

Article

Solving Stochastic Reaction Networks with Maximum Entropy Lagrange Multipliers

Michail Vlysidis * and Yiannis N. Kaznessis †

Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA

* Correspondence: vlysi001@umn.edu

† Current address: General Probiotics Inc., St. Paul, MN 55114, USA; yiannis@umn.edu.

Received: 3 August 2018; Accepted: 6 September 2018; Published: 12 September 2018



Abstract: The time evolution of stochastic reaction networks can be modeled with the chemical master equation of the probability distribution. Alternatively, the numerical problem can be reformulated in terms of probability moment equations. Herein we present a new alternative method for numerically solving the time evolution of stochastic reaction networks. Based on the assumption that the entropy of the reaction network is maximum, Lagrange multipliers are introduced. The proposed method derives equations that model the time derivatives of these Lagrange multipliers. We present detailed steps to transform moment equations to Lagrange multiplier equations. In order to demonstrate the method, we present examples of non-linear stochastic reaction networks of varying degrees of complexity, including multistable and oscillatory systems. We find that the new approach is as accurate and significantly more efficient than Gillespie's original exact algorithm for systems with small number of interacting species. This work is a step towards solving stochastic reaction networks accurately and efficiently.

Keywords: stochastic chemical reactions; probability distributions; maximum entropy; lagrange multipliers

1. Introduction

The traditional approach to describe chemically reacting systems involves the use of deterministic rate laws. This macroscopic, continuous-deterministic modeling formalism is appropriate at the thermodynamic limit, when the volume of the system and the numbers of molecules of reactants and products all tend to very large values [1]. However, this approach fails to describe molecular populations that are small, finite and countable [2]. Such populations must be described by variables that can only change stochastically and by discrete amounts [3–8].

Markov chain models can be used to describe chemical reactions away from the thermodynamic limit [9]. When the system evolves stochastically, the all-encompassing chemical master equation (CME) can model the probability distribution of the system being at a particular state at time t [3,4].

The historical difficulties in solving the CME are well documented in the literature [5,9–11]. The CME is not a single equation but an infinite set of coupled equations. As a result the CME is analytically unsolvable for the majority of systems, with only few exceptions [12,13]. An established way to sample the master probability distribution is to use Gillespie's Stochastic Simulation Algorithm (SSA) [6]. Despite the accuracy of SSA, the method is a kinetic Monte Carlo algorithm and hence computationally expensive, especially for systems with large numbers of molecules, or with reaction kinetics that span multiple time scales [14,15].

A proposed alternative to solve the CME is by calculating the probability moments [3]. Moments are expected values, e.g., the mean and the variance of the number of molecules [5,16–18].

This approach is based on the concept that any probability distribution can be completely described by its moments [19]. The starting point of moment equations is to rewrite the CME in terms of moments describing the master probability distribution [18,20]. Instead of a master equation that governs the probability distribution in time, one can then write a set of ordinary differential equations evolving the moments of the probability distribution [15,21]. Efficient algorithms have been developed to generate these moment equations for arbitrary networks [10,15,17,18,21]. Yet, because the dynamics of lower-order moments depend on the higher ones for non-linear reaction networks, the system of ODEs needs to be closed or somehow truncated in order to be solved [18,22].

A common approach to tackle the moment equation closure challenge is to a priori assume a specific form of the species behavior. By assuming a molecular component's behavior one can relate the higher order moments to the lower order ones [19,23,24]. An important subset of the literature approaches the issue by assuming a specific form of the system's probability distribution, such as normal, Poisson, lognormal, Gamma etc. [10,21,24–26]. In cases when the molecular components behave in a known fashion, such distribution assumptions may produce reasonably accurate results. However, in most systems, there is a lack of knowledge of the species time evolution and thus such assumptions may break down or become impractical.

Recent advances in probability moment equation closure schemes have made it possible to efficiently calculate the stationary probability distribution [22]. These methods may quite accurately calculate the stationary behavior, however, they still face significant numerical challenges calculating the time evolution of the moments [15,19,27].

To circumvent the challenges of solving moment equations in time, we propose the use of Lagrange multipliers [22,28,29]. The connection between moments and Lagrange multipliers relies on the maximum entropy principle, which states the system attains a probability distribution that maximizes its entropy [20,22,28,30,31].

By Shannon's definition entropy is given by $S = -\sum_{\mathbf{X}} P(\mathbf{X}) \log P(\mathbf{X})$ [30], where \mathbf{X} is the number of molecules for each component and P the probability function. Using this definition and assuming that S is maximum, the system of moment equations can be transformed into a system of equations that depend only on the Lagrange multipliers [20,22]. We refer to the new system as "Lagrange multiplier equations (LMEs)". The LMEs set, unlike moment equations [3,18], is a closed system, i.e., it has the same number of unknowns with equations. Thus, an initial value problem numerical technique, like Runge-Kutta [32], can be used to solve the system in time.

Herein, we unpack in full mathematical detail the derivation of a general system of LMEs in the next section. We then apply the LMEs method to multiple examples of non-linear reaction networks, including the bistable Schlögl model [33], the multistable Wilhelm model [34], the oscillatory Brusselator system [35,36] and the viral infection model [37,38]. After presenting a numerical procedure for the implementation of the method, we present results comparing the accuracy of LMEs solutions to SSA simulations. We conclude that the significance of this new method is that LMEs can solve reaction networks as accurate as SSA with significantly less computational cost. SSA has been used widely in the past two decades to capture the stochastic nature of chemical and biological systems away from the thermodynamic limit. LMEs may offer a potentially more efficient alternative to SSA, enabling practitioners to model stochastic reacting networks.

2. Materials and Methods/Theory

2.1. Lagrange Multiplier Equations

2.1.1. Connect Moments to Lagrange Multipliers

The general form of the moment equations for any arbitrary chemical reaction network is:

$$\frac{\partial \mu}{\partial t} = A\mu + A'\mu' + \mu_c \quad (1)$$

where $\boldsymbol{\mu}$ is the lower-order moment vector (not including the zeroth-order moment, which is always equal to 1), $\boldsymbol{\mu}'$ is the higher-order moment vector, $\boldsymbol{\mu}_c$ is a vector of constants and t represents time. A and A' are constant matrices that represent the linear and non-linear components of the network, respectively. In most cases, the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ have different dimensions [20]. For non-linear reaction networks, the $\boldsymbol{\mu}'$ vector is nonempty.

For a chemical reaction network with N reactants and products, the moment μ_i is connected to the probability function P through the following expression:

$$\mu_i = \sum_{\Omega} f_{\mu_i}(\mathbf{X})P(\mathbf{X}) \quad (2)$$

where f_{μ_i} is the functional form of the i th moment μ_i . For example, for a one component system the functional form of the 3rd polynomial moment is $f_{\mu_3} = Y^3$, where Y is the system's component. The number of molecules of each component is contained in the matrix $\mathbf{X} = (X_1 \dots X_N)$ and Ω corresponds to the N -dimensional state space for all the possible values of $(X_1 \dots X_N)$. The differences between \mathbf{X} and Ω becomes clear if one compares Equation (2) and the definition of entropy ($S = -\sum_{\mathbf{X}} P(\mathbf{X})\log P(\mathbf{X})$).

One can connect the probability distribution to Lagrange multipliers $\boldsymbol{\lambda}$ by maximizing the entropy [22]:

$$P(\mathbf{X}) = \exp \left[-\sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \quad (3)$$

where λ_j is the j th Lagrange multiplier and M is the number of lower-order moments and also the size of vector $\boldsymbol{\mu}$. Thus, moments can be related to Lagrange multipliers by combing Equations (2) and (3):

$$\mu_i = \sum_{\Omega} \left\{ f_{\mu_i}(\mathbf{X}) \exp \left[-\sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \right\} \quad (4)$$

2.1.2. Time Derivatives

In Equation (3), only the Lagrange multipliers depend on time; the state space and the functional form of the moments are time-independent. Hence, the time derivative of the probability distribution is:

$$\begin{aligned} \frac{\partial P(\mathbf{X})}{\partial t} &= \frac{\partial \exp \left[-\sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right]}{\partial t} \\ &= -\exp \left[-\sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \left[\sum_{j=0}^M \frac{\partial \lambda_j}{\partial t} f_{\mu_j}(\mathbf{X}) \right] \end{aligned} \quad (5)$$

This equation and the definition of moments (Equation (2)) connect the time derivative of the moments to the time derivative of the Lagrange multipliers:

$$\begin{aligned} \frac{\partial \mu_i}{\partial t} &= \frac{\partial \sum_{\Omega} f_{\mu_i}(\mathbf{X})P(\mathbf{X})}{\partial t} \\ &= \sum_{\Omega} f_{\mu_i}(\mathbf{X}) \frac{\partial P(\mathbf{X})}{\partial t} \\ &= -\sum_{\Omega} \left\{ f_{\mu_i}(\mathbf{X}) \exp \left[-\sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \left[\sum_{j=0}^M \frac{\partial \lambda_j}{\partial t} f_{\mu_j}(\mathbf{X}) \right] \right\} \\ &= -\sum_{j=0}^M \left(\sum_{\Omega} \left\{ f_{\mu_i} f_{\mu_j}(\mathbf{X}) \exp \left[-\sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \right\} \frac{\partial \lambda_j}{\partial t} \right) \end{aligned} \quad (6)$$

Thus, the moments' time derivative is transformed to:

$$\frac{\partial \mu_i}{\partial t} = - \sum_{j=0}^M \left(\mu_{i,j} \frac{\partial \lambda_j}{\partial t} \right) \quad (7)$$

where $\mu_{i,j}$ is the combined moment of f_{μ_i} and f_{μ_j} , given by

$$\mu_{i,j} = \sum_{\Omega} \left\{ f_{\mu_i} f_{\mu_j}(\mathbf{X}) \exp \left[- \sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \right\} \quad (8)$$

The sum of the probabilities across the whole state space is always 1 by definition. As a result, the zeroth-order moment is 1 ($\mu_0 = 1$) and its time derivative is zero ($\frac{\partial \mu_0}{\partial t} = 0$). This results in the zeroth Lagrange multiplier λ_0 to be dependent on the rest of Lagrange multipliers

$$\begin{aligned} 0 &= \frac{\partial \mu_0}{\partial t} \\ &= - \sum_{j=0}^M \left(\mu_j \frac{\partial \lambda_j}{\partial t} \right) \\ &= - \mu_0 \frac{\partial \lambda_0}{\partial t} - \sum_{j=1}^M \left(\mu_j \frac{\partial \lambda_j}{\partial t} \right) \\ &= - \frac{\partial \lambda_0}{\partial t} - \sum_{j=1}^M \left(\mu_j \frac{\partial \lambda_j}{\partial t} \right) \end{aligned} \quad (9)$$

Hence, the time derivative of the zeroth Lagrange multiplier depends on the rest of the time derivatives as follows:

$$\frac{\partial \lambda_0}{\partial t} = - \sum_{j=1}^M \left(\mu_j \frac{\partial \lambda_j}{\partial t} \right) \quad (10)$$

Based on this relation, Equation (7) can be modified as:

$$\begin{aligned} \frac{\partial \mu_i}{\partial t} &= - \sum_{j=0}^M \left(\mu_{i,j} \frac{\partial \lambda_j}{\partial t} \right) \\ &= - \mu_i \frac{\partial \lambda_0}{\partial t} - \sum_{j=1}^M \left(\mu_{i,j} \frac{\partial \lambda_j}{\partial t} \right) \\ &= \mu_i \sum_{j=1}^M \left(\mu_j \frac{\partial \lambda_j}{\partial t} \right) - \sum_{j=1}^M \left(\mu_{i,j} \frac{\partial \lambda_j}{\partial t} \right) \\ &= \sum_{j=1}^M \left(-\mu_{i,j} + \mu_i \mu_j \right) \frac{\partial \lambda_j}{\partial t} \end{aligned} \quad (11)$$

We have previously proved [39] that:

$$- \mu_{i,j} + \mu_i \mu_j = \frac{\partial \mu_i}{\partial \lambda_j} \quad (12)$$

Hence, the time derivative of one moment is given by (Equations (11) and (12)):

$$\frac{\partial \mu_i}{\partial t} = \sum_{j=1}^M \frac{\partial \mu_i}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial t} \quad (13)$$

or in matrix form:

$$\frac{\partial \boldsymbol{\mu}}{\partial t} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\lambda}}{\partial t} = J \frac{\partial \boldsymbol{\lambda}}{\partial t} \tag{14}$$

J represents the Jacobian matrix of the system. Equation (14) also represents the chain rule of differentiation. There is now a way to connect the moment equations to equations that involve the Lagrange multipliers. The Lagrange multiplier equations can be derived by Equation (14) and the moment equations (Equation (1)):

$$\frac{\partial \boldsymbol{\lambda}}{\partial t} = J^{-1} \frac{\partial \boldsymbol{\mu}}{\partial t} = J(\boldsymbol{\lambda})^{-1} [A\boldsymbol{\mu}(\boldsymbol{\lambda}) + A'\boldsymbol{\mu}'(\boldsymbol{\lambda}) + \boldsymbol{\mu}_c] \tag{15}$$

where J^{-1} is the inverse of the Jacobian matrix. In the Appendix A, we discuss a computationally efficient way to calculate it. Even though the above equation also includes the moments of the probability distribution, it only depends on the Lagrange multipliers since the moments are directly correlated with Lagrange multipliers (Equation (4)).

All the Lagrange multipliers can be calculated through Equation (15) except the zeroth-order one λ_0 . As discussed earlier, λ_0 depends on the rest of the Lagrange multipliers through the relation [39]:

$$\lambda_0 = \log \left\{ \sum_{\Omega} \left[\exp \left(- \sum_{j=1}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right) \right] \right\} \tag{16}$$

We refer to Equation (15) as the ‘‘Lagrange multiplier equations (LMEs)’’. To our knowledge, this is the first time that LMEs are being discussed and derived in the literature. Through this step, the problem of calculating the probability for each point of the state space of stochastic networks has been transformed into a problem of calculating a finite set of Lagrange multipliers. The system is closed, it has the same number of equations and unknowns, and it is an initial value problem. An iterative numerical method can be used to solve this system. Appendix B outlines a proposed numerical approach to solve LMEs based on the Dormand-Prince Runge-Kutta RK5(4)7M method [32].

2.2. Initial Value from Moments to Lagrange Multipliers

The LMEs problem is an initial value problem; the values of the Lagrange multipliers for the initial state are required. However, Lagrange multipliers are numerical variables with no physical meaning and as a result an initial value is unknown. What is known is either the initial probability distribution or the values of its moments. Hence, a way to calculate Lagrange multipliers from moments is required. It should be noted that there is not a universally acceptable way to back calculate probability distributions from moments and the problem is not trivial [40].

Let’s assume that the moments, $\boldsymbol{\mu}(t = 0)$ of the initial distribution, $P(t = 0)$ are known. If only the initial distribution is known, its moments can be calculated from Equation (2). The moments are related to the Lagrange multipliers by Equation (4). Hence:

$$\begin{aligned} \boldsymbol{\mu}(t = 0) &= \begin{bmatrix} \mu_1(t = 0) \\ \mu_2(t = 0) \\ \vdots \end{bmatrix} \\ &= \begin{bmatrix} \sum_{\Omega} \left\{ f_{\mu_1(t=0)}(\mathbf{X}) \exp \left[- \sum_{j=0}^M \lambda_j(t = 0) f_{\mu_j(t=0)(t=0)}(\mathbf{X}) \right] \right\} \\ \sum_{\Omega} \left\{ f_{\mu_2(t=0)}(\mathbf{X}) \exp \left[- \sum_{j=0}^M \lambda_j(t = 0) f_{\mu_j(t=0)(t=0)}(\mathbf{X}) \right] \right\} \\ \vdots \end{bmatrix} \\ &= \mathbf{G}[\boldsymbol{\lambda}(t = 0)] \end{aligned} \tag{17}$$

G denotes the functional form of the matrix with dependences of the Lagrange multipliers. To calculate the Lagrange multipliers at the initial time from the moments, one has to solve a problem of the form: $G(\lambda) - \mu = 0$. This is an infinite set and in order to be numerically solvable the user should decide the necessary number of lower-order moments (i.e., to specify the closure order [22]). When closure order is specified, the system has the same number of equations and unknowns (the Lagrange multipliers) and a root-finding method such as Newton-Raphson can be used. The residual of the method is $R = G(\lambda) - \mu$ and the Jacobian matrix is $J_{i,j} = \frac{\partial F_i}{\partial \lambda_j} = \frac{\partial \mu_i}{\partial \lambda_j} = -\mu_{i,j} + \mu_i \mu_j$ [39]. With this approach all the Lagrange multipliers can be calculated other than the zeroth one. The zeroth one can be calculated from Equation (16).

The knowledge of moments is used here to calculate the Lagrange multipliers, which then can be used to calculate the probability distribution based on Equation (3). This approach allows to calculate probability distributions from its moments. The novelty of the method is the connection of the probability distribution with Lagrange multipliers by maximizing the entropy of the system.

3. Results

To demonstrate the proposed Lagrange multiplier equations (LMEs) approach, we employ four different example networks: the bistable Schlögl model [33], the multistable Wilhelm’s system [34], the oscillatory Brusselator [35,36] and the viral infection [37,38]. The reaction networks and kinetic constants for each model are reported in Table 1. To assess the accuracy and computational efficiency of this method, the results are compared to SSA results using the same initial condition and kinetic constants.

Table 1. The table shows the reaction networks with their kinetic constants for the bistable Schlögl, multistable Wilhelm, oscillatory Brusselator and viral infection networks. The kinetic constants values and initial condition for each system are also presented. For the third order reactions the kinetic constants are in $(\text{molecules}^2 \cdot \text{s})^{-1}$ units, for the second order in $(\text{molecules} \cdot \text{s})^{-1}$, for the first order in s^{-1} and for the zeroth order in $\text{molecules} \cdot \text{s}^{-1}$. For the Schlögl model the initial distribution is bimodal. All the other systems have a unimodal initial distribution. The first and second-order factorial moments of the initial distribution is reported for each network.

Models	Reactions	Kinetic Constants	Initial Moments	
			First-Order	Second-Order
Schlögl	$3X \xrightarrow{k_1} 2X$	$k_1 = 1.5 \times 10^{-3}$	$\{X\} = 38.01$	$\{X^2\} = 2.10 \times 10^3$
	$2X \xrightarrow{k_2} 3X$	$k_2 = 1.5 \times 10^{-1}$		
	$X \xrightarrow{k_3} \emptyset$	$k_3 = 3.5$		
	$\emptyset \xrightarrow{k_4} X$	$k_4 = 22$		
Wilhelm	$Y \xrightarrow{k_1} 2X$	$k_1 = 35$	$\{X\} = 15.94$ $\{Y\} = 7.14$	$\{X^2\} = 2.55 \times 10^2$ $\{XY\} = 1.24 \times 10^2$ $\{Y^2\} = 50.93$
	$2X \xrightarrow{k_2} X + Y$	$k_2 = 1$		
	$X + Y \xrightarrow{k_3} Y$	$k_3 = 1$		
	$X \xrightarrow{k_4} \emptyset$	$k_4 = 9.74$		
	$\emptyset \xrightarrow{k_5} X$	$k_5 = 30$		
Brusselator	$\emptyset \xrightarrow{k_1} X$	$k_1 = 10$	$\{X\} = 10.68$ $\{Y\} = 23.40$	$\{X^2\} = 1.46 \times 10^2$ $\{XY\} = 2.11 \times 10^2$ $\{Y^2\} = 5.79 \times 10^2$
	$2X + Y \xrightarrow{k_2} 3X$	$k_2 = 9 \times 10^{-3}$		
	$X \xrightarrow{k_3} Y$	$k_3 = 3$		
	$X \xrightarrow{k_4} \emptyset$	$k_4 = 1$		
Viral Infection*	$D + P \xrightarrow{k_1} \emptyset$	$k_1 = 1$	$\{D\} = 15.41$ $\{P\} = 25.53$ $\{R\} = 49.56$	$\{D^2\} = 2.37 \times 10^2$ $\{DP\} = 3.89 \times 10^2$ $\{DR\} = 7.65 \times 10^2$ $\{P^2\} = 6.42 \times 10^2$ $\{PR\} = 1.26 \times 10^3$ $\{R^2\} = 2.41 \times 10^3$
	$D \xrightarrow{k_2} R + D$	$k_2 = 3$		
	$R \xrightarrow{k_3} \emptyset$	$k_3 = 1$		
	$R \xrightarrow{k_4} D + R$	$k_4 = 10$		
	$R \xrightarrow{k_5} P + R$	$k_5 = 110$		
	$P \xrightarrow{k_6} \emptyset$	$k_6 = 200$		

For all the systems the initial condition is a distribution generated with SSA, away from the steady state. The first and second-order factorial moments [17] of each initial distribution can be found in Table 1. Using the algorithm in Section 2.2, the initial distribution was transformed into an initial set of Lagrange multipliers. Based on the initial Lagrange multipliers and the Prince-Dormant algorithm outlined in Appendix B, the Lagrange multipliers for every time point were calculated until steady state was reached. Equations (3) and (4) were then used to calculate the probability distribution and factorial moments for each time point based on the calculated Lagrange multipliers.

All the LMEs results obtained with RK5(4)M are compared with SSA results on Table 2, in order to draw conclusions about the accuracy and the time efficiency of the method. The solution for each time point was compared to the one obtained from SSA simulations with 500,000 trajectories. The average difference among all time points between the two methods can be found in Table 2. Results for both the probability distribution and the first two orders of factorial moments are presented. For the system with more than one components (i.e., Wilhelm's, Brusselator and viral infection models), the moments of the same order are averaged. The difference between LMEs results and the SSA probability distribution is calculated with the Kullback-Leiber divergence [41], whereas the second norm is used to calculate the errors in the moments.

Table 2. The table presents the results of the LMEs solved with RK5(4)M compared to the ones obtained from SSA alongside the required computational time for each reaction network of Table 1. Each SSA simulation used 500,000 trajectories. This number of trajectories is sufficient for the errors for the mean and variance to be at most $O(1)$ molecules and $O(1)$ molecules², respectively. The first three columns show the average error between the two methods, which is calculated for each time point until steady state is reached and then averaged across all of them. For the probability distribution ($P(\mathbf{X})$), the Kullback-Leibler divergence [41] is used. For the first and second-order moments the second norm is used. The last three columns present the time that is required for each method. Each system has different initial and final time as well as different step size. To take into account these differences, we report the solving time per each method's time step in CPU seconds. The last column presents the ratio of SSA time required over the RK5(4)M time.

Network	Average Error (%) of			Time per Step (CPU s)		
	$P(\mathbf{X})$	1st-Order Moments	2nd-Order Moments	LMEs	SSA	Time Ratio
Schlögl	0.05	0.22	0.22	0.02	419.2	20,960
Wilhelm	0.61	0.56	0.71	0.15	732.75	4,885
Brusselator	9.90	2.81	6.07	0.74	68.05	92
Viral Infection	0.43	0.01	0.02	8.97	64.96	7

3.1. Multistable Systems: The Schlögl and Wilhelm Models

The Schlögl model is one of the most commonly used theoretical bistable systems and is the simplest single-component system that can exhibit bistability [9,34]. Biological systems exhibiting bistability, both natural and synthetic, have gained an increasing interest recently [42–44], and the Schlögl model is a simple prototypical bistable model.

For different kinetic constants, the model has either a unimodal or a bimodal distribution. The kinetic constants selected in Table 1 result in a bimodal network. Results for the bistable model are reported on Figure 1. The probability distribution of the model is a bimodal distribution that changes form with time.

Another example of a theoretical system that can exhibit bistability is Wilhelm's network. The network is a non-linear two-component model and it was created by Wilhelm as the smallest known bistable chemical reaction system [34]. Depending on the kinetic constants, both components can exhibit bistability simultaneously which results in a multistable system overall. This is the case for the kinetic constants reported in Table 1. The probability distribution for different time points

for the two component of the network can be found in Figure 2. The system starts with a unimodal distribution for both components. As the time progresses, both components have a different bimodal distribution resulting in a multistable network. Close to the steady state, both components still have bimodal distributions but different than before.

For these two example systems, the LMEs solutions are as accurate as the SSA results (Table 2). All the probability and moments errors are lower than 1%. The accuracy of the LMEs approach does not seem to be affected by the number of stability points of the system. There is also no significant difference between the first and second-order moment errors.

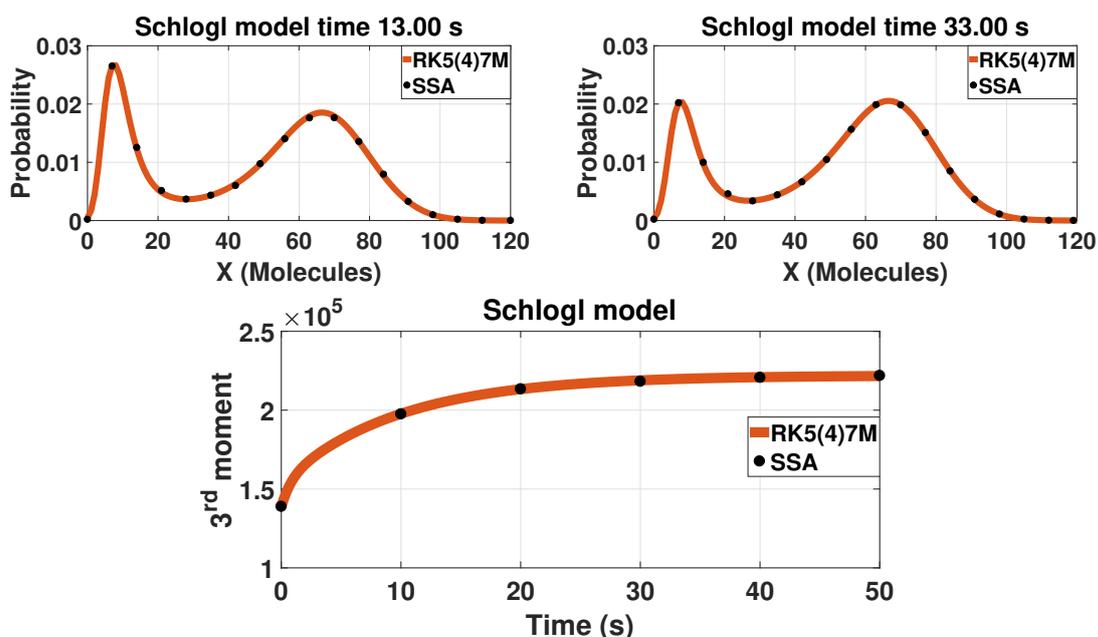


Figure 1. The probability distribution and the third-order moment of the bistable Schlögl model. The first row presents the probability distribution of the single component for two different time points (13 s after the initial condition the left and 33 s the right one). The second row presents the time evolution of the third moment. The solid lines are solutions calculated based on LMEs with tenth-order closure ($M = 10$). The closure order is chosen based on the findings of [29]. The dots represent results from 500,000 SSA trajectories.

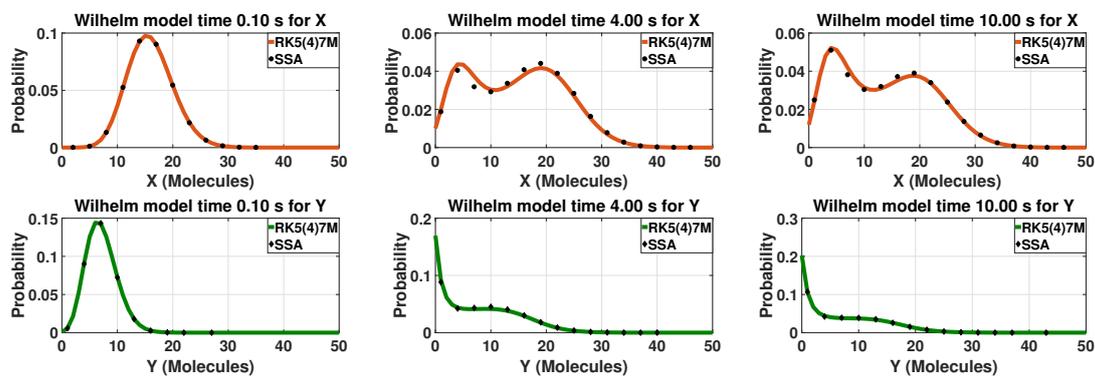


Figure 2. The probability distributions of the multistable Wilhelm’s model. The first row presents results for component X and the second row for component Y. The first column presents the time close to the initial condition (0.1 s), the second a mid-time point (4 s) and the last column a time point close to steady-state (10 s). The solid lines are solutions calculated based on LMEs and the dots/diamonds SSA solutions. The LMEs are calculated with up to sixth-order moments ($M = 27$). The closure order is chosen based on the findings of [39,45]. For SSA, 500,000 trajectories were used.

3.2. Oscillatory System: The Brusselator Model

Oscillatory behaviors are particularly challenging to capture with stochastic models [46]. We have reported earlier how methods that can capture multistable behaviors fail to appropriately capture oscillatory ones [39].

We used the LMEs to investigate the Brusselator as a simple example of a theoretical oscillatory network [36]. We study the network in its oscillatory region as shown in Figure 3. Both network's first-order moments oscillate with respect to time.

As observed, the LMEs can capture the damped oscillatory behavior of the Brusselator. However, the method exhibits limited accuracy for solving this model (Table 2). The solution by LMEs of the Brusselator network has an increased error in all categories compared to the other networks. The error in both probability and moments reaches up to 10%. In this case, unlike in the other networks, there is a significant increase in the error of the second-order moments compared to the first-order ones.

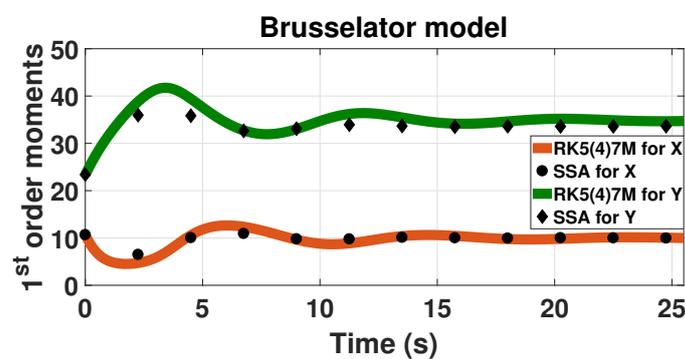


Figure 3. Time evolution of the first-order probability moments for the oscillatory Brusselator model. The solid-line upper green line presents the moment for Y component calculated based on LMEs. The solid-line lower orange line represents the moment for X. SSA results are also included from 500,000 trajectories; dots represent component X and diamonds component Y. The results are calculated with up to fourth-order moments ($M = 14$).

The difference in the accuracy can be associated with the order of closure that is used to generate the LMEs (in this case 4 or $M = 14$, Figure 3). The accuracy of the Lagrange multipliers approach depends on the number of lower-order moments, M [29]. The higher the value of M , the more accurate results the method can produce. After, a certain value of M , the improvement in accuracy is insignificant. For more information on how the lower-order moments affects probability distributions the reader is directed to the literature [29]. It has been reported that the Brusselator requires more than fourth-order moments for accurate results [46].

A low order is used for this system due to the numerical difficulties of the algorithm that generates the initial condition (Section 2.2). In the LMEs approach, the closure order is specified through the initial condition. This is the number of Lagrange multipliers with known values at the initial time. For the rest of the time points, the number of Lagrange multipliers is the same as the initial time. For this system, the initial probability distribution was created with SSA and then, based on the algorithm described in Section 2.2, it was translated into the initial Lagrange multipliers. However, the algorithm in Section 2.2 faced significant numerical issues for moment orders higher than 4 and we were not able to generate initial conditions for higher lower-order moments. Thus, we did not provide the LMEs approach with the necessary number of lower-order moments that produce accurate results. The high error was generated by the algorithm that calculates the initial condition and not the Runge-Kutta algorithm of the LMEs. The Brusselator results are not ideal; however, given the difficulties of current methods to solve oscillatory systems, this is a step in the right direction.

3.3. Multicomponent System: The Viral Infection Model

The LMEs approach is only limited to theoretical systems with complex behavior but can also be applied to natural and synthetic biological networks. One example is the viral infection model. The model was first created by Haseltine [37] as a general model of a cell infection by a virus. The model was then revised and simplified by Goutsias [38]. The network in Table 1 reflects the revised version of Goutsias. Results for the mixed first-order moments (moments that depend on more than one component) are reported in Figure 4. The LMEs approach accurately captures the different dynamics of all components and their combinations throughout the entire simulated time (Table 2).

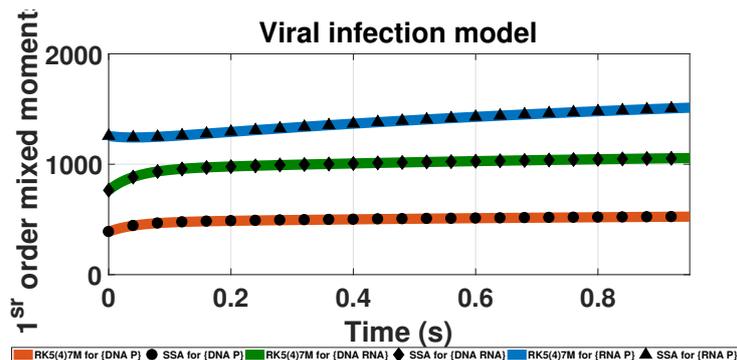


Figure 4. Time evolution of the first-order mixed probability moments for the Viral infection model. The solid-lines present results calculated based on LMEs and the dotted points present SSA results. The results are calculated with up to second-order moment closure ($M = 9$) and 500,000 SSA trajectories. The upper blue line with triangles present the $\{DNA\ Protein\}$ mixed moment, the mid green line with diamonds corresponds to the $\{DNA\ RNA\}$ mixed moment and the lower orange line with dots the $\{RNA\ Protein\}$ mixed moment.

3.4. Computational Cost of LMEs

Aside from the accuracy of the approach, Table 2 also includes the computational time that is required for RK5(4)M to solve the LMEs. In all the systems, the LMEs approach is significantly faster than SSA. Each model has different computational time steps that are used; in order to take this difference into account, the table shows the time required for one time point per system. For each system, at least 10,000 time steps were used. As observed, the total amount of time that is required for LMEs approach is significantly lower than for SSA.

Among all systems, the LMEs approach requires less computational time to solve the two multistable systems (the Schlögl and the Wilhelm's models). The LMEs approach is significantly faster than SSA (at least 4000 times faster, Table 2) and advantageous especially for multistable systems. As for oscillatory dynamics, they do not seem to affect the computational time of the LMEs approach.

The number of components, on the other hand, affects significantly the computational time of the LMEs approach. The Wilhelm's and Brusselator models (both two-component systems) are slower than the Schlögl model (one-component system) and faster than the viral infection model (three-component system).

3.5. Dependence on State Space Size

The proposed method is a numerical method to solve stochastic reaction networks. The LMEs include summations across all the possible values of the state space. The state space needs to be specified and be finite in order for the algorithm to perform numerical calculations. This is a known case with numerical methods [22,47] and it can raise concerns about the accuracy of a method, especially in the case of systems with unknown state space. Some ways to calculate the appropriate state space of stochastic networks have been discussed in the literature [11,47].

To evaluate the stability of the Lagrange multipliers approach with the state space size, we use the Schögl model as an example. The model has one component and thus the connection between state space size and solution accuracy can be drawn easily; yet it has a complex enough bistable behavior. Figure 5 shows results of the LMEs approach for different state space sizes. The solutions are also compared with the SSA results. The accuracy of the method is not affected by the state space size in this example. All different state space lengths provide the same value for the moments of the probability.

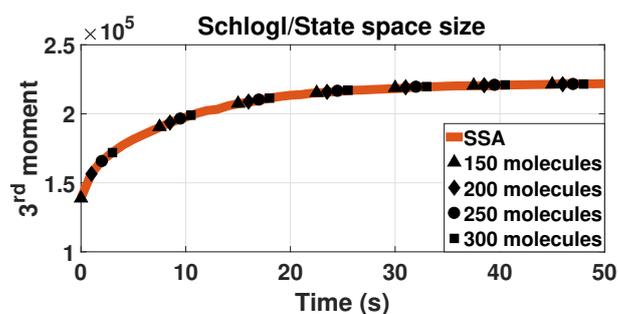


Figure 5. Time evolution of the third-order moment for the Schögl model. The plot is similar to the bottom plot of Figure 1. Here, four different state space lengths are plotted. All the cases have zero as the lower bound of the state space. The coloring in this picture is reverse than the rest of them. The solid orange line is the SSA solution for 500,000 trajectories. The black dots are results with tenth-closure LMEs ($M = 10$). The triangles represent the solution with 150 molecules as the upper bound of the state space, diamonds use 200 molecules limit, circles 250 molecules and the squares 300 molecules.

4. Discussion

Stochasticity governs reaction networks away from the thermodynamic limit. A common approach to solve stochastic chemically reacting systems is through moment equations. However, moment equation systems are not closed for non-linear chemical reactions, and require important assumptions about the form of the probability function.

Here, we present an alternative method to solve stochastic reaction networks by using Lagrange multipliers. An approach to transform moment equations into a closed system of Lagrange multiplier equations (LMEs) is described. Since, LMEs is a closed system, Runge-Kutta methods (e.g., Durmand-Prince RK5(4)7M) can be employed to solve the transient problem and calculate the Lagrange multipliers for different time points. With the knowledge of the Lagrange multipliers, one then can calculate the probability distribution and its moments at any given time.

We show that this is a novel approach to bypass numerical challenges with chemical master equations and moment equations. The only assumption used is that the entropy of the system is maximum at all times. The LMEs solutions are as accurate as alternative methods, such as SSA, with significant computational advantages. Four different non-linear reaction networks are employed to demonstrate the advantages of the new LMEs approach. The networks vary in numbers of components (from one to three) and complexity in behavior (including multistability and oscillations). The number of stable solutions of the network does not affect the accuracy of the LMEs approach. On the other hand, the number of the network's components can affect the method's computational requirements. Future studies may be required to assess how the method scales with the size of the reaction network.

For all of the examples studied, the LMEs approach is faster than SSA. Notably, the method is computationally efficient for multistable systems, which are of interest to a large community [42–44]. The method also appears unaffected by the size of the system's state space; the method is stable and accurate for multiple state space sizes. On the other hand, oscillatory behaviors can reduce the method's accuracy.

Author Contributions: M.V. and Y.N.K. designed the study and wrote the manuscript. M.V. conducted the simulations. Conceptualization, M.V. and Y.N.K.; Formal analysis, M.V.; Funding acquisition, Y.N.K.; Investigation, M.V.; Methodology, M.V.; Project administration, Y.N.K.; Resources, Y.N.K.; Software, M.V.; Validation, M.V.; Visualization, M.V.; Writing—original draft, M.V. and Y.N.K.

Funding: This work was supported by the National Institutes of Health [Grant No. GM111358]; National Science Foundation [Grant No. CBET-1412283]; Extreme Science and Engineering Discovery Environment (XSEDE) [National Science Foundation Grant No. ACI-10535753]; Minnesota Supercomputing Institute (MSI); and the University of Minnesota Digital Technology Center.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Inverse of Jacobian

The Jacobian matrix J of the moment equations for a stochastic reaction network is calculated as:

$$J_{i,j} = \frac{\partial \mu_i}{\partial \lambda_j} = -\mu_{i,j} + \mu_i \mu_j \quad (\text{A1})$$

By the definition (Equation (8)), it is true that:

$$\mu_{i,j} = \sum_{\Omega} \left\{ f_{\mu_i} f_{\mu_j}(\mathbf{X}) \exp \left[- \sum_{j=0}^M \lambda_j f_{\mu_j}(\mathbf{X}) \right] \right\} \quad (\text{A2})$$

Thus, $\mu_{i,j} = \mu_{j,i}$. It is then apparent that the Jacobian matrix is symmetric since:

$$J_{i,j} = -\mu_{i,j} + \mu_i \mu_j = -\mu_{j,i} + \mu_j \mu_i = J_{j,i} \quad (\text{A3})$$

As a symmetric matrix, J can be calculated as [48]:

$$J = Q^T \Lambda Q \quad (\text{A4})$$

where Λ is a diagonal matrix with the diagonal elements being the eigenvalues of J and Q the associated eigenvector matrix. Q is also real orthogonal [48], i.e., $Q^{-1} = Q^T$. The inverse of the Jacobian matrix, J^{-1} can be calculated as:

$$J^{-1} = (Q^T \Lambda Q)^{-1} = (Q)^{-1} (\Lambda)^{-1} (Q^T)^{-1} = Q^T (\Lambda)^{-1} Q \quad (\text{A5})$$

Since Λ is diagonal, $(\Lambda)^{-1}$ is also diagonal with elements the reciprocal elements of Λ ($(\Lambda_{i,j})^{-1} = \frac{1}{\Lambda_{i,j}}$). Additionally, each element of Λ , $\Lambda_{i,j}$, represents each eigenvalue of the Jacobian J .

Based on Equation (A5), in order to calculate the inverse of the Jacobian matrix, one only needs to solve the eigenproblem of the Jacobian.

Appendix B. Runge-Kutta

As mentioned, the LMEs problem is an initial value problem. A proposed way to solve the system is the Dormand-Prince RKF5(4)7M Runge-Kutta method [32]. RKF5(4)7M is an explicit numerical method with adaptive step size. Dormand and Prince formulated the method in order to create a stable approach for local extrapolations; this is the reason it is chosen in this work as it will help to create more stable solution for time calculations by extrapolating the knowledge of past time solutions. A table that includes all the constants of the RKF5(4)7M method (often called "Butcher tableau" [49] or simply "tableau") can be found in the Appendix B.1.

As derived in Equation (15), the LMEs are:

$$\frac{\partial \lambda}{\partial t} = J(\lambda)^{-1} [A\mu(\lambda) + A'\mu'(\lambda) + \mu_c] = F(\lambda) \quad (\text{A6})$$

where F denotes the functional form of the LMEs.

The calculation of Lagrange multipliers, λ , is desired for multiple time points. To calculate the multipliers at the time point $n + 1$, the knowledge of the multipliers at time n is required. The Lagrange multipliers at time $n + 1$, λ_{n+1} , are calculated as:

$$\lambda_{n+1} = \lambda_n + \sum_{i=1}^7 \hat{b}_i k_i \tag{A7}$$

where λ_{n+1} are the Lagrange multipliers at time n , \hat{b}_i are constants of RKF5(4)7M and k_i are the Runge-Kutta coefficients:

$$k_i = hF \left(\lambda_n + \sum_{j=1}^{i-1} a_{ij} k_j \right) \tag{A8}$$

h is the adaptive step size and a_{ij} are RKF5(4)7M constants.

In RKF5(4)7M, a lower-order solution, λ^* , is used to evaluate the error of the solution λ and change accordingly the adaptive step size h . The lower-order solution at time $n + 1$, λ_{n+1}^* , are calculated as:

$$\lambda_{n+1}^* = \lambda_n^* + \sum_{i=1}^7 b_i k_i \tag{A9}$$

where b_i 's are another set of RKF5(4)7M constants. All the constants (a_{ij}, \hat{b}_i & b_i) can be found in the Appendix B.1.

More details about the algorithm of RKF5(4)7M used in this work can be found in the Appendix B.2.

Appendix B.1. Runge-Kutta Tableau

In this section we list the Runge-Kutta RK5(4)7M tableau. The method was first described by Dormand and Prince [32] and we just list it here for the convenience of the reader. The Runge-Kutta tableau of the method is:

i	c_i	$j=1$	2	3	4	5	6	\hat{b}_i	b_i
1	0	a_{ij}						35	5179
2	$\frac{1}{5}$	$\frac{1}{5}$						384	57600
3	$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$					0	0
4	$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$				500	7571
5	$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$			1113	16695
6	1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$		125	393
7	1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	192	640
								$-\frac{2187}{6784}$	$-\frac{92097}{339200}$
								11	187
								84	2100
								0	$\frac{1}{40}$

Appendix B.2. Runge-Kutta Algorithm

The proposed algorithm to solve LMEs based on the Dormand-Prince RKF5(4)M method is:

1. Calculate vector μ_c and matrices A and A' as in [17]. These matrices depend only on the kinetic constants of the system and they are constant for all time points.
2. Introduce the time step size h . The time point of this iteration is $t_{n+1} = t_n + h$, where t_n is the time point of the previous iteration.
3. Increase the iteration counter by 1. For the rest of the algorithm the iteration step will be denoted by $n + 1$. At the first iteration, $n = 0$.
4. Introduce the previous step (iteration n) Lagrange multipliers λ_n . At the first iteration, introduce the initial condition λ_0 . The subscripts here refers to the time step and are applied on the whole Lagrange multiplier vector.

5. Calculate the k_i 's coefficients from equation $k_i = hF\left(\lambda_n + \sum_{j=1}^{i-1} a_{ij}k_j\right)$ and the tableau (Appendix B.1). Each coefficient k_i depends on the functional form F of the Lagrange multiplier equations ($F(\lambda) = J(\lambda)^{-1} [A\mu(\lambda) + A'\mu'(\lambda) + \mu_c]$) and on different set of Lagrange multipliers ($\lambda_n + \sum_{j=1}^{i-1} a_{ij}k_j$). To account for the change of the Lagrange multipliers, there is a need for a subroutine to calculate k_i .
6. This is a proposed subroutine for k_i
 - (a) Calculate the augmented Lagrange multipliers for the specific k_i , $\lambda_{n,k_i} = \lambda_n + \sum_{j=1}^{i-1} a_{ij}k_j$. The values of a_{ij} can be found in the tableau in Appendix B.1.
 - (b) Calculate the lower ($\mu(\lambda_{n,k_i})$) and higher-order moments ($\mu'(\lambda_{n,k_i})$) for the augmented Lagrange multipliers based on equation $\mu_i = \sum_{\Omega} f_{\mu_i}(\mathbf{X}) \exp\left(-1 - \sum_j \lambda_j f_{\mu_j}(\mathbf{X})\right)$.
 - (c) Calculate the inverse of the Jacobian, $J(\lambda_{n,k_i})^{-1}$ for the augmented multipliers as described in Appendix A.
 - (d) Calculate the functional form $F(\lambda_{n,k_i})$ of the Lagrange multiplier equations ($F(\lambda) = J(\lambda)^{-1} [A\mu(\lambda) + A'\mu'(\lambda) + \mu_c]$).
 - (e) Calculate k_i based on equations $k_i = hF(\lambda_{n,k_i})$.
7. Repeat step 6 for all seven k_i coefficients, $i = 1, \dots, 7$. Its coefficient k_i depends on all the previous coefficients, thus the calculation of each one needs to be in sequence and include the subroutine. For example, the calculations of step 6 should first performed for $i = 1$, then based on k_1 to be performed for $i = 2$, then based on k_2 to be performed for $i = 3$ etc.
8. Calculate the Lagrange multipliers for the current iteration $n + 1$, λ_{n+1} , based on equation $\lambda_{n+1} = \lambda_n + \sum_{i=1}^7 \hat{b}_i k_i$ and the tableau (Appendix B.1).
9. Calculate the lower-order solution for the current iteration $n + 1$, λ_{n+1}^* , based on equation $\lambda_{n+1}^* = \lambda_n + \sum_{i=1}^7 b_i k_i$ and the tableau (Appendix B.1).
10. Check the accuracy of the solution by calculating the error ϵ :

$$\epsilon = \left| \frac{\lambda_{n+1} - \lambda_{n+1}^*}{\lambda_{n+1}} \right|_2 \quad (\text{A10})$$

11. If the error ϵ is smaller than the wanted tolerance accept the solution and proceed to the next step. Otherwise, set a new time step size $h = h \cdot s$ and go back to step 5 and re-perform the calculation with the new step size. s is used to adjust the step size based on the accuracy of the solution and is given by:

$$s = \left(\frac{\text{tolerance} \cdot h}{2 \cdot |\lambda_{n+1}^* - \lambda_{n+1}|_2} \right)^{\frac{1}{4}} \quad (\text{A11})$$

12. Accept the solution of Lagrange multipliers for the current iteration ($n + 1$) and current time point (t_{n+1}) as calculated at step 8.
13. Set the new step size to $h = h \cdot s$ and proceed to step 3 to calculate the solution for the rest of the time points until the final is reached.
14. For any given time t , the probability distribution and its moments can be calculated based on the Lagrange multipliers ($\lambda(t)$) and equations $P(\mathbf{X}) = \exp\left[-1 - \sum_j \lambda_j f_{\mu_j}(\mathbf{X})\right]$ & $\mu_i = \sum_{\Omega} f_{\mu_i}(\mathbf{X}) \exp\left(-1 - \sum_j \lambda_j f_{\mu_j}(\mathbf{X})\right)$.

References

1. Folger, H. *Elements of Chemical Reaction Engineering*, 4th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2016.
2. Schnoerr, D.; Sanguinetti, G.; Grima, R. Approximation and inference methods for stochastic biochemical kinetics-A tutorial review. *J. Phys. A* **2017**, *50*, 093001. [[CrossRef](#)]
3. McQuarrie, D.A. Stochastic approach to chemical kinetics. *J. Appl. Probab.* **1967**, *4*, 413–478. [[CrossRef](#)]
4. Oppenheim, I.; Shuler, K.E. Master Equations and Markov Processes. *Phys. Rev.* **1965**, *138*, B1007. [[CrossRef](#)]

5. Kampen, N.V. *Stochastic Processes in Physics and Chemistry*, 5th ed.; Elsevier: Amsterdam, The Netherlands, 2004; ISBN 978-0-444-89349-0.
6. Gillespie, D.T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **1976**, *22*, 403–434. [[CrossRef](#)]
7. Gibson, M.A.; Bruck, J. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J. Phys. Chem. A* **2000**, *104*, 1876–1889. [[CrossRef](#)]
8. Cao, Y.; Li, H.; Petzold, L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys.* **2004**, *121*, 4059–4067. [[CrossRef](#)] [[PubMed](#)]
9. Gillespie, D.T. *Markov Processes, An Introduction for Physical Scientists*; Academic Press: San Diego, CA, USA, 1992.
10. Lakatos, E.; Ale, A.; Kirk, P.D.W.; Stumpf, M.P.H. Multivariate moment closure techniques for stochastic kinetic models. *J. Chem. Phys.* **2015**, *143*, 094107. [[CrossRef](#)] [[PubMed](#)]
11. Munsky, B.; Khammash, M. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* **2006**, *124*, 044104. [[CrossRef](#)] [[PubMed](#)]
12. Jahnke, T.; Huisinga, W. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Bio.* **2006**, *54*, 1–26. [[CrossRef](#)] [[PubMed](#)]
13. Grima, R.; Schmidt, D.R.; Newman, T.J. Steady-state fluctuations of a genetic feedback loop: An exact solution. *J. Chem. Phys.* **2012**, *137*, 035104. [[CrossRef](#)] [[PubMed](#)]
14. Sotiropoulos, V.; Kaznessis, Y.N. An adaptive time step scheme for a system of stochastic differential equations with multiple multiplicative noise: Chemical Langevin equation, a proof of concept. *J. Chem. Phys.* **2008**, *128*, 014103. [[CrossRef](#)] [[PubMed](#)]
15. Ale, A.; Kirk, P.; Stumpf, M.P.H. A general moment expansion method for stochastic kinetic models. *J. Chem. Phys.* **2013**, *138*, 174101. [[CrossRef](#)] [[PubMed](#)]
16. Sotiropoulos, V.; Kaznessis, Y.N. Analytical Derivation of Moment Equations in Stochastic Chemical Kinetics. *Chem. Eng. Sci.* **2011**, *66*, 268–277. [[CrossRef](#)] [[PubMed](#)]
17. Smadbeck, P.; Kaznessis, Y.N. Efficient Moment Matrix Generation for Arbitrary Chemical Networks. *Chem. Eng. Sci.* **2012**, *84*, 612–618. [[CrossRef](#)] [[PubMed](#)]
18. Gillespie, C. Moment-closure approximations for mass-action models. *IET Syst. Biol.* **2009**, *3*, 52–58. [[CrossRef](#)] [[PubMed](#)]
19. Grima, R. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **2012**, *136*, 154105. [[CrossRef](#)] [[PubMed](#)]
20. Constantino, P.H.; Vlysidis, M.; Smadbeck, P.; Kaznessis, Y.N. Modeling stochasticity in biochemical reaction networks. *J. Phys. D Appl. Phys.* **2016**, *49*, 093001. [[CrossRef](#)]
21. Milner, P.; Gillespie, C.S.; Wilkinson, D.J. Moment closure based parameter inference of stochastic kinetic models. *Stat. Comput.* **2013**, *23*, 287–295. [[CrossRef](#)]
22. Smadbeck, P.; Kaznessis, Y.N. A closure scheme for chemical master equations. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14261–14265. [[CrossRef](#)] [[PubMed](#)]
23. Zechner, C.; Ruess, J.; Krenn, P.; Pelet, S.; Peter, M.; Lygeros, J.; Koepl, H. Moment-based inference predicts bimodality in transient gene expression. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8340–8345. [[CrossRef](#)] [[PubMed](#)]
24. Singh, A.; Hespanha, J. Moment closure techniques for stochastic models in population biology. In *Proceeding of the 2006 American Control Conference*, Minneapolis, MN, USA, 14–16 June 2006.
25. Krishnarajah, I.; Cook, A.; Marion, G.; Gibson, G. Novel moment closure approximations in stochastic epidemics. *Bull. Math. Biol.* **2005**, *67*, 855–873. [[CrossRef](#)] [[PubMed](#)]
26. Schnoerr, D.; Sanguinetti, G.; Grima, R. Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **2015**. [[CrossRef](#)] [[PubMed](#)]
27. Goutsias, J.; Jenkinson, G. Markovian dynamics on complex reaction networks. *Phys. Rep.* **2013**, *2*, 199–264. [[CrossRef](#)]
28. Kapur, J. *Maximum-Entropy Models in Science and Engineering*; Wiley Eastern Ltd.: Darya Ganj, New Delhi, 1989.

29. Vlysidis, M.; Constantino, P.H.; Kaznessis, Y.N. ZI-Closure Scheme: A Method to Solve and Study Stochastic Reaction Networks. In *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*; Springer International Publishing: Cham, Switzerland, 2017; pp. 159–174, ISBN: 978-3-319-62627-7.
30. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
31. Sutter, T.; Sutter, D.; Esfahani, P.M.; Lygeros, J. Generalized Maximum Entropy Estimation. *arXiv* **2004**, arXiv:1708.07311. Available online: <https://arxiv.org/abs/1708.07311> (accessed on 6 September 2018).
32. Dormand, J.; Prince, P. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.* **1980**, *6*, 19–26. [[CrossRef](#)]
33. Schlogl, F. On Thermodynamics Near a Steady State. *Z. Physik* **1971**, *458*, 446–458. [[CrossRef](#)]
34. Wilhelm, T. The smallest chemical reaction system with bistability. *BMC Syst. Biol.* **2009**, *3*, 90. [[CrossRef](#)] [[PubMed](#)]
35. Lefever, R.; Nicolis, G. Chemical instabilities and sustained oscillations. *J. Theor. Biol.* **1971**, *30*, 267–284. [[CrossRef](#)]
36. Lavenda, B.; Nicolis, G.; Herschkowitz-Kaufman, M. Chemical instabilities and relaxation oscillations. *J. Theor. Biol.* **1971**, *32*, 283–292. [[CrossRef](#)]
37. Haseltine, E.L.; Rawlings, J.B. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* **2002**, *117*, 6959–6969. [[CrossRef](#)]
38. Goutsias, J. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.* **2005**, *122*, 184102. [[CrossRef](#)] [[PubMed](#)]
39. Vlysidis, M.; Kaznessis, Y.N. A linearization method for probability moment equations. *Comput. Chem. Eng.* **2018**, *112*, 1–5. [[CrossRef](#)]
40. Munkhammar, J.; Mattsson, L.; Rydén, J. Polynomial probability distribution estimation using the method of moments. *PLoS ONE* **2017**, *12*, e0174573. [[CrossRef](#)] [[PubMed](#)]
41. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
42. Gardner, T.S.; Cantor, C.R.; Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **2000**, *403*, 339–342. [[CrossRef](#)] [[PubMed](#)]
43. Angeli, D.; Ferrell, J.E.; Sontag, E.D.; Sontag, E.D. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1822–1827. [[CrossRef](#)] [[PubMed](#)]
44. Artyomov, M.N.; Das, J.; Kardar, M.; Chakraborty, A.K. Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18958–18963. [[CrossRef](#)] [[PubMed](#)]
45. Vlysidis, M.; Schiek, A.C.; Kaznessis, Y.N. ZICS: An Application for Calculating the Stationary Probability Distribution of Stochastic Reaction Networks. *arXiv* **2018**, arXiv:1806.06428. Available online: <https://arxiv.org/abs/1806.06428> (accessed on 6 September 2018).
46. Constantino, P.H.; Kaznessis, Y.N. Maximum entropy prediction of non-equilibrium stationary distributions for stochastic reaction networks with oscillatory dynamics. *Chem. Eng. Sci.* **2017**, *171*, 139–148. [[CrossRef](#)]
47. Peleš, S.; Munsky, B.; Khammash, M. Reduction and solution of the chemical master equation using time scale separation and finite state projection. *J. Chem. Phys.* **2006**, *125*, 204104. [[CrossRef](#)] [[PubMed](#)]
48. Golub, G.H.; Van Loan, C.F. *Matrix Computations*, 4th ed.; The Johns Hopkins University Press: Baltimore, MD, USA, 2013; ISBN 978-1-4214-0794-4.
49. Butcher, J.C. A stability property of implicit Runge-Kutta methods. *BIT* **1975**, *15*, 358–361. [[CrossRef](#)]

