_1] \). Furthermore, as the moment dynamics (B9) and (B10) are linear in the variables \( \mathbb{E}[\psi_{x_i}], \mathbb{E}[\hat{\psi}_{x_i}], \mathbb{E}[\psi_{x_i}^T], \) there exists a unique, bounded solution to (B9) and (B10) for \( t \in [0, t_1] \). Thus, the assumptions of Tikhonov’s theorem are satisfied and applying the theorem to the set of moment equations in (B1)–(B5) and (B9)–(B12) yields the result (19)–(25).

**APPENDIX C: CONVERGENCE OF MOMENTS OF THE ORIGINAL VARIABLES**

Here, we prove that the moments of the reduced system (14)–(17) can provide a good approximation for the moments of the original variables \( v \) and \( \xi \) in the system (2) and (3). We only provide a complete proof for the variable \( v \) since the proof for \( \xi \) and \( \mathbb{E}[\xi^2] \) can be derived in a similar manner. From Eqs. (2) and (3), we have that \( v \) and \( \xi \) represent the original variables, and from Claim 1 we have that

\[
v = T^{-1} [x^T, \zeta^T]^T \quad \text{and} \quad \xi = T^{-1} [\psi_{x_1}^T, \psi_{x_2}^T]^T.
\]

Then, let \( \bar{v} \) and \( \bar{\xi} \) represent the species concentrations obtained using the reduced system (14)–(17). Therefore, we have that

\[
\bar{v} = T^{-1} [x^T, \bar{\zeta}^T]^T \quad \text{and} \quad \bar{\xi} = T^{-1} [\bar{\psi}_{x_1}^T, \bar{\psi}_{x_2}^T]^T.
\]

Then, computing the error between the moments of the original species concentrations and the moments obtained using the reduced system, we have

\[
\| v - \bar{v} \| = \left\| T^{-1} \begin{bmatrix} x \\ \zeta \end{bmatrix} - T^{-1} \begin{bmatrix} \bar{x} \\ \bar{\zeta} \end{bmatrix} \right\|
\]

and using the definition of the Euclidean norm, we obtain

\[
\| v - \bar{v} \| \leq \| T^{-1} \| \left\| \begin{bmatrix} x - \bar{x} \\ \zeta - \bar{\zeta} \end{bmatrix} \right\| ,
\]

from Theorem 1, we have that \( \| x - \bar{x} \| \leq c_1 \epsilon \) and \( \| \zeta - \bar{\zeta} \| \leq c_2 \epsilon \) for appropriate constants \( c_1 \) and \( c_2 \). Thus, from (C1), it follows that

\[
\| v - \bar{v} \| \leq C \epsilon,
\]

for an appropriate constant \( C \). Similar result holds for the variables \( \mathbb{E}[\xi] \) and \( \mathbb{E}[\xi^2] \).

**APPENDIX D: DERIVATION OF THE REDUCED-ORDER MOMENT DYNAMICS OF PROTEIN M AND PROTEIN G**

In this section, we present the derivation of the reduced-order moment dynamics of protein M and protein G given in (44)–(48) using the reduced system (38). For ease of analysis,
yields coordinate transforms \( \bar{m} \) system (38) in terms of the variables \( \bar{m} \) and \( \bar{g} \) using the coordinate transforms \( \bar{w} = \bar{m} - \bar{c}_2 \) and \( \bar{u} = \bar{g} - \bar{c}_1 \). This yields

\[
\bar{c}_1 = \frac{\bar{g} p_{r1}}{\bar{g} + k_{d1}}, \quad (D1)
\]

\[
\bar{c}_2 = \frac{\bar{m} p_{r2}}{\bar{m} + k_{d2}}, \quad (D2)
\]

\[
\psi_{\bar{c}_1} = \frac{R_1(\bar{g})}{1 + R_1(\bar{g})} \bar{w} + \sqrt{\frac{\bar{g} R_1(\bar{g})}{1 + R_1(\bar{g})}} N_1(0, 1), \quad (D3)
\]

\[
\psi_{\bar{c}_2} = \frac{R_2(\bar{m})}{1 + R_2(\bar{m})} \bar{w} + \sqrt{\frac{\bar{m} R_2(\bar{m})}{1 + R_2(\bar{m})}} N_2(0, 1), \quad (D4)
\]

where \( R_1(\bar{g}) = \frac{p_{r1} k_{d2}}{\bar{g} (\bar{m} + k_{d2})^2} \) and \( R_2(\bar{m}) = \frac{p_{r2} k_{d1}}{\bar{m} (\bar{g} + k_{d1})^2} \) as defined in the main text.

First considering the deterministic dynamics for protein \( \bar{m} \), we have that \( \dot{\bar{m}} = \bar{w} + \bar{c}_2 \), which yields

\[
\frac{d\bar{m}}{dt} = \left( \frac{1}{1 + \frac{\bar{c}_2}{\bar{m}}} \right) \frac{d\bar{w}}{dt} = \left( \frac{1}{1 + R_2(\bar{m})} \right) \left( \beta_1 \frac{p_{r1} k_{d1}}{\bar{g} + k_{d1}} - \delta_1 \bar{m} \right),
\]

Then simplifying further and using Eq. (D1) and the reduced dynamics for \( \bar{w} \) given in (38), we obtain

\[
\frac{d\bar{w}}{dt} = \left( \frac{1}{1 + \frac{\bar{c}_2}{\bar{m}}} \right) \frac{d\bar{w}}{dt} + \frac{d\bar{c}_2}{dt} = \frac{d\bar{w}}{dt} + \frac{d\bar{c}_2}{dt}.
\]

Next, we consider the dynamics of \( \bar{w} \) using the fast variable approximations (D3) and (D4) and the fact the normal random variables \( N_1(0, 1) \) and \( N_2(0, 1) \) are independent of each other and of \( \bar{w} \) and \( \bar{u} \), we have

\[
E \left[ \bar{w} \bar{w} \right] = E \left[ (\bar{w} - \bar{w}_f)^2 \right] = \left( \frac{1}{1 + R_2(\bar{m})} \right) \left( \frac{1 + R_1(\bar{g})}{1 + R_1(\bar{g})} \right) E \left[ \bar{w} \bar{w} \right]. \quad (D5)
\]

\[
E \left[ \bar{w}_f \right] = E \left[ (\bar{w}_f - \bar{w}_f)^2 \right] = \left( \frac{1}{1 + R_1(\bar{g})} \right) \left( \frac{1}{1 + R_1(\bar{g})} \right) E \left[ \bar{w}_f \bar{w}_f \right]. \quad (D6)
\]

In a similar manner, we derive the dynamics for \( E \left[ \bar{w}_f \right] \) and \( E \left[ \bar{w}_f \right] \), which yields

\[
\frac{dE \left[ \bar{w}_f \bar{w}_f \right]}{dt} = \frac{-2\delta_1}{1 + R_2(\bar{m})} E \left[ \bar{w}^2 \right] - 2\beta_1 R_1(\bar{g}) \frac{1}{1 + R_1(\bar{g})} E \left[ \bar{w}_f \bar{w}_f \right] + \beta_1 \frac{p_{r1} k_{d1}}{\bar{g} + k_{d1}} + \delta_1 \bar{m}, \quad (D11)
\]

\[
\frac{dE \left[ \bar{w}_f \bar{w}_f \right]}{dt} = -2\delta_2 E \left[ \bar{w}^2 \right] = 2\beta_2 R_1(\bar{g}) \frac{1}{1 + R_1(\bar{g})} E \left[ \bar{w}_f \bar{w}_f \right] + \beta_2 \frac{m p_{r2}}{\bar{m} + k_{d2}} + \delta_2 \bar{g}, \quad (D12)
\]
where the deterministic variables \( \bar{w} \) and \( \bar{u} \) are represented in terms of \( \bar{\bar{m}} \) and \( \bar{\bar{g}} \) using the coordinate transforms \( \bar{w} = \bar{\bar{m}} - \bar{c}_2 \) and \( \bar{u} = \bar{\bar{g}} - \bar{c}_1 \).

Then, substituting the momentum dynamics (D11)–(D13) in Eqs. (D8)–(D10) and simplifying further using the expressions (D5)–(D7) yields the reduced-order moment dynamics of protein M and protein G given in (44)–(48).


