Mathematical Approaches for Analysis of Biochemical Reaction Networks

Chang Hyeong Lee

Abstract. A biochemical reaction network is the system in which biochemical species interact through various reaction channels. Many chemical or biochemical systems such as signal transduction pathways, gene regulatory networks and enzymatic reaction networks are modelled as biochemical reaction networks. To describe the time evolution of such systems quantitatively or qualitatively, mathematical methods and computational tools have been used. In this paper, we introduce mathematical and computational methods for analysing biochemical reaction networks. We review recent advances and discuss future works in the area of biochemical reaction networks.

1. Overview of Biochemical Reaction Networks

In a cell of living organisms, changes in the concentration or molecular number of biochemical species occur through various reaction channels, and they affect the evolution and the mutation of the cell. To describe such changes, researchers have tried to find elaborate mathematical and computational ways of modelling. A common way of describing a chemical reaction is to use the form

$$\sum_i a_{ij} S_i \xrightarrow{k} \sum_i b_{ij} S_i \,,$$

where S_i is ith species, and a_{ij}, b_{ij} are the stoichiometric coefficients of ith species as the reactant and the product of jth reaction, respectively. For example, a reaction $A+B \xrightarrow{k} C$ means one molecule or one mole of species A and that of B react with the reaction rate constant k and they produce one molecule or one mole of species C. Here the reaction rate is determined by mass-action kinetics. The mass-action kinetics means that the reaction rate is proportional to concentration or molecular number of reactant [1]. The proportional constant for the reaction rate is assumed to be determined by experiments. Since most of real chemical systems are complex, the graph theory is useful for analysis of complex networks; In the graph theory, each reaction can be regarded as a directed arrow, and each reactant or product as a node. For example, if we consider a reaction $A+B \xrightarrow{k} C$, the reactant $A+B$ and the produce C can be regarded as nodes, and the reaction as a directed edge. The graph is embedded with reaction rate and stoichiometric amount. This approach can be used for analysis of large scale reaction networks such as complex metabolic networks [2].

In Sec. 2, we describe ways of mathematical modelling for biochemical reaction networks. Throughout this paper, a vector is denoted by a boldfaced small letter.

2. Mathematical Modelling of Biochemical Reaction Networks

2.1. Deterministic modelling

The time-dependent dynamics of reaction networks has been traditionally described by deterministic differential equations (that is, given initial conditions, the system dynamics is completely determined);

$$\frac{d\mathbf{c}}{dt} = VR(\mathbf{c}), \tag{1}$$

where $\mathbf{c}(t)$ is the vector of concentration or number of molecules of species at time t, $R(\mathbf{c})$ is the vector of the reaction rate functions, and V is the stoichiometric matrix. More precisely, $R_j(\mathbf{c})$, the jth entry of $R(\mathbf{c})$, denotes the reaction rate of jth reaction, (i,j) entry of V is the stoichiometric amount of ith species changed by an occurrence of jth reaction. For large scale biological models which have sufficiently many species and reactions, deterministic description is generally accurate, and mathematical and computational methods have been developed for finding the solution of the governing equation (1) for deterministic models.

The investigation on the steady-state solution as well as the time-dependent solution is also important in that it shows how the system will be at the final stage and the stability of the system at the equilibrium. Throughout the 1960's and 1970's, many researchers have made progresses

on the existence and uniqueness of the steady-state solution by combining graph theory, differential equation and chemical reaction network theory [3–6]. Especially, Feinberg proved a very important property about the steady-state solution [6];

Theorem 1. *(Feinberg) Suppose that a reaction network is strongly connected with deficiency zero. Then, independent of the choice of reaction rate constants, there is precisely one equilibrium value that is locally asymptotically stable.*

This theorem implies that one can determine the existence and uniqueness of the stable steady-state for a general class of reaction networks that satisfy an easy-checkable topological property. It can be usefully applied to large complex networks whose dynamical properties at the equilibrium are difficult to analyse.

2.2. Stochastic modelling

As researches in biological sciences have recently been directed to small biological systems, more elaborate ways of modelling have been required. In a reaction network with small number of reactions and molecular species, random movement of molecules causes important effects such as mutation and evolution [7]. Deterministic differential equations cannot capture random and probabilistic events, and one has to use stochastic modelling and probabilistic methods for describing them. In the discrete stochastic modelling, the vector of the number of molecules of every species is the state variable $\mathbf{n}$ and ith entry $\mathbf{n}_i$ of $\mathbf{n}$ is the number of molecules of ith species. If the network has r reactions in total, the governing equation is written as

$$\frac{dp(\mathbf{n},t)}{dt} = \sum_{k=1}^{r} [a_k(\mathbf{n}-V_k)p(\mathbf{n}-V_k,t) - a_k(\mathbf{n})p(\mathbf{n},t)]\,. \tag{2}$$

In (2), V_k denotes kth column vector of the stoichiometric matrix V and a_k is so-called propensity function that is the probability that kth reaction occurs per unit time [8]. Generally the propensity is a nonlinear function of $\mathbf{n}$ if the system has nonlinear reactions, and it is very difficult, if not impossible, to find the solution of (2) due to high dimensionality of the state variables.

If the number of reactions and species is small, given initial conditions, one can identify all possible states and transition probability between the states, and one can construct a Markov chain that describes the stochastic dynamics of the state variables. The governing equation of the Markov chain is so-called Kolmogorov equation

$$\frac{d\mathbf{p}}{dt} = K\mathbf{p}\,, \tag{3}$$

where $\mathbf{p}$ is the vector of probability of all states $\mathbf{n}$ and K is the transition matrix whose entries are transition probabilities between states [9]. If the number of possible states is small, one can find the solution of (3) analytically or computationally. However, real complex biological models have large number of reactions and species, and it is very difficult or impossible to find the solution of (2) and (3), because the dimension of state variable $\mathbf{n}$ is high or infinite. To avoid such difficulties in finding the solution of the governing equation, alternatively researchers have been simulating stochastic trajectory of the evolution of the state variables by using Monte-Carlo type simulation algorithms. The well-known stochastic algorithm is the Gillespie stochastic simulation algorithm which is a simple but elegant Monte-Carlo type algorithm. Here we end this section by introducing the Gillespie algorithm [10];

Exact stochastic algorithm (Gillespie)

0. Set initial condition $\mathbf{n}(0)$.

1. Calculate

- the reaction rates $R_\ell(\mathbf{n})$ for each ℓ
- the sum of reaction rates $R_{tot} = \sum_{\ell=1}^{r} R_\ell(\mathbf{n})$

2. Generate two random numbers r_1 and r_2 from the uniform distribution $(0,1)$.
Set $\tau = -\frac{\log(r_1)}{R_{tot}}$ and choose k such that $\sum_{l=1}^{k-1} R_\ell(\mathbf{n}) < r_2 R_{tot} \le \sum_{l=1}^{k} R_\ell(\mathbf{n})$

3. Let $t \leftarrow t + \tau$

Let $\mathbf{n} \leftarrow \mathbf{n} + V_k$

Go to 1.

3. Recent Research Directions

In this section we introduce research topics that have been getting more attention recently.

3.1. Multiple time scale reaction networks

In many reaction networks in cells, while some reactions occur quickly and they reach an equilibrium, others occur slowly and dominate the long time dynamics. Generally, the former is called fast

reactions and the latter slow reactions. The network with the fast and slow reactions is called a two time scale reaction network. If there are more time scale discrepancies of reactions occurring, the network is called a multiple time scale reaction network. In a multiple time scale reaction network, fast reactions dominate the initial dynamics and slow reactions determine the long time dynamics. The dynamics of the reaction networks with fast and slow reactions is determined by the system of differential equations with various magnitude of kinetic parameters and many variables. Thus, it is important but difficult to analyse or compute the dynamics of large scale reaction networks such as metabolic networks in living organism and signal transduction pathways. In many cases, researchers are interested in the slow dynamics and computations for the slow dynamics need less computational cost than fast dynamics. To obtain the governing equation on the slow time scale, one can use proper approximation methods by applying perturbation method. By the quasi-steady-state assumption, slow dynamics is approximated after elimination of fast dynamics [11–14].

For analysis of stochastic models of multiple time scale reaction networks, one has to solve Eq. (3), but it is very difficult, if not impossible, to find its solution analytically or computationally due to high dimensionality. Alternatively, if we use stochastic simulation algorithm (SSA) described in Sec. 2, fast reactions occur very often in any small time interval and randomly chosen time step τ in SSA should be very small. This implies that intensive and expensive computations are required to obtain simulation results. However, if long time dynamics is the main interest, one can reduce computational cost by obtaining and computing the reduced governing equation on slow time scale. To obtain the reduced governing equation of stochastic models on slow time scale, one needs to approximate fast dynamics. One method for approximating the fast dynamics is to find the steady-state probability of fast dynamics and use it for reducing the governing equation on slow time scale [9, 7, 16, 17]. Generally, the steady-state probability of fast dynamics can be obtained by solving difficult partial differential equation. Recently, it has been shown in [18] that steady-state probability for a strongly connected reaction network with deficiency 0 is given as the product of Poisson distributions with parameters as steady-state value of deterministic models. This result can be used to find the quasi-steady solution of general reaction networks with proper conditions.

3.2. Stochastic models of nonlinear reactions

Nonlinear reactions occur in most living organisms and they are essential reactions for which the number of reactant species is 2 or more. According to the number of reactants, nonlinear reactions are classified as bimolecular reactions, trimolecular reactions, and so on. Since a reaction involving more than three molecular entities is very rare and many trimolecular reactions can be considered as two consecutive bimolecular reactions, researches on nonlinear reactions have been focused on bimolecular reactions. Indeed, many bimolecular reactions are known to explain important biological processes. For example, the bimolecular reaction $A + B \to C$ describes basic reactions such as ligand binding and enzyme-substrate binding. The autocatalysis $A + B \to 2A$ is also a common bimolecular reaction and it is known that it is relevant to the mechanism of Creutzfeldt-Jakob disease [19]. Due to high dimensionality of state variables and nonlinearity of reactions, it is very difficult to find the explicit form of solutions for stochastic models of nonlinear reaction networks except for simple cases. Since nonlinear reactions are essential and important reactions for complex biological and chemical reaction systems such as enzyme reaction networks, signal transduction pathways and gene networks, the analysis of nonlinear reactions can give mathematical tools and insights for analysis of stochastic models of more complex reaction networks.

3.3. Reaction-Diffusion Network

In previous sections, we discussed the methods for analysing the reaction networks. The assumption on the reaction network theory is that the given network is spatially well-mixed and the spatial effect can be ignored. However, in most of living cells, diffusive transports of molecular species occur as well as reactions and play important roles of evolution of cell. By adding the diffusive transports to reaction networks,

researchers constructed the reaction-diffusion network model. In the 1950's, Turing recognised the importance of the diffusive transport, when he studied the dynamics of pattern formation [20]. He considered a chemical system of two species, an activator and an inhibitor, and he showed that the system is stable to any small perturbations if diffusion is not present, but it is unstable when diffusion is present. This mechanism is called Turing instability, which causes the creation of spatial patterns. After his work, there have been many studies about the reaction-diffusion models of pattern formation. If the network has large-scale, reaction-diffusion partial differential equations have been used to analysing its dynamics and many computational methods have been developed for solving the equations. However, in case that the network is sufficiently small and random collision of molecules affects its dynamics, one has to model the network stochastically. One common way of modelling reaction-diffusion networks stochastically is by construction of localised reaction compartments. In the stochastic compartment model, reactions between species occur in each compartment and diffusive transport of molecules of species occurs between adjacent compartments. A detailed description of the stochastic compartment model is as follows; We first suppose that the system has s species which interact through r reactions in each of m identical compartments and perform diffusive transport between adjacent compartments. To derive the governing equation, we denote the number of molecules of ith species in jth compartment by N_i^j. Since the diffusive transport between adjacent compartments can be considered as a first-order reaction, the diffusion of ith species from jth to kth compartment is modelled as a first-order reaction with the reaction rate d_i^{jk}. The rate d_i^{jk} is the jump rate of ith species from jth to kth compartment and it can be computed as

$$d_i^{jk} = \frac{D_i}{h^2},$$

where D_i is the bulk diffusion constant for ith species and h is the uniform mesh size [21]. Let $\mathbf{N} = (\mathbf{N}^1, \mathbf{N}^2, \ldots, \mathbf{N}^m)^T$ denote the vector of numbers of molecules of species in the network. Here $\mathbf{N}^i = (N_1^i, \ldots, N_s^i)$ denotes a vector of numbers of molecules of species in ith compartment. We let $\tilde{\mathbf{e}}_k$ and $\mathbf{e}_k$ be an $m \times 1$ and an $s \times 1$ unit vector, respectively, whose kth entry is 1 and 0 otherwise, and we define $\mathbf{N}(t) = \mathbf{n}$. Using these notation, we can write the master equation for reaction-diffusion system by

$$\begin{aligned}\frac{dp(\mathbf{n})}{dt} &= \sum_{l=1}^{r}\sum_{k=1}^{m} a_l^k(\mathbf{n}^k - V_l)p(\mathbf{n} - \tilde{\mathbf{e}}_k \otimes V_l) - a_l^k(\mathbf{n}^k)p(\mathbf{n}) \\ &\quad + \sum_{k=1}^{m}\sum_{j=1}^{m}\sum_{i=1}^{s} d_i^{jk}(n_i^j + 1)p(\mathbf{n} + \tilde{\mathbf{e}}_j \otimes \mathbf{e}_i - \tilde{\mathbf{e}}_k \otimes \mathbf{e}_i) \\ &\quad - d_i^{jk} n_i^j p(\mathbf{n}),\end{aligned}$$

where $a_l^k(\cdot)$ is the propensity of lth reaction in kth compartment, and $\otimes$ denotes the tensor product. Due to high dimensionality of variable $\mathbf{n}$ and complexity of network structure of reaction-diffusion system, it is very difficult, if not impossible, to find the solution of (4) analytically and also it needs heavy computations with high performance computing machines to find computational solutions. Another issue is that the basis for the choice of compartment size h has not been clarified; Suppose a reaction-diffusion system contains species with diffusion rates of different time scales. If the compartment size is chosen for fast-diffusing species, the spatially homogeneous assumption for each compartment may not be true for slowly-diffusing species. If one chooses the compartment size for slowly-diffusing species, any stochastic simulation algorithm will be computationally inefficient [22].

References

[1] F. Horn and R. Jackson, General mass action kinetics, *Archive for Rational Mechanics and Analysis* **47**(2) (1972) 81–116.

[2] H. G. Othmer, *A Graph-theoretic Analysis of Chemical Reaction Networks*, Lecture Notes, Rutgers University (1979).

[3] F. Heineken, H. Tsuchiya and R. Aris, On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics, *Mathematical Biosciences* **1** (1967) 95–113.

[4] F. Horn, Necessary and sufficient conditions for complex balancing in chemical kinetics, *Archive for Rational Mechanics and Analysis* **49**(3) (1972) 172–186.

[5] H. G. Othmer, Nonuniqueness of equilibria in closed reacting systems, *Chemical Engineering Science* **31** (1976) 993–1003.

[6] M. Feinberg, Lecture Notes on Chemical Reaction Networks (1979).

[7] C. V. Rao, M. W. Wolf and A. P. Arkin, Control, exploitation and tolerance of intracellular noise, *Nature* **420** (2002) 231–237.

[8] D. T. Gillespie, A rigorous derivation of the chemical master equation, *Physica A* **188** (1992) 404–425.

[9] C. H. Lee and R. Lui, A reduction method for multiple time scale stochastic reaction networks, *J. Mathematical Chemistry* **46** (2009) 1292–1321.

[10] D. T. Gillespie, Exact simulation of coupled chemical reactions, *J. Physical Chemistry* **81** (1977) 2340–2361.
[11] L. A. Segel and M. Slemrod, The quasi-steady-state assumption: a case study in perturbation, *SIAM Review* **31** (1989) 446–477.
[12] A. Kumar and P. Daoutidis, *Control of Nonlinear Differential Algebraic Equation Systems* (Chapman and Hall/CRC, London, 1999).
[13] M. Stiefenhofer, Quasi-steady-state approximation for chemical reaction networks, *J. Mathematical Biology* **36** (1998) 593–609.
[14] C. H. Lee and H. G. Othmer, A multi-time-scale analysis of chemical reaction networks: I. Deterministic systems, *J. Mathematical Biology* **60** (2010) 387–450.
[15] C. V. Rao and A. P. Arkin, Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm, *J. Chemical Physics* **118** (2003) 4999.
[16] E. L. Haseltine and J. B. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics, *J. Chemical Physics* **117** (2002) 6959.
[17] J. Goutsias, Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems, *J. Chemical Physics* **122** (2005) 184102.
[18] D. F. Anderson, G. Craciun and T. G. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, *Bulletin of Mathematical Biology* **72** (2010) 1947–1970.
[19] J. Bieschke, P. Weber, N. Sarafoff, M. Beekes, A. Giese and H. Kretzschmar, Autocatalytic self-propagation of misfolded prion protein, *Proc. Natl. Acad. Sci. U.S.A* **101** (2004) 12207.
[20] A. M. Turing, The chemical basis of morphogenesis, *Phil. Trans. R. Soc. Lond. B* **237** (1952) 37.
[21] A. B. Stundzia and C. J. Lumsden, Stochastic simulation of coupled reaction-diffusion processes, *J. Computational Physics* **127** (1996) 196.
[22] S. Isaacson and C. Peskin, Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations, *SIAM J. Sci. Comput.* **28** (2005) 47–74.

Translated from the Newsletter of Korean Mathematical Society ***136*** *(2011) 7 –12*

Chang Hyeong Lee
Ulsan National Institute of Science and Technology (UNIST), Korea
chlee@unist.ac.kr

Professor Chang Hyeong Lee received the BS degree in mathematics from Korea University, Seoul, Korea, and the PhD degree in mathematics from the University of Minnesota, Minneapolis, US in 2006. He was a visiting assistant professor at the Department of Mathematical Sciences, WPI (Worcester Polytechnic Institutes) from 2006–2009. He has been an assistant professor at School of Technology Management and Applied Mathematics Graduate Program at UNIST (Ulsan National Institute of Science and Technology), Ulsan, Korea from 2009–present. His research interests are in stochastic analysis and computation of complex networks, epidemic modeling and financial engineering.