Stochastic Hybrid System for Chemical Master Equation

Luonan Chen[1,2,3,4]   Ruiqi Wang[4]   Kazuyuki Aihara[2,3]

[1]Department of Electrical Engineering and Electronics, Osaka Sangyo University, Osaka, Japan
[2]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
[3]ERATO Aihara Complexity Modelling Project, JST, Tokyo, Japan
[4]Institute of Systems Biology, Shanghai University, Shanghai, China

1   Introduction

Cellular processes at a molecular level are unavoidably stochastic. In particular, gene regulation is an inherently noisy process, from transcriptional control, alternative splicing, translation, diffusion to chemical modification reactions of transcriptional factors, which all involve stochastic fluctuations. Such stochastic noise may not only affect the dynamics of biological systems but also be exploited by living organisms to actively facilitate certain functions. From the evolution viewpoint, noises are also assumed to be used for the cellular and population variability control. Due to very low copy numbers for many species in living cells, the origin of stochasticity can be traced to random transitions among discrete chemical states, which implies that a model of gene regulatory networks should be able to present such discrete nature of small numbers both qualitatively and quantitatively [1].

Explicitly considering all variables and chemical reactions in a cell is unrealistic for a gene regulatory network from modeling, analyzing, and computing viewpoint. However, many different time scales in a cell characterize the gene regulatory processes, which can be exploited to reduce the complexity of the mathematical models. For instance, the transcription and translation processes generally evolve on a time scale that is much slower than that of phosphorylation, dimerization or binding reactions of transcription factors. Moreover, in biological systems, a large class of biological models can be approximately by stochastic hybrid systems in which some state components are discrete and other are continuous. Continuous state components are usually involved in fast reactions with high copy numbers of molecules, whereas discrete state components are in slow processes and have low copy numbers of molecules. In this work, based on the partial Kramers-Moyal expansion [2] with the central limit theorem, we exploit such properties to simplify a complicated molecular network to a hybrid system by giving several models, which can be applied to the quantitative simulation of a large cellular system. In other words, we aim to propose a novel stochastic hybrid model for representing chemical master equations, and provide several computational algorithms to efficiently simulate the stochastically cellular dynamics.

Notice that there are two different implications for hybrid systems, i.e., one is a hybrid system with both discrete and continuous variables, and the other is a hybrid system with both stochastic and deterministic dynamics, or stochastic hybrid processes. Hybrid simplification approximates partial or all discrete variables by continuous values, and therefore can drastically reduces computational time by eliminating the simulated discrete events from computational viewpoint because the computational complexity depends on the number of the discrete jumps. Besides the hybrid simplification based on central limit theorem, other simplification schemes, such as averaging approximation, state-aggregation scheme, and stochastic quasi-steady-state approximation, can also be adopted to simplify the complexity of the complicated molecular networks.

2　Master Equation and Molecular Network

2.1　Reaction-Diffusion Master Equation

The chemical master equation models discrete biochemical reactions within biological cells, and represents dynamic but primarily non-spatial, treating the cell as a well-mixed volume or compartment [3]. However, biological cells contain incredibly complex spatial environments, comprised of numerous organelles, irregular membrane structures, fibrous actin networks, long directed microtubule bundles, and many other geometrically complex structures. To model such discrete but spatial system, the reaction-diffusion master equation is able to discretize space, approximating the diffusion of molecules as a continuous time random walk on a lattice, with bimolecular reactions occurring with a fixed probability per unit time for molecules within the same voxel (i.e. volumetric and pixel in a three dimensional space). In particular, exact realizations of the reaction-diffusion master equation can be created by using the Gillespie algorithm [3].

Let Ω denote a closed volume in $\mathscr{R}^3$ for the system. In the reaction-diffusion master equation model, Ω is divided by a mesh into a collection voxels labeled by vectors i in an index set $I = \{1, ..., L\}$, i.e., $i \in I$. Consider a system containing m chemical reactions with n molecular species of z. Let $z = (z^1, \cdots, z^L) \in \mathscr{R}^{n \times L}$, where z^i is the concentrations of molecules at voxel-i or compartment i. Let $1_i^l \in \mathscr{R}^{n \times L}$ be the state, where the number of all chemical species at all locations is zero, except for the lth chemical species at the ith location, which is one. Let k_{ij}^l denote the diffusive jump rate for each individual molecule of the lth chemical species into voxel i from voxel j, for $i \neq j$. Since diffusion is treated as a first order reaction and molecules are assumed to diffuse independently, the total probability per unit time at time t for one molecule of species l to jump from voxel j to voxel i is $k_{ij}^l z_j^l(t)$, where $k_{ii}^l = 0$.

Define $r_k = (r_k^1, ..., r_k^L) \in \mathscr{R}^{n \times L}$, where r_k^i is a vector for the change of the state, i.e., $r_{k,j}^i$ is the change in the number of the jth molecule by the kth reaction in the ith voxel. $w_k^i(z^i)$ is the transition rate (≥ 0) from state z^i to state $z^i + r_k$ by the kth reaction occurring in the ith voxel. The corresponding state change in z due to an occurrence of the kh reaction in the ith voxel will be denoted by $r_k 1_i$, i.e. $z(t) \rightarrow z(t) + r_k 1_i$. Therefore, the reaction-diffusion master equation for the time evolution of $P(z;t)$ with initial state $z(0)$ at $t = 0$ is represented as

$$\frac{\partial P(z;t)}{\partial t} = \sum_{i \in I}\sum_{j \in I}\sum_{l=1}^{L}[k_{ij}^l(z_j^l+1)P(z+1_j^l-1_i^l;t)-k_{ij}^l z_j^l P(z;t)],$$

$$-\sum_{i \in I}\sum_{k=1}^{m}[w_k^i(z_i-r_k)P(z-r_k 1_i;t)-w_k^i(z_i)P(z;t)].$$

Clearly, diffusion is then modelled as a chemical reaction or jumps between the neighbouring compartments. Hence, the Gillespie algorithm can be directly adopted to simulate the reaction-diffusion master equation numerically.

2.2 Chemical Master Equation

Next we ignore the diffusion effect by focusing the dynamic features of the chemical reaction network to simplify the model. Consider a system containing m chemical reactions with n molecular species, where m and n are positive integer numbers. Define $(x,y) \in \mathscr{R}^n$ and (X,Y) to be concentrations and numbers of molecules, respectively. $n_x + n_y = n$, where x_i or y_i is the concentration of a molecule, i.e., the number divided by the system size or volume V. Then the dynamics of the system is described by the master equation [2] with initial state (x_0,y_0) at $t=0$

$$\frac{\partial P(x,y;t)}{\partial t} = \sum_{k=1}^{m}[w_k(x-\phi_k/V,y-\theta_k/V)P(x-\phi_k/V,y-\theta_k/V;t)$$

$$-w_k(x,y)P(x,y;t)], \tag{1}$$

where $(\phi_k,\theta_k) \in \mathscr{R}^n$ is a vector for the change of the state. $w_k(x,y)$ is the transition rate (≥ 0) from state (x,y) to state $(x+\phi_k/V,y+\theta_k/V)$ by the kth reaction. Equation (1) theoretically provides full information of system performances, but only a few simple cases are amenable to exact solution due to its complexity for a large number of variables. Next we exploit the properties of molecular networks to simplify the master equation.

3 Stochastic Hybrid Systems

3.1 Hybrid System with Deterministic Process

Assuming that the number of X is much bigger than that of Y, we can approximate X by continuous values x, i.e., $x=X/V$ by keeping Y as discrete variables, i.e., $y=Y/V$. Therefore, by partial Kramers-Moyal expansion of (1) with respect to x and ϕ_k/V up to the first order (i.e., zero-th order and first order), we have the following hybrid representation

$$\frac{\partial P(x,y;t)}{\partial t} = \sum_{k=1}^{m}[w_k(x,y-\theta_k/V)P(x,y-\theta_k/V;t)-w_k(x,y)P(x,y;t)]$$

$$-\sum_{j=1}^{n_x}\frac{\partial}{\partial x_j}\left[\left(\sum_{k=1}^{m}\phi_{kj}W_k(x,y)\right)P(x,y;t)\right]+\mathscr{O}(\frac{1}{V}), \tag{2}$$

where $w_k = W_k V$ is rates of the reactions which are proportional to the volume V based on mass action law. $\mathscr{O}(\frac{1}{V})$ implies that the order of the term is higher than $1/V$. Clearly,

the term of the Taylor zero-th order expansion, i.e., the term $\sum_{k=1}^{m}[\cdot]$ in (2), is the discrete dynamics or the master equation for discrete variables y, and the term of the Taylor first order expansion, i.e., the term $\sum_{j=1}^{n_x}\frac{\partial}{\partial x_j}[\cdot]$ in (2), is the deterministic kinetic dynamics or corresponds to the Langevin equation for the continuous variables x. The term of the Taylor second order expansion, i.e., $\mathscr{O}(\frac{1}{V})$, corresponds to the diffusion process and when $V \to \infty$, it approaches zero. Therefore, (2) is a hybrid system with both discrete and continuous dynamics, or with both stochastic and deterministic processes. Specifically, the stochastic system for discrete variables Y is

$$
\begin{aligned}
\frac{\partial P(x,y;t)}{\partial t} &= \sum_{k=1}^{m} [w_k(x,y-\theta_k/V)P(x,y-\theta_k/V;t) \\
&\quad -w_k(x,y)P(x,y;t)]
\end{aligned}
\tag{3}
$$

and the deterministic system for continuous variables x is for $j = 1,\cdots,n_x$

$$
\frac{dx_j(t)}{dt} = \sum_{k=1}^{m} \phi_{kj}W_k(x(t),y).
\tag{4}
$$

Let $\delta(\theta_k) = 0$ if θ_k is a zero vector; otherwise $\delta(\theta_k) = 1$. Therefore, defining the jump intensity $w_0 = \sum_{k=1}^{m} w_k(x,y)\delta(\theta_k)$, we have the following algorithm of stochastic simulation [4, 5, 6, 7] based on Piecewise deterministic Markov Process (PDMP) [8, 9].

Table 1: Algorithm of stochastic simulation for (2) based on PDMP

Step 1. Initialization: set $t_0 = 0$ and fix initial number of molecules $(X_0/V, Y_0/V)$.

Step 2. Calculate the propensity function w_k, $k = 1,\cdots,m$.

Step 3. Generate one random number r_1 uniformly distributed in $[0,1)$.

Step 4. Integrate the following differential equations

$$
\frac{dx_j(t)}{dt} = \sum_{k=1}^{m} \phi_{kj}W_k(x(t),y) \text{ for } j = 1,\cdots,n_x
\tag{5}
$$

$$
\frac{dq(t)}{dt} = -w_0(x(t),y)q(t)
$$

$$
\text{with } x(t_i) = x_i,\ q(t_i) = 1
\tag{6}
$$

between t_i and $t_i + \Delta t_i$ with the stopping condition $q(t_i + \Delta t_i) = r_1$. Then we have Δt_i.

Step 5. Generate a second random number r_2 uniformly distributed in $[0,1)$. Choose μ_i so that $\mu_i =$ the smallest integer satisfying $\sum_{j'=1}^{\mu_i} w_{j'}(x,y) > r_2 w_0(x,y)$. Step 6. Execute the reaction μ_i, i.e., update (x,y). If $t_i > T_{max}$, terminate the computation. Otherwise, goto Step 2.

Clearly, the continuous variables x are governed by the deterministic system (4) and

change continuously during each time interval $[t_i, t_i + \Delta t_i)$ while the discrete variables y remain constant. Therefore, x are piecewise continuous variables and may change discretely at $t_i + \Delta t_i$. On the other hand, y evolve discretely with stochastic motion punctuated by a sequence of random waiting times Δt_i due to the master equation (3).

3.2 Hybrid System with Diffusion Process

Moreover, we can expand (1) to the second order of V to consider the diffusion effect [10], i.e.,

$$\frac{\partial P(x,y;t)}{\partial t} = \sum_{k=1}^{m} [w_k(x, y - \theta_k/V)P(x, y - \theta_k/V; t) - w_k(x,y)P(x,y;t)] \tag{7}$$

$$-\sum_{k=1}^{m} \sum_{j=1}^{n_x} \frac{\partial}{\partial x_j} [g_{jk}(x,y;t)P(x,y;t)] \tag{8}$$

$$+\sum_{j=1}^{n_x} \sum_{l=1}^{n_n} \frac{\partial^2}{\partial x_j \partial x_k} \left(\sum_{k=1}^{m} \frac{\phi_{kj}\phi_{kl}}{2V} W_k(x,y)P(x,y;t) \right) \tag{9}$$

$$+\mathcal{O}(\frac{1}{V^2}) \tag{10}$$

where

$$g_{jk}(x,y;t) = \phi_{kj}W_k(x,y) - \sum_{l=1}^{n_y} \frac{\phi_{kj}\theta_{kl}}{VP(x,y;t)} \frac{\partial W_k(x,y)P(x,y;t)}{\partial y_l}$$

$$= \phi_{kj}W_k(x,y) - \sum_{l=1}^{n_y} \frac{\phi_{kj}\theta_{kl}}{V} [\frac{\partial W_k(x,y)}{\partial y_l}$$

$$+ W_k(x,y)\frac{\partial lnP(x,y;t)}{\partial y_l}].$$

Therefore, (8)-(9) can be also expressed by Langevin equations instead of the differential equations (4), i.e., stochastic differential equations with the continuous variables x for $j = 1, \cdots, n$

$$\frac{dx_j(t)}{dt} = \sum_{k=1}^{m} g_{jk}(x(t),y;t) + \sum_{k=1}^{m} \frac{\phi_{kj}}{\sqrt{V}} \sqrt{W_k(x(t),y)} \Gamma_k(t), \tag{11}$$

where $\Gamma_k(t)$ are temporally uncorrelated, statistically independent Gaussian white noises, and are formally defined by

$$\Gamma_k(t) = \lim_{dt \to \infty} \mathcal{N}(0, 1/dt), \tag{12}$$

where $\mathcal{N}(m, \sigma^2)$ denotes the normal random variable with mean m and variance σ^2.

For this case, the hybrid system is the combination of discrete stochastic system (3) for y and continuous stochastic system (11) for x, which can be simulated similarly as the algorithm of Table 1, i.e., Table 2.

In Table 2, $V_j(t)$ are independent one-dimensional Wiener processes. The stochastic differential equation can be calculated by Itô integration. Clearly, there is $P(x(t), y; t)$ in

Table 2: Algorithm of stochastic simulation for (7)-(10) based on PDMP

Step 1. Initialization: set $t_0 = 0$ and fix initial number of molecules $(X_0/V, Y_0/V)$.

Step 2. Calculate the propensity function w_k, $k = 1, \cdots, m$.

Step 3. Generate one random number r_1 uniformly distributed in $[0, 1)$.

Step 4. Integrate the following stochastic differential equations

$$dx_j(t) = \sum_{k=1}^{m} g_{jk}(x(t), y; t)dt + \sum_{k=1}^{m} \frac{\phi_{kj}}{\sqrt{V}} \sqrt{W_k(x(t), y)}dV_k(t)$$

$$\text{for } j = 1, \cdots, n_x, \tag{13}$$

$$dq(t) = -w_0(x(t), y)q(t)dt \text{ with } x(t_i) = x_i, q(t_i) = 1 \tag{14}$$

between t_i and $t_i + \Delta t_i$ with the stopping condition $q(t_i + \Delta t_i) = r_1$. Then we have Δt_i.

Step 5. Generate a second random number r_2 uniformly distributed in $[0, 1)$. Choose μ_i so that μ_i = the smallest integer satisfying $\sum_{j'=1}^{\mu_i} w_{j'}(x, y) > r_2 w_0(x, y)$.

Step 6. Execute the reaction μ_i, i.e., update (x, y). If $t_i > T_{max}$, terminate the computation. Otherwise, goto Step 2.

(11) or g_{jk}, which is required to be estimated during the integration. There are many ways to approximate $P(x(t), y; t)$ such as by finite state projection approach, Gaussian distribution assumption for the continuous variables, or the equilibrium probability distribution.

In this paper, we consider the following scheme to approximately estimate $\partial lnP(x, y; t)/\partial y_l$. Since y corresponds to discrete variables which are expected to change dynamics in a slow manner in contrast to the continuous variables x, we assume $\partial P(x(t), y; t)/\partial y_l \approx 0$ or $\partial lnP(x(t), y; t)/\partial y_l \approx 0$. Specifically, we have

$$g_{jk}(x, y; t) = \phi_{kj}W_k(x, y) - \sum_{l=1}^{n_y} \frac{\phi_{kj}\theta_{kl}}{V} \frac{\partial W_k(x, y)}{\partial y_l}. \tag{15}$$

4 Conclusion

In this paper, we theoretically provide a general framework to derive molecular networks with stochasticity based on hybrid systems. We exploit the fast-slow dynamics of biological systems to reduce the dimensionality, and take advantage of special interaction structure of continuous - discrete variables to simplify the mathematical model, which significantly reduce the complexity of molecular networks.

References

[1] T. B. Kepler and T. C. Elston (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations, *Biophysical Journal*, **81**, 3116-3136.

[2] N. G. van Kampen (1992) Stochastic processes in physics and chemistry, *North-Holland*.

[3] S. A. Isaacson, (2009) The Reaction-Diffusion Master Equation as an Asymptotic Approximation of Diffusion to a Small Target. SIAM J. Appl. Math. **70**, 77-111.

[4] D. T. Gillespie (1976) *J. Comput. Phys.*, **22**, 403-1976.

[5] D. T. Gillespie (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, **115**, 1716-1733.

[6] D. T. Gillespie (2000). The chemical Langevin equation, *Journal of Chemical Physics*, **113**, pp.297-306.

[7] D. T. Gillespie. (2002) The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction, *Journal of Phys. Chem. A*, **106**, 5063-5071.

[8] S. Zeiser, U. Franz, O. Wittich, V. Liebscher, (2008) Simulation of genetic networks modelled by piecewise deterministic Markov processes. *IET Systems Biology*, **2**, 113-135.

[9] A. Crudu, A. Debussche, O. Radulescu, (2009) Hybrid stochastic simplifications for multiscale gene networks. BMC Systems Biology. In press.

[10] H. Risken (1989) The Fokker-Planck equations, *Springer-Verlag*.