Computers and Chemical Engineering 90 (2016) 111–120

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Regularized maximum likelihood estimation of sparse stochastic monomolecular biochemical reaction networks☆

Hong Jang[a], Kwang-Ki K. Kim[b], Richard D. Braatz[c], R. Bhushan Gopaluni[d], Jay H. Lee[a,*]

[a] Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
[b] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta 30332, GA, USA
[c] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge 02139, MA, USA
[d] Department of Chemical and Biological Engineering, University of British Columbia, 2360 East Mall, Vancouver BC V6T1Z3, Canada

ARTICLE INFO

Article history:
Received 5 November 2015
Received in revised form 21 March 2016
Accepted 21 March 2016
Available online 26 April 2016

Keywords:
Sparse parameter estimation
Regularized maximum likelihood estimation
Mono-molecular biochemical reaction network
Chemical master equation
Stochastic simulation algorithm

ABSTRACT

A sparse parameter matrix estimation method is proposed for identifying a stochastic monomolecular biochemical reaction network system. Identification of a reaction network can be achieved by estimating a sparse parameter matrix containing the reaction network structure and kinetics information. Stochastic dynamics of a biochemical reaction network system is usually modeled by a chemical master equation (CME) describing the time evolution of probability distributions for all possible states. This paper considers closed monomolecular reaction systems for which an exact analytical solution of the corresponding chemical master equation can be derived. The estimation method presented in this paper incorporates the closed-form solution into a regularized maximum likelihood estimation (MLE) for which model complexity is penalized. A simulation result is provided to verify performance improvement of regularized MLE over least-square estimation (LSE), which is based on a deterministic mass-average model, in the case of a small population size.

1. Introduction

Stochastic dynamics of biological systems have lately received increased attention from researchers in the field of biological engineering (Raj and van Oudenaarden, 2009). In the past, such studies were greatly hampered by lack of qualitative measurement data. Nowadays, quantitative but noisy data can be obtained using the microarray technology. Recent developments in sensing techniques that can provide real-time observations of intrinsic stochastic dynamics at small length scales have motivated many scientific investigations (Raj et al., 2010; Xie et al., 2008). One such sensing technique for biological systems is bio-imaging using fluorescent proteins (García-Parajó et al., 2001). By grafting a fluorescent protein into gene expression, proteins and mRNA expressions originating from targeted DNA can be detected quantitatively in real time. Specifically, yellow fluorescent protein (yfp) (Elowitz et al., 2002; Li and Xie, 2011; Yu et al., 2006) has been widely used for detecting changes with single-macromolecule sensitivity in individual live cells.

In the real-time data reported in the aforementioned papers, strongly stochastic behavior has been observed. For example, bursts of transcribed protein molecules from the cell, controlled by an identical messenger RNA molecule, have different copy numbers (Elowitz et al., 2002). The total population number of species in the system within the detectable range is small, ranging from tens to thousands of copies. Stochastic dynamics of systems with discrete states can be modeled by the chemical master equation (CME) (Feinberg, 1979; Fichthorn and Weinberg, 1991),

$$\frac{\partial P(\boldsymbol{\sigma}, t)}{\partial t} = \sum_{\boldsymbol{\sigma}'} W\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}\right) P\left(\boldsymbol{\sigma}', t\right) - \sum_{\boldsymbol{\sigma}'} W\left(\boldsymbol{\sigma}, \boldsymbol{\sigma}'\right) P(\boldsymbol{\sigma}, t) \tag{1}$$

where $P(\boldsymbol{\sigma}, t)$ is the probability of the system being in discrete state $\boldsymbol{\sigma}$ at time t, and $W(\boldsymbol{\sigma}', \boldsymbol{\sigma})$ is the transition rate from state $\boldsymbol{\sigma}'$ to state $\boldsymbol{\sigma}$. The CME describes the time evolution of the probability distri-

☆ Preliminary results in this manuscript were published in conference papers "H. Jang, K. K. K. Kim, J. H. Lee, and R. D. Braatz, Regularized Maximum Likelihood Estimation of Sparse Stochastic Monomolecular Biochemical Reaction Networks, Proceedings of the IFAC World Congress, Cape Town, South Africa, 2014" and "K. K. K. Kim, H. Jang, R. B. Gopaluni, J. H. Lee and R. D. Braatz, Sparse Identification in Chemical Master Equations for Monomolecular Reaction Networks, Proceedings of the American Control Conference, Portland, Oregon, 2014."

* Corresponding author. Fax: +82 42 350 3910.
E-mail address: jayhlee@kaist.ac.kr (J.H. Lee).

http://dx.doi.org/10.1016/j.compchemeng.2016.03.018
0098-1354/

bution among all possible configurations, and can be written as (MacNamara et al., 2008; Munsky and Khammash, 2006)

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{A}\left(t;\theta\right)\boldsymbol{P}(t) \tag{2}$$

where $\boldsymbol{P}(t)$ is the state vector containing all the state probability variables and $\boldsymbol{A}\left(t;\theta\right)$ is a matrix containing all the transition rate constants with dependence on the model parameter vector θ related with physicochemical phenomena, e.g., biochemical reactions. Many numerical algorithms have been developed for solving the matrix ordinary differential Eq. (2), which can be divided into direct methods (MacNamara et al., 2008; Munsky and Khammash, 2006) and indirect methods (Gibson and Bruck, 2000; Gillespie, 1977). Direct methods, such as the finite state projection (FSP) algorithm, attempt to evaluate the matrix exponential directly. In practice, given the large size of the state space, indirect methods that use stochastic simulation algorithms (SSAs) to generate approximate probability distributions have been more popular.

A reaction network is one of fundamental models for describing biological mechanisms. Different types of biological reaction networks exist and interact to accomplish needed functions including gene regulatory networks, metabolic networks, and signaling networks. Network identification is an important aspect of studying biological network systems. Given time series experimental data from sensors, a parameter estimation method can be used to identify the parameters of a given reaction network model. Typically, the estimation is formulated to find parameter values minimizing the difference between the experimental data and their model predictions. Most of the literature has employed least-squares estimation (LSE) approaches, which fit stochastic data to a deterministic continuum model (Gennemark and Wedelin, 2009).

The LSE method based on deterministic models, however, can provide poor parameter estimates for highly stochastic systems (Tian et al., 2007). Several stochastic parameter estimation methods based on the stochastic differential equation of type (1) have shown improved estimation performance (Munsky et al., 2012). These methods attempt to solve for the probability density functions (PDFs) of the CME and use them for estimation. Previous studies employed the moment-based method (Munsky et al., 2009; Zechner et al., 2012), the Bayesian method (Boys et al., 2008; Golightly and Wilkinson, 2011; Lillacci and Khammash, 2010), the maximum likelihood estimation (MLE) method (Daigle et al., 2012; Tian et al., 2007), and the density function distance (DFD) method (Lillacci and Khammash, 2013; Poovathingal and Gunawan, 2010). However, these published methods are based on approximated PDFs of the CME. In many cases, PDFs are approximated using the SSA approach (Gillespie, 2007), which typically demands a very large number of simulations to be performed for an accurate estimation.

In many cases, the exact network structure of a biological system is not known. An important problem in biological systems is to use measurement data to identify the topology of reaction networks and estimate associated parameters at the same time. Craciun and Pantea, 2008 discussed the issue of identifiability of reaction networks with the fact that there might exist more than one parameter set that fits the measurement data with the same level of accuracy. To cope with such challenges, model discrimination/invalidation of reaction networks has been studied (Conradi et al., 2005; Kremling et al., 2004). Non-uniqueness of an estimation algorithm can be overcome by using the principle of parsimony and choosing simple models over complex models (Jefferys and Berger, 1991). An approach to reduce model complexity is to use penalization term for the number of parameters such as ridge and ℓ_1 regularization (Hesterberg et al., 2008).

Taking the above issues into consideration altogether, this paper considers stochastic monomolecular reaction systems in a sparse parameter matrix estimation problem. Stochastic monomolecular reaction system, described by an exact probability distribution solution of the CME (Jahnke and Huisinga, 2007), can represent realistic gene regulatory networks or metabolic networks. The exact solution enables formulation of a regularized MLE method rather than LSE. Improved performance over the LSE method is verified with a simulation study involving a small scale reaction network system. Applicability of the proposed MLE method to a more stochastic and larger scale reaction network system is also tested.

2. Biochemical reaction network system

In a living organism, most biological functions arise from complex interactions between numerous components such as genes, metabolites, and proteins. The interactions form a large biochemical reaction network system involving thousands of components. Gene expression is the main mechanism by which cells regulate the interaction webs to perform functions. The gene expression process occurs in two steps: (1) transcription of genes into mRNAs initiated by specialized proteins and (2) translation of mRNAs into biochemically active proteins. By detecting the gene expression time-profiles, we can construct a model of the gene–gene interaction which can be a subset of a larger network (Fig. 1) (Gardner et al., 2003). This gene network identification has important potential applications, for example, in drug discovery to identify candidate pathways to be targeted (Schreiber, 2000).

Gene regulatory networks as well as other complex biochemical reaction networks, such as metabolic networks (Feist et al., 2008; Jeong et al., 2000) and signal networks (Hyduke and Palsson, 2010), can be described by a combination of monomolecular reactions (Radulescu et al., 2012). Possible monomolecular reactions can be categorized with conversion, inflow, and outflow reactions with a set of $n \in \mathbb{N}$ different species or complexes denoted by S_i, $i = 1, \ldots, n$. The reactions are given by

$$\begin{aligned} &S_i \xrightarrow{k_{ij}} S_j \quad \text{Conversion}(i \neq j) \\ &S_o \xrightarrow{k_{0j}} S_j \quad \text{Inflow} \\ &S_i \xrightarrow{k_{i0}} S_* \quad \text{Outflow} \end{aligned} \tag{3}$$

where S_o and S_* are pseudo-species outside the system and k_{ij} is a nonnegative rate constant for reaction from S_i to S_j for $i \neq j$ and can be time-varying. The monomolecular conversion reaction excludes catalytic or splitting reactions of the types

$$\begin{aligned} &S_i \xrightarrow{k} S_i + S_j \quad \text{Catalytic}(i \neq j) \\ &S_i \xrightarrow{k} S_j + S_r \quad \text{Splitting}(i \neq j \neq r) \end{aligned} \tag{4}$$

If the number of species in the system is sufficiently large, the dynamics of the system can be described by the deterministic ordinary differential equations

$$\frac{\mathrm{d}C_i(t)}{\mathrm{d}t} = \sum_{j \neq i} k_{ji}(t)\, C_j(t) - \sum_{j \neq i} k_{ij}(t)\, C_i(t), \; i = 1, \ldots, n \tag{5}$$

where $C_i(t)$ is the population density or concentration of the species S_i and continuous variable. Eq. (5) can be written in a vector form as

$$\frac{\mathrm{d}\boldsymbol{C}(t)}{\mathrm{d}t} = \boldsymbol{A}(t)\,\boldsymbol{C}(t) \tag{6}$$

$$[\boldsymbol{A}]_{ij}(t) = k_{ji}(t), \; j \neq i \tag{7}$$

$$[\boldsymbol{A}]_{ii}(t) = -\sum_{j \neq i} k_{ij}(t) \tag{8}$$

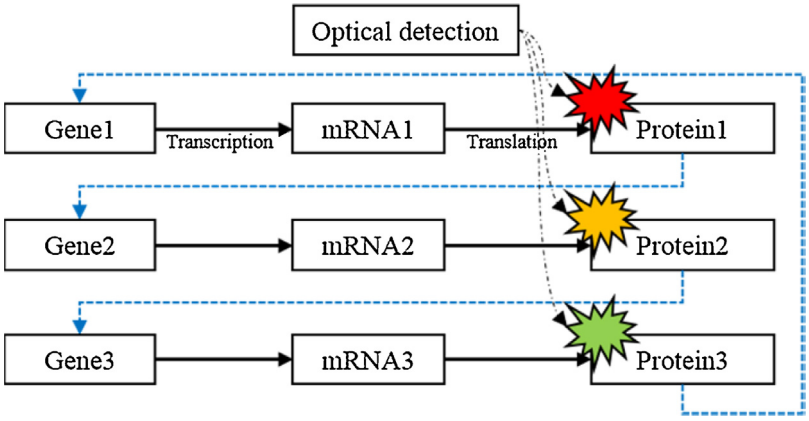

Fig. 1. A simple gene network system. Black solid arrows indicate the two-step gene expression process. Blue dotted arrows indicate the gene-gene interaction regulated by expressed proteins. Single-molecule-level gene expression can be detected by fluorescent proteins encoded with target genes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\boldsymbol{A}(t) \in \mathbb{R}^{n\times n}$ is the kinetic parameter matrix, $[\boldsymbol{A}]_{ij}$ is the element at the ith row and jth column of the matrix $\boldsymbol{A}(t)$, and $[\boldsymbol{A}]_{ii}$ is the ith diagonal element of the matrix $\boldsymbol{A}(t)$. If the reaction rates are constant, its analytical solution can be computed with initial concentration, $\boldsymbol{C}_0$, as

$$\boldsymbol{C}(t) = e^{\boldsymbol{A}t}\boldsymbol{C}_0 \tag{9}$$

In contrast, the concentrations in biochemical reactions are highly stochastic when there are only a small number of each species. For such small population systems, the CME describes the stochastic dynamics of the same reaction by (Gadgil et al., 2005; Jahnke and Huisinga, 2007),

$$\begin{aligned}\frac{\partial P(\boldsymbol{x},t)}{\partial t} &= \sum_{i=1}^{n} k_{0i}(t)\left(P(\boldsymbol{x}-\boldsymbol{e}_i,t)-P(\boldsymbol{x},t)\right)\\ &+\sum_{j=1}^{n} k_{j0}(t)\left(\left(x_j+1\right)P\left(\boldsymbol{x}+\boldsymbol{e}_j,t\right)-x_jP(\boldsymbol{x},t)\right)\\ &+\sum_{j=1}^{n}\sum_{i=1}^{n} k_{ji}(t)\left(\left(x_j+1\right)P\left(\boldsymbol{x}+\boldsymbol{e}_j-\boldsymbol{e}_i,t\right)-x_jP(\boldsymbol{x},t)\right)\end{aligned} \tag{10}$$

where $P(\boldsymbol{x},t)$ is a probability for the integer state vector $\boldsymbol{x} \in \mathbb{Z}^n$ with x_i as the population of the ith species and $\boldsymbol{e}_i \in \mathbb{R}^n$ denotes the standard basis vector (with 1 for the ith element and zero for the rest). In the right-hand side of (10), the first, second, and third terms describe the influences of the inflow, outflow, and conversion reactions, respectively. This paper considers an estimation problem of constant parameters for closed monomolecular reaction system with no inflow and outflow ($k_{0i} = 0$ for all i and $k_{j0} = 0$ for all j).

3. Exact maximum likelihood estimation

3.1. Exact Solution of the CME

In Jahnke and Huisinga (2007), analytical solutions for the CME (10) for both open and closed systems are derived. In this section, results in Jahnke and Huisinga (2007) for closed systems are highlighted. The multinomial distribution, $\mathcal{M}(\boldsymbol{x}, N, \boldsymbol{\lambda}(t))$, is defined by

$$\mathcal{M}(\boldsymbol{x}, N, \boldsymbol{\lambda}(t)) = N!\prod_{i=1}^{N}\frac{\lambda_i^{x_i}(t)}{x_i!} \tag{11}$$

where $\boldsymbol{\lambda}(t) \in \mathbb{R}^n$ is the probability parameter (also called the hyper-parameter).

Proposition 1 in Jahnke and Huisinga (2007): Suppose that $P(\boldsymbol{x},0) = \mathcal{M}(\boldsymbol{x}, N, \boldsymbol{\lambda}_0)$. Then $P(\boldsymbol{x},t) = \mathcal{M}(\boldsymbol{x}, N, \boldsymbol{\lambda}(t))$ solves the CME (10), where N is the total number of entities in the population in the system and $\boldsymbol{\lambda}(t) \in \mathbb{R}^n$ is the solution of the mass average rate-reaction Eq. (6) $\dot{\lambda}(t) = A\lambda(t)$ with the initial condition $\boldsymbol{\lambda}(0) = \boldsymbol{\lambda}_0$ and$\boldsymbol{A}$ where $[\boldsymbol{A}]_{ij,i\neq j} = k_{ji}$ and $[\boldsymbol{A}]_{ii} = -\sum_{j\neq i} k_{ij}$ for all $t \geq 0$ and $i, j = 1, \ldots, n$.

Proposition 1 in Jahnke and Huisinga (2007) states that a multinomial initial condition will remain multinomial as the species distribution evolves. In practice, ordinary biochemical reaction systems rarely have a multinomial distribution as an initial condition, but can have arbitrary deterministic initial condition defined by the delta function,

$$P(\boldsymbol{x}, t_0) = \delta_{\boldsymbol{\xi}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \boldsymbol{x} = \boldsymbol{\xi} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\delta_{\boldsymbol{\xi}}(x)$ is the Kronecker delta and $\boldsymbol{\xi}$ is a particular deterministic initial state.

Theorem 1 in Jahnke and Huisinga (2007): Suppose that the monomolecular reaction system has the initial distribution $P(\boldsymbol{x},0) = \delta_{\boldsymbol{\xi}}(x)$ for some $\boldsymbol{\xi} = \begin{bmatrix}\xi_1 & \cdots & \xi_n\end{bmatrix}^T \in \mathbb{R}^n$. The probability distribution at time t is defined by $P(\boldsymbol{x},t) = \mathcal{M}(\boldsymbol{x}, \xi_1, \boldsymbol{\lambda}^{(1)}(t)) * \mathcal{M}(\boldsymbol{x}, \xi_2, \boldsymbol{\lambda}^{(2)}(t)) * \cdots * \mathcal{M}(\boldsymbol{x}, \xi_n, \boldsymbol{\lambda}^{(n)}(t))$, where $\boldsymbol{\lambda}^{(i)}(t)$ is the solution of the ODE $\dot{\boldsymbol{\lambda}}^{(i)}(t) = \boldsymbol{A}\boldsymbol{\lambda}^{(i)}(t)$ with $\boldsymbol{\lambda}^{(i)}(0) = \boldsymbol{e}_i$ and $\boldsymbol{A}$ where $[\boldsymbol{A}]_{ij,i\neq j} = k_{ji}$ and $[\boldsymbol{A}]_{ii} = -\sum_{j\neq i} k_{ij}$ for all $t \geq 0$ and $i, j = 1, \ldots, n$.

The asterisk, $*$, is the convolution operator. The multidimensional discrete convolution of two mappings P_1 and P_2 is defined by

$$(P_1 * P_2)(\boldsymbol{x}) = \sum_{\boldsymbol{z}} P_1(\boldsymbol{z})P_2(\boldsymbol{x}-\boldsymbol{z}) = \sum_{\boldsymbol{z}} P_1(\boldsymbol{x}-\boldsymbol{z})P_2(\boldsymbol{z}) \tag{13}$$

where the sum is taken over all $\boldsymbol{z} \in \{0, 1, \ldots, N\}^n$ such that $(\boldsymbol{x}-\boldsymbol{z}) \in \{0, 1, \ldots, N\}^n$.

For the reaction system where an analytical solution of the CME is hard to derive, an approximate solution obtained by the finite state projection algorithm (Munsky and Khammash, 2006) or the stochastic simulation algorithms (Gillespie, 1977) can be used in the estimation. However, obtaining an approximate solution sufficiently close to an exact solution for an accurate estimation can significantly increase the computational load. Therefore, we recommend using the proposed estimation method only for a reaction

system where an analytical solution of the CME can be derived, such as the monomolecular reaction network system covered in this paper.

3.2. Maximum likelihood estimation formulation

The formulation of the parameter estimation method is based on the likelihood function, which is primarily defined by a conditional PDF of the measured sequence, $\hat{\boldsymbol{x}}_j(t_k) \in \mathbb{R}^n$ for the run index, $j = 1, \ldots, r$ and time index, $k = 1, \ldots, m$ for given the reaction parameter matrix, $\boldsymbol{K} \in \mathbb{R}^{n\times n}$,

$$L\left(\left\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\right\} | \boldsymbol{K}\right) \tag{14}$$

where $[\boldsymbol{K}]_{ij}$ is equal to k_{ij} and assumed to be time invariant. This paper assumes that the time-series data $\hat{\boldsymbol{x}}_j(t_k)$ of all species in the reaction network system can be obtained with minimal error, noise, and missing information. Although this may not be entirely realistic at this point, sensing techniques under development currently offer this possibility in the near future. The matrix, $\boldsymbol{K}$, is related to the kinetic matrix $\boldsymbol{A}$ in (6) given by

$$\boldsymbol{A} = \boldsymbol{K}^T - \mathrm{diag}(\boldsymbol{K}\boldsymbol{e}) \tag{15}$$

where $\boldsymbol{e} \in \mathbb{R}^n$ is a vector of all ones. With the Markov process assumption, the likelihood function can be defined by

$$L\left(\left\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\right\} | \boldsymbol{K}\right) = \prod_{j=1}^{r}\prod_{k=1}^{m-1} P\left(\hat{\boldsymbol{x}}_j(t_{k+1}) | \hat{\boldsymbol{x}}_j(t_k), \boldsymbol{K}\right) \tag{16}$$

With the assumption of the monomolecular reaction, the overall population $\hat{\boldsymbol{x}}_j(t_k)$ can be divided into independent subsets for each species, $\hat{\boldsymbol{x}}_j^{(i)}(t_k) \in \mathbb{R}^n$ for $i = 1, \ldots, n$, at time t_k,

$$\hat{\boldsymbol{x}}_j(t_k) = \hat{\boldsymbol{x}}_j^{(1)}(t_k) + \hat{\boldsymbol{x}}_j^{(2)}(t_k) + \ldots + \hat{\boldsymbol{x}}_j^{(n)}(t_k) \tag{17}$$

$$\hat{\boldsymbol{x}}_j^{(1)}(t_k) = \begin{bmatrix} \hat{x}_{j,1}(t_k) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \hat{\boldsymbol{x}}_j^{(2)}(t_k) = \begin{bmatrix} 0 \\ \hat{x}_{j,2}(t_k) \\ \vdots \\ 0 \end{bmatrix}, \ldots, \hat{\boldsymbol{x}}_j^{(n)}(t_k) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \hat{x}_{j,n}(t_k) \end{bmatrix} \tag{18}$$

The initial distribution of each subset, $\hat{\boldsymbol{x}}_j^{(i)}(t_k)$, can be defined by the multinomial distribution based on (12) and can be evolved to the next time step t_{k+1} independently. Then the joint probability for all subset at time t_{k+1} is obtained by convoluting all PDFs, which remain as multinomial distributions based on *Theorem* 1 in Jahnke and Huisinga (2007). Then, each conditional PDF appearing in the multiplication of (16) follows from the exact solution of the CME for an arbitrary deterministic initial condition,

$$\mathrm{L}(\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\}|\boldsymbol{K}) \prod_{j=1}^{r}\prod_{k=1}^{m-1} \begin{pmatrix} \mathrm{M}(\hat{\boldsymbol{x}}_j(t_{k+1}), \hat{x}_{j,1}(t_k), \boldsymbol{\lambda}^{(1)}(t_{k+1})) \\ *\mathrm{M}(\hat{\boldsymbol{x}}_j(t_{k+1}), \hat{x}_{j,2}(t_k), \boldsymbol{\lambda}^{(2)}(t_{k+1})) \\ \vdots \\ *\mathrm{M}(\hat{\boldsymbol{x}}_j(t_{k+1}), \hat{x}_{j,n}(t_k), \boldsymbol{\lambda}^{(n)}(t_{k+1})) \end{pmatrix} \tag{19}$$

Finally, the standard MLE that finds a parameter matrix having maximum value for the likelihood function is defined by

$$\max_{\boldsymbol{K}} L\left(\left\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\right\} | \boldsymbol{K}\right) \tag{20}$$

3.3. Sparse parameter estimation

Our approach to the sparse parameter estimation is an optimality criterion with the principle of parsimony (Jefferys and Berger, 1991). Let $\boldsymbol{k}$ denote the concatenation of the element of $\boldsymbol{K}$, a vectorized form of $\boldsymbol{K}$. For model selection, two conflicting objective functions of the parameter vector $\boldsymbol{k}$ can be considered: energy function and model complexity function. Energy function is the likelihood function in (19), which depends on the training dataset. The model complexity function is a function of ℓ_0 norm of $\boldsymbol{k}$, which is defined by

$$M(\boldsymbol{k}) = g(\|k\|_0) \tag{21}$$

where g is a monotonically increasing function. Note that model complexity function is independent of the training data set. A common objective function for the sparse parameter estimation is a combination of the energy function and model complexity function (August and Papachristodoulou, 2009; Boyd and Vandenberghe, 2004),

$$J(\boldsymbol{k}) = L\left(\left\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\right\} | \boldsymbol{k}\right) - \gamma g(\|k\|_0), \tag{22}$$

where γ is a nonnegative weight specifying the trade-off between the energy function and model complexity function. By replacing the ℓ_0 norm by the ℓ_1 norm (Julius et al., 2009), the model selection problem is considered to find $\boldsymbol{k}$ that maximizes the objective function

$$\max_{\boldsymbol{k}} L(\left\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\right\} | \boldsymbol{k}) - \gamma g(\|k\|_1) \tag{23}$$

An optimal $\boldsymbol{k}$ for the above can alternatively be found by solving a more numerically convenient optimization by exploiting monotonicity of the logarithm. In addition, constraints for the parameter matrix from prior knowledge can be included. The final constrained optimization is formulated by

$$\min_{\boldsymbol{k}} -\log L(\left\{\hat{\boldsymbol{x}}_j(t_k) : j = 1, \ldots, r, k = 1, \ldots, m\right\} | \boldsymbol{k}) + \gamma\|k\|_1 \text{ s.t. } \boldsymbol{k} \in \mathcal{F} \tag{24}$$

where $\mathcal{F}$ denotes the structure constraint on the network available from prior knowledge. In this paper, the set $\mathcal{F}$ is defined as a non-negative set, $\boldsymbol{k} \geq 0$. No prior parametric or structural information is assumed to be provided. The multinomial distribution in the exponential family is log-concave (Lehmann and Romano, 2006), which implies that its negative logarithm is convex.

4. Simulation study and discussion

4.1. 3 Species reaction system

The proposed parameter estimation formulation is applied to a simple reaction system containing a population size of 100 for all 3 species. Potential trajectories are simulated by SSA and used in the proposed parameter estimation method. The sampling interval is 1 s and final simulation time is 50 s so that the population of species A–C at 51 time points are calculated in each SSA run. Species A and B have a reversible reaction connectivity and species A and C have an irreversible reaction connectivity as shown below,

$$B \leftrightarrows A \rightarrow C \tag{25}$$

The true parameter matrix, $\boldsymbol{K}_{\mathrm{true}} \in \mathbb{R}^{3\times 3}$, having zero and nonzero elements according to the connectivity of the system, is defined by

$$\boldsymbol{K}_{\mathrm{true}} = \begin{bmatrix} 0 & 0.2770 & 0.4 \\ 0.1667 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{26}$$

For measurable species A–C, a large number of simulation runs with different initial conditions and different time-profiles were generated by using different random seeds. In Fig. 2, four representative data sets are shown. The data for a population size of 10^6 are almost the same as the data from the deterministic simulation, as

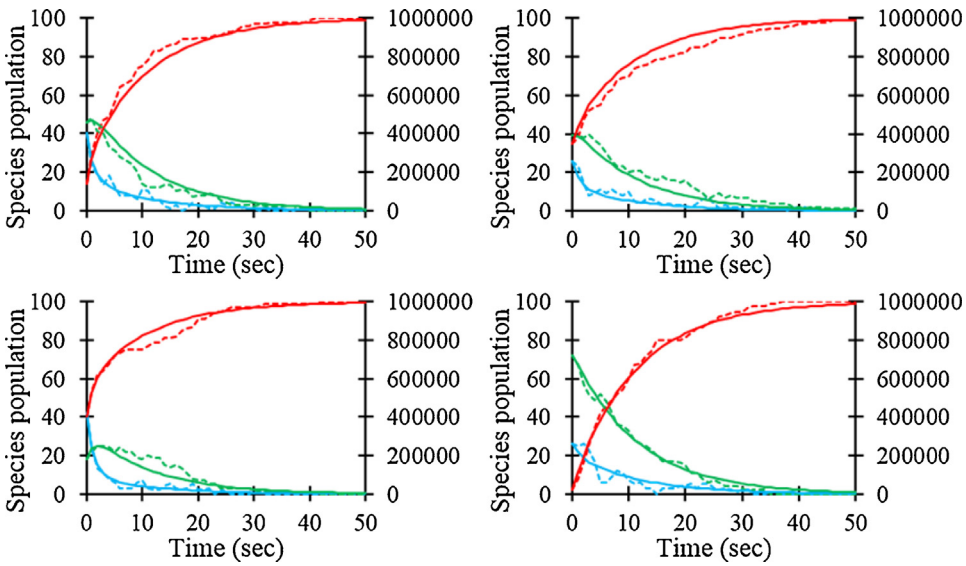

Fig. 2. Four representative sets of in silico data generated from SSA for 3-species reaction network system. Blue, green, and red lines indicate species A, B, and C respectively. Solid lines depict time-profiles for a population size of 10^6, which is near the deterministic limit. And dashed lines depict time-profiles for a population size of 100. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

stochastic fluctuations tend to average out over the large population size. On the other hand, for a population of 100, the data exhibit significant stochastic fluctuations relative to total population size.

In the parameter estimation, the actual connections among species A–C are assumed to be unknown, in which case the reaction parameter matrix, $\boldsymbol{K}$, to be estimated has the structure of

$$\boldsymbol{K} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} \rightarrow \begin{bmatrix} 0 & k_{12} & k_{13} \\ k_{21} & 0 & k_{23} \\ k_{31} & k_{32} & 0 \end{bmatrix} \tag{27}$$

The diagonal elements of the matrix,k_{ii}, are set to zero since those elements correspond to non-reactions that do not affect any of the governing reaction equations. This assumption makes the matrix, $\boldsymbol{K}$, stable since it satisfies the condition of Geršgorin theorem (Zavlanos et al., 2011). Even when the assumption of zero diagonal elements is relaxed, the stability constraints can be incorporated in the parameter estimation formulation.

4.1.1. Performance evaluation

With the generated dataset, the objective is to determine the connectivity among the species in the reactions and associated reaction rate constants using the MLE method. First, the performance of the MLE method without regularization is compared with the LSE method, which serves as a reference benchmark. The LSE method simply finds the parameter matrix $\boldsymbol{K}$ minimizing the sum of the squared errors between the stochastic data and predictions based on the deterministic model,

$$\min_{\boldsymbol{K}} \sum_{j=1}^{r} \sum_{k=1}^{m-1} \left(\hat{\boldsymbol{x}}_j(t_{k+1}) - e^{\delta t \boldsymbol{A}} \hat{\boldsymbol{x}}_j(t_k) \right)^2 \tag{28}$$

In the above, the prediction of current state based on previous measurement data and the kinetic matrix $\boldsymbol{A}$ is obtained by using the matrix differential Eq. (6). The exact discretization of the matrix differential equation yields the matrix exponential. It can be calculated by the forward Euler method or the bilinear transform based on an approximate discrete model with small time steps. The least-squares problem (28) is a multivariate nonlinear optimization problem for unknown matrix $\boldsymbol{K}$.

The constrained optimization for the MLE and the LSE were implemented in MATLAB. The interior-point algorithm in `fmincon` in Optimization Toolbox was used for the optimization. For a population size of 10^6, the estimated parameter matrix from LSE, averaged over the 10,000 datasets (one run in one dataset) was

$$\hat{\boldsymbol{K}}_{\text{LSE,Avg}} = \begin{bmatrix} 0 & 0.2780 & 0.3990 \\ 0.1667 & 0 & 0.0002 \\ 0.0000 & 0.0000 & 0 \end{bmatrix} \tag{29}$$

which is reasonably close to the true parameter matrix (26). On the other hand, for a population size of 100, the estimated parameter matrix from LSE, averaged over the 10,000 datasets was

$$\hat{\boldsymbol{K}}_{\text{LSE,Avg}} = \begin{bmatrix} 0 & 0.3919 & 0.3783 \\ 0.1923 & 0 & 0.0214 \\ 0.0006 & 0.0007 & 0 \end{bmatrix} \tag{30}$$

which is significantly different from the true parameter matrix. MSE (mean square error) over the 10,000 datasets was 2.44×10^{-2}. On the other hand, the averaged parameter matrix estimated from MLE with the same datasets without regularization was

$$\hat{\boldsymbol{K}}_{\text{MLE,Avg}} = \begin{bmatrix} 0 & 0.2927 & 0.3826 \\ 0.1649 & 0 & 0.0092 \\ 0.0000 & 0.0000 & 0 \end{bmatrix} \tag{31}$$

which is much closer to the true parameter matrix. MSE over the 10,000 datasets was 5.19×10^{-3}, which means the estimated parameter matrices from MLE were much more narrowly distributed around the true value than those from LSE. This example indicates the importance of using the stochastic model in parameter estimation in the case of a small population size. In the case of a large population size (e.g., 10^6), use of the deterministic model appears well justified.

For this stochastic case, the comparison of MSE averaged over 10,000 trials with different datasets might not be sufficient to judge the accuracy of the estimates from two estimation methods. Fig. 3 shows the asymptotical behavior of MSEs of MLE and LSE averaged over increasing number of trials from 1 to 50,000. The averaged MSEs of MLE have smaller bias than those of LSE.

Other key performance indices that can be used in comparing the results from LSE and exact MLE are correctness with respect to

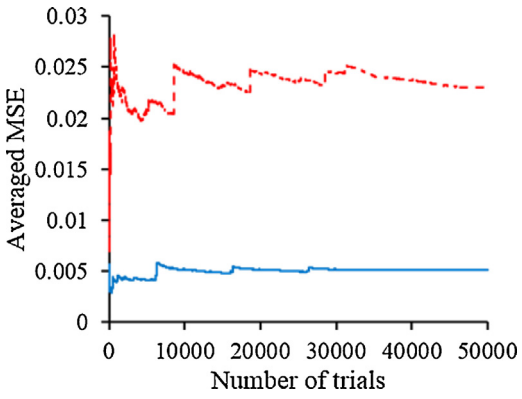

Fig. 3. Averaged MSE for increasing number of trials. Blue solid line depicts the averaged MSEs from MLE and red dashed line depicts the averaged MSEs from LSE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

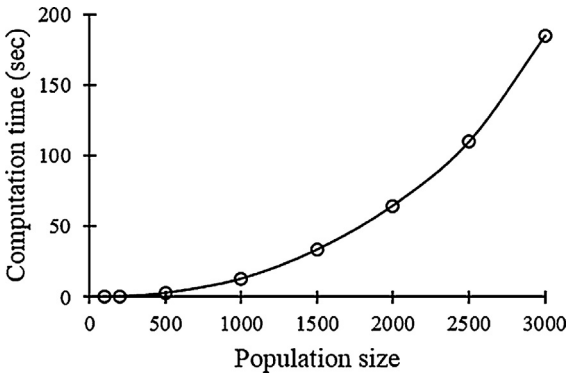

Fig. 4. Computation time per one likelihood calculation with increasing population size. Black circles indicate the calculated points and black line is smooth connection of circles.

Table 1
Two performance indices of $\hat{\boldsymbol{K}}$ from MLE and LSE for 3-species reaction system.

Method	MLE	LSE
P_{zero}[a]	1.22×10^{-4}	5.15×10^{-4}
$P_{\text{non-zero}}$[a]	1.02×10^{-2}	4.83×10^{-2}

[a] The values were averaged over 10,000 datasets for the 3-species reaction system with the population size of 100.

sparsity of the parameter matrix $\boldsymbol{K}$ and estimation of the parameter values k_{ij} for the connected reactions. Firstly, zero elements in an estimated matrix can be detected by using a tolerance criterion which defines an upper limit μ on the magnitude of the estimated value. For example, if the tolerance is set as $\mu = 0.1$, both LSE and MSE correctly identify the zero elements. However, if the tolerance is tightened to $\mu = 0.01$, LSE misses one zero element, k_{23}, while the MSE still correctly identifies all of the zero elements.

Alternatively, accuracy of zero parameters (indicating the sparsity detection) and accuracy of non-zero parameters can be quantified by the averaged MSE for the zero and non-zero elements of the parameter matrix over P runs,

$$P_{\text{zero}} = \sum_{i=1}^{P} \frac{\hat{K}_{31,i}^2 + \hat{K}_{32,i}^2 + \hat{K}_{23,i}^2}{3P} \tag{32}$$

$$P_{\text{non-zero}} = \sum_{i=1}^{P} \frac{\left(K_{21} - \hat{K}_{21,i}\right)^2 + \left(K_{12} - \hat{K}_{12,i}\right)^2 + \left(K_{13} - \hat{K}_{13,i}\right)^2}{3P} \tag{33}$$

In Table 1, the MLE shows lower values for both performance indices calculated by (32) and (33). That is, the proposed method shows better detection of the sparsity and more accurate estimation for the non-zero reaction rate constants even without regularization.

4.1.2. Effect of inputs on performance and computation time

In actual biological systems, availability of repeated runs under identical conditions is limited. The simulation study with a large number of datasets for the 3-species reaction system is somewhat artificial, but was useful for checking the basic performance of the estimation methods. If possible, it would be better to estimate the parameter matrix from a dataset containing multiple runs having different dynamics. Table 2 shows MSEs of the estimates from both the MLE and LSE methods with increasing number of runs in the dataset. The results indicate that accuracy of the estimated parameter matrix is improved when large datasets are provided for both methods. In addition, MLE provides more accurate estimates compared to LSE in all cases.

In Table 2, the averaged computation times of MLE are significantly longer than those of LSE. Almost 90% of the total computation time of MLE is taken up by formulating objective function (19) which includes generating all possible state configurations (`meshgrid`), computing multinomial PDFs (`mnpdf`) and multi-dimensional convolutions (`convn`). The call number of functions, `meshgrid`, `mnpdf`, and `convn` in the beginning of the optimization increases linearly with increasing number of runs (with a fixed number of samples with the each run), resulting in the linear dependency of total computation time with respect to dataset size. The total computation time related with the main parts of the optimization algorithm implementation, such as the updating of the Jacobian and Hessian matrix, and the line search, is less than 0.1 s. This article considers the information-limited case (1 run in the dataset) where the benefit of using a stochastic model in the estimation formulation is maximized in terms of both accuracy and computation time. Indeed, the information limited case reflects the real experimental situation better.

Optimal estimation based on a model that tracks probabilities over the discrete state space is computationally expensive when the population size is large. Fig. 4 shows the computation time vs. population size with a 3.40 GHz CPU computing machine. The computation time is exponentially increasing with growing population size. The computation time is based on just a single calculation of likelihood function (19) given a parameter set.

Invoking the central limit theorem, we may hypothesize that, with a large population size the actual distribution of states is closely approximated by a normal distribution. A general rule is that the approximation is adequate as long as the population size is larger than 30 (Hayter, 2012). The result (29) for a population size of 10^6 also showed that the least-squares estimation based on the mass–average model assuming normal distributions provide a sufficiently accurate solution within very short computation time (<1 s). In addition to this, "stochasticity" (the degree of randomness or noise variance) can be measured by a normalized root mean squares error (RMSE) of the stochastic data from the deterministic data,

$$\text{Stochasticity} = \sqrt{\sum_{j=1}^{r}\sum_{k=2}^{m} \frac{\left(\left(\hat{\boldsymbol{x}}_j(t_k) - \tilde{\boldsymbol{x}}_j(t_k)\right)/\tilde{\boldsymbol{x}}_j(t_k)\right)^2}{r(m-1)}} \tag{34}$$

where the stochastic data $\hat{\boldsymbol{x}}_j(t_k)$ are generated from the SSA and the deterministic data $\tilde{\boldsymbol{x}}_j(t_k)$ are obtained from Eq. (9). Fig. 5 shows that the stochasticity rapidly drops in the earlier range of population size from 100 to 1000. For a population size of 1000 or greater, LSE is clearly a better choice in terms of computation time and accuracy.

Table 2
Three performance indices of $\hat{\boldsymbol{K}}$ and computation time from MLE and LSE with multiple runs for 3-species reaction system.

Size of dataset		1-run	5-run	10-run	20-run
MSE[a]	MLE	3.06×10^{-3}	6.69×10^{-4}	2.74×10^{-4}	1.34×10^{-4}
	LSE	3.06×10^{-2}	1.57×10^{-3}	5.62×10^{-4}	3.27×10^{-4}
P_{zero}[a]	MLE	1.15×10^{-4}	3.09×10^{-5}	1.04×10^{-5}	4.41×10^{-6}
	LSE	6.88×10^{-4}	1.08×10^{-4}	4.49×10^{-5}	1.84×10^{-5}
$P_{non-zero}$[a]	MLE	6.01×10^{-3}	1.31×10^{-3}	5.37×10^{-4}	2.64×10^{-4}
	LSE	6.05×10^{-2}	3.02×10^{-3}	1.08×10^{-3}	6.35×10^{-4}
Computation time (s)[b]	MLE	46.64	176.57	363.16	758.64
	LSE	0.97	1.27	1.83	6.05

[a] The values were averaged over 100 datasets for the 3-species reaction system with the population size of 100.
[b] The computation time was recorded using a workstation with 3.40 GHz.

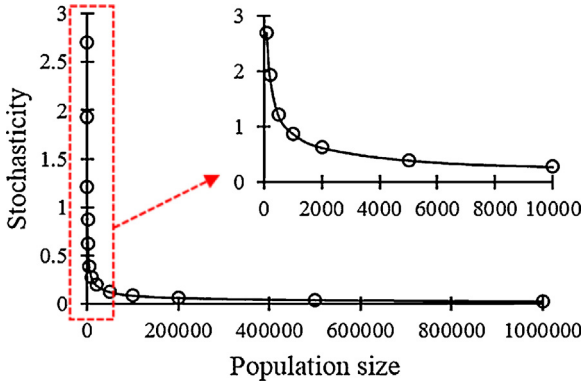

Fig. 5. Stochasticity of the SSA data with increasing population size. Black circles indicate the calculated points and black line is smooth connection of circles. Smaller plot depicts a zoom-in of the population size from 0 to 10,000.

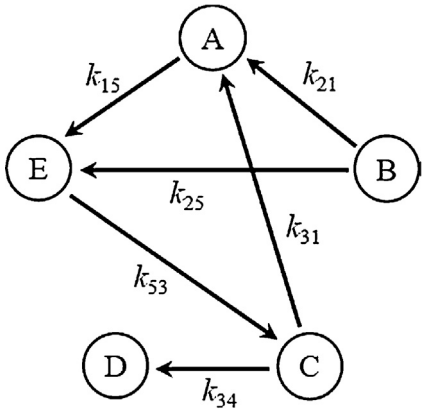

Fig. 6. A 5-species biochemical reaction network. 5 circles indicate species A–E, and each arrow indicates a monomolecular reaction with rate constants k_{ji} where i and j denote the ith species as a reactant and the jth species as a product. For example, k_{15} is a rate constant of a reaction from the first species, A, to the fifth species, E.

4.2. Published case of a 5-species reaction network

To evaluate the performance advantage of proposed method against more realistic biological problems, a reaction network system containing 5 species and 6 monomolecular reactions studied in August and Papachristodoulou, 2009 is considered with modification. To make the problem more challenging, the population size for all species in the system is assumed to be much smaller than the 3-species case, resulting in more stochastic dynamics (Elowitz et al., 2002). For this problem, the regularized MLE is used for reducing the model complexity.

Consider the biochemical reaction network depicted in Fig. 6 where the reaction kinetics involve $n = 5$ species $S = \{A–E\}$, and the associated reactions are monomolecular reactions. The true parameter matrix, $\boldsymbol{K}_{true} \in \mathbb{R}^{5\times5}$ is given as

$$\boldsymbol{K}_{true} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.8492 \\ 0.3386 & 0 & 0 & 0 & 0.4290 \\ 0.8244 & 0 & 0 & 0.0563 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7364 & 0 & 0 \end{bmatrix} \quad (35)$$

SSA simulations were used to generate data, in which the total population is 20 (Fig. 7). Total simulation time is 50 s with 1 s sampling time and the initial state is assumed to be uniformly distributed. The data from the SSA simulations are used in the regularized MLE to identify the network structure and the associated reaction parameters. The parameter matrix $\boldsymbol{K}$ to be estimated is

$$K = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} & k_{15} \\ k_{21} & k_{22} & k_{23} & k_{24} & k_{25} \\ k_{31} & k_{32} & k_{33} & k_{34} & k_{35} \\ k_{41} & k_{42} & k_{43} & k_{44} & k_{45} \\ k_{51} & k_{52} & k_{53} & k_{54} & k_{55} \end{bmatrix} \rightarrow \begin{bmatrix} 0 & k_{12} & k_{13} & k_{14} & k_{15} \\ k_{21} & 0 & k_{23} & k_{24} & k_{25} \\ k_{31} & k_{32} & 0 & k_{34} & k_{35} \\ k_{41} & k_{42} & k_{43} & 0 & k_{45} \\ k_{51} & k_{52} & k_{53} & k_{54} & 0 \end{bmatrix} \quad (36)$$

where the diagonal elements of the matrix,k_{ii}, are set to zero and matrix $\boldsymbol{K}$ satisfies the stability matrix condition (Zavlanos et al., 2011). The number of parameters to be estimated is 20.

The performance of MLE without regularization can be simply compared with conventional LSE through MSE of the estimated parameter matrices over 100 datasets. Table 3 shows the estimation performance including the accuracy of all parameters, zero and non-zero parameters, which are measured by MSEs for zero as well as non-zero elements of the parameter matrix, similarly in (32) and (33). Averaged MSEs from MLE are less than those from LSE, which means that the estimated parameter matrices from MLE are more narrowly distributed from the true parameter matrix than for LSE. Also, the results from MLE show better detection of the sparsity and more accurate estimation for the reaction rate constants.

The estimation errors for the 5-species reaction network system are larger than those for the 3-species reaction system, which can be attributed to the higher number of parameters to be estimated as well as the higher degree of stochasticity in the data. The level of model complexity should be controlled in estimating parameter matrix by adding a regularization term to the MLE formulation.

Table 3
Three performance indices of $\hat{\boldsymbol{K}}$ from MLE and LSE for 5-species reaction system.

Method	MLE	LSE
MSE[a]	4.09×10^{-2}	8.93×10^{-2}
P_{zero}[a]	9.64×10^{-2}	1.01×10^{-1}
$P_{non-zero}$[a]	1.71×10^{-2}	8.45×10^{-2}

[a] The values were averaged over 100 datasets for the 5-species reaction system with the population size of 20.

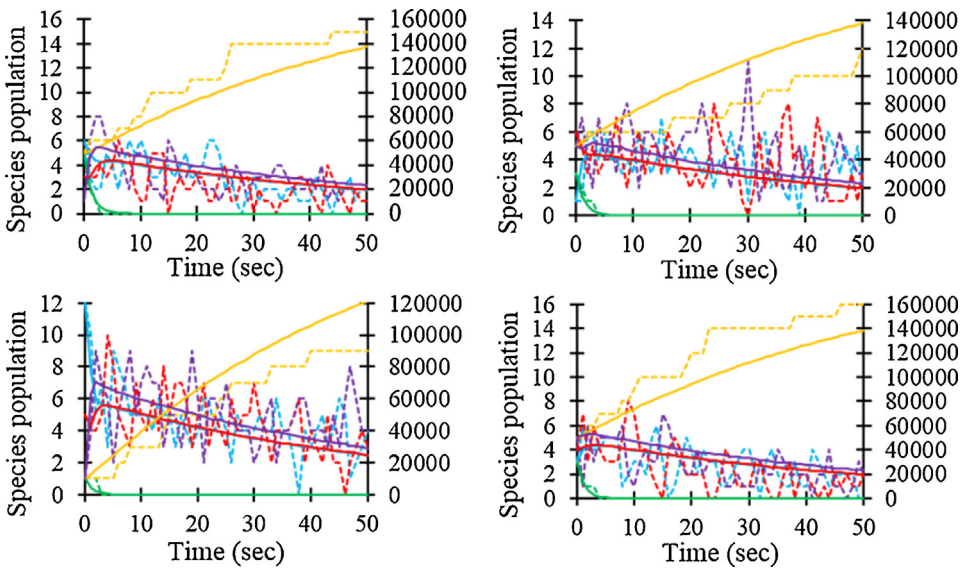

Fig. 7. Four representative runs generated from SSA for 5-species reaction network system. Blue, green, red, orange, and violet lines indicate species A, B, C, D, and E respectively. Solid lines depict time-profiles for the system having a population size of 2×10^5, which is near the deterministic limit. Dashed lines depict time-profiles for the system having a population size of 20. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Consider the estimation performance for an increasing γ in ℓ_1-regularized MLE of (24). With an appropriate value of $\gamma > 0$, some of the network connections are expected to disappear without significantly affecting the estimation performance. By following the recursive algorithms (August and Papachristodoulou, 2009; Julius et al., 2009), which generate a new sparse structure by repeating the same optimization with an updated sparse structure, the reaction parameter matrix can be further improved.

This approach is called the *non-weighted recursive algorithm* (August and Papachristodoulou, 2009). This two-step estimation process involves ℓ_1-regularized optimization followed by a recursive optimization. In each iteration of the ℓ_1-regularized optimization with a fixed γ, entries that are small (less than a threshold factor δ) are set to zero in the subsequent iterations. The recursive optimization is performed until estimated parameter matrix $\hat{\boldsymbol{K}}$ converges. The estimation results are truly dependent on the value of δ which can be tuned to adjust the degree of sparsity.

Three different parameter estimation methods are performed with the dataset containing a single run: (1) MLE, (2) ℓ_1-regularized exact MLE, and (3) ℓ_1-regularized exact MLE with non-weighted recursive algorithm. The datasets containing multiple runs are not tested here as they rarely occur in practice and the computational time of the MLE calculation grows rapidly.

As a reference, the sparse parameter estimation is started with the MLE in (24) with $\gamma = 0$, which gives

$$\boldsymbol{K}_{\mathrm{MLE},\gamma=0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.8913 \\ 0.3514 & 0 & 0 & 0 & 0.1671 \\ 0.5664 & 0 & 0 & 0.0339 & 0.1917 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1877 & 0 & 0.6354 & 0.0162 & 0 \end{bmatrix} \tag{37}$$

where eleven entries in the parameter matrix are detected as zero parameters. With $\gamma = 10$, the estimate becomes

$$\boldsymbol{K}_{\mathrm{MLE},\gamma=10} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.5653 \\ 0.2234 & 0 & 0 & 0 & 0.0774 \\ 0.4913 & 0 & 0 & 0.0304 & 0.0264 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0065 & 0 & 0.4479 & 0.0166 & 0 \end{bmatrix} \tag{38}$$

where two more entries, k_{51} and k_{35} shrink dramatically. When the weight parameter is further increased to $\gamma = 20$,

$$\boldsymbol{K}_{\mathrm{MLE},\gamma=20} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.4623 \\ 0.1540 & 0 & 0 & 0 & 0.0563 \\ 0.4065 & 0 & 0 & 0.0277 & 0.0160 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3732 & 0.0165 & 0 \end{bmatrix} \tag{39}$$

where k_{51} finally becomes zero and k_{35} further shrinks. In this manner, the relationship between increasing the γ value and the estimated parameter matrix can be observed. For larger γ values,

$$\boldsymbol{K}_{\mathrm{MLE},\gamma=50} = \begin{bmatrix} 0 & 0 & 0.0290 & 0.0035 & 0.2845 \\ 0.0760 & 0 & 0 & 0 & 0.0345 \\ 0.2588 & 0 & 0 & 0.0212 & 0.0322 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0159 & 0 & 0.2449 & 0.0136 & 0 \end{bmatrix} \tag{40}$$

$$\boldsymbol{K}_{\mathrm{MLE},\gamma=100} = \begin{bmatrix} 0 & 0 & 0.0475 & 0.0059 & 0.1601 \\ 0.0400 & 0 & 0 & 0 & 0.0217 \\ 0.1510 & 0 & 0 & 0.0155 & 0.0435 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0311 & 0 & 0.1483 & 0.0098 & 0 \end{bmatrix} \tag{41}$$

Beyond $\gamma = 50$, three wrong connections, k_{51}, k_{13}, and k_{14}, suddenly appear and the number of connections are increased. The effect of γ value also can be seen by comparing MSE, P_{zero}, and $P_{\mathrm{non-zero}}$ of the estimation results (Fig. 8). The MSE and $P_{\mathrm{non-zero}}$ have the lowest values when there is no regularization ($\gamma = 0$) in the parameter estimation. The estimated parameter matrix for $\gamma = 20$ shows the lowest P_{zero} value among the tried cases, as plotted in Fig. 8.

In practice, the MSE, P_{zero}, and $P_{\mathrm{non-zero}}$ cannot be used to choose a γ value because the true parameter matrix is not known. Instead, plot of the negative log-likelihood function term in (24) versus the ℓ_1-norm penalty term in (24) (Fig. 9). The negative log-likelihood function indicates the fitness of data for the model and the ℓ_1-norm penalty for the model parameter matrix indicates the model complexity. The curve shows that a γ value around 20 reduces the model complexity significantly with relatively a small compensation for the likelihood function in this case.

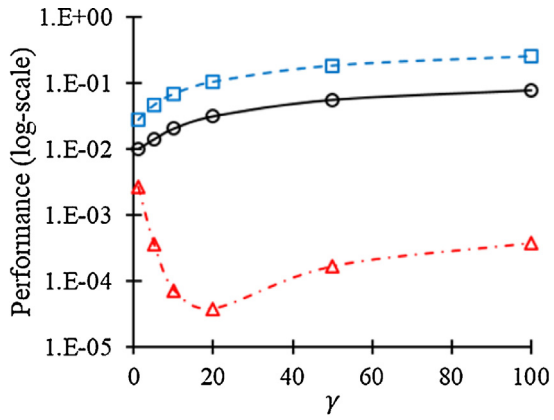

Fig. 8. A plot of three performance indices with increasing γ value. Black circles and solid line depict MSE. Red triangles and dash-dotted line depict P_{zero}. Blue squares and dashed line depict $P_{non-zero}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

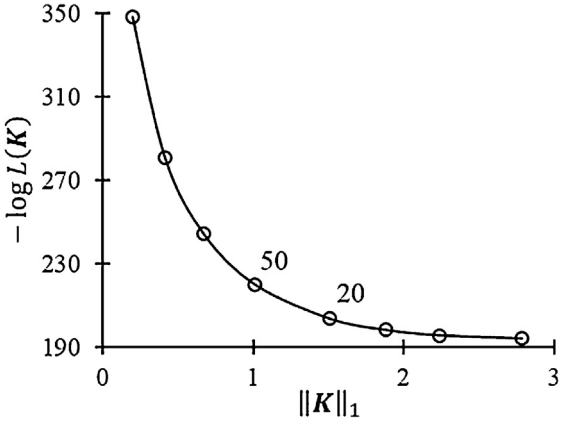

Fig. 9. The negative log-likelihood function vs. the ℓ_1-norm penalty for the model parameter matrix. Black circles indicate the calculated points and black line is smooth connection of circles.

After choosing a γ value ($\gamma = 20$), we can use a recursive algorithm, which repeats the regularized MLE calculations with the updated information about a parameter matrix structure. The estimated parameter matrix from the non-weighted recursive algorithm with $\delta = 0.02$ (which is slightly larger than the minimum value in (39) of 0.016) converges to

$$\boldsymbol{K}_{\mathrm{MLE},\gamma=20,\delta=0.02} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.4730 \\ 0.1541 & 0 & 0 & 0 & 0.0558 \\ 0.4164 & 0 & 0 & 0.0481 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3859 & 0 & 0 \end{bmatrix} \quad (42)$$

where the two additional entries, k_{54} and k_{35}, become zero parameters. Ultimately, 6 correct connections in the reaction network are detected. The regularized MLE and recursive algorithms were more effective in identifying the reaction topology, compared to the result of MLE of (37). For the fixed network structure in (42), the corresponding parameters can be estimated by using the MLE method without considering a regularization term, to give

$$\boldsymbol{K}_{\mathrm{MLE}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.8338 \\ 0.3980 & 0 & 0 & 0 & 0.1205 \\ 0.7438 & 0 & 0 & 0.0540 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6455 & 0 & 0 \end{bmatrix} \quad (43)$$

which is much closer to the true parameter matrix. Following the same procedure, the final structures and parameter values obtained with different datasets are

$$\boldsymbol{K}_{\mathrm{MLE}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.7823 \\ 0.3064 & 0 & 0 & 0 & 0.6269 \\ 0.7398 & 0 & 0 & 0.0716 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6363 & 0 & 0 \end{bmatrix} \quad (44)$$

$$\boldsymbol{K}_{\mathrm{MLE}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.9333 & 0 & 0 & 0 & 0.7707 \\ 0.7033 & 0 & 0 & 0.0716 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6069 & 0 & 0 \end{bmatrix} \quad (45)$$

$$\boldsymbol{K}_{\mathrm{MLE}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.9884 \\ 0.7692 & 0 & 0 & 0 & 0 \\ 0.8870 & 0 & 0 & 0.0909 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9068 & 0 & 0 \end{bmatrix} \quad (46)$$

$$\boldsymbol{K}_{\mathrm{MLE}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1.0619 \\ 0 & 0 & 0 & 0 & 0.8571 \\ 0.5436 & 0 & 0 & 0.0552 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.2112 & 0 & 0.5169 & 0 & 0 \end{bmatrix} \quad (47)$$

where (45)–(47) have some misses in connectivity determination. The connectivity is correct in the middle three columns for all of the matrices, and with nonzero values of similar values to the exact matrix. The connectivity is correct in the first column for all but one matrix, and one or both of the nonzero elements are correctly identified in the last column.

System identification of a biochemical system which typically has a complex reaction network structure and highly stochastic dynamics is a challenging problem. However, the exact MLE formulation, a regularization method, and a recursive algorithm can be combined to discover better solutions for such problems. The proposed strategy is computationally efficient for the reaction network system having a small population size (~100) and a small number of components (~5). Furthermore, many biological network systems such as metabolic networks or gene regulatory networks come with a priori known connectivity structures inferred from biological knowledge. This prior information about the network structure can be easily incorporated into the optimization formulation as constraints. This procedure would enable us to obtain more accurate solutions.

5. Conclusions

A regularized maximum likelihood estimation (MLE) method is presented for determining the interaction topology of a biochemical reaction network system. The regularized MLE method is formulated by combining a closed-form solution of the chemical master equations that describe stochastic monomolecular biochemical reaction systems with an ℓ_1-penalty term for the parameter matrix. Improved performance of the MLE method and recursive algorithms is demonstrated by using stochastic simulation data for 3- and 5-species monomolecular reaction network systems. The proposed method showed an improved ability to identify a sparse structure of the parameter matrix and estimate the associated reaction rate constants. The proposed method can potentially be used for robust reaction network identification prob-

lems such as found in metabolic networks or gene regulatory networks.

Acknowledgement

This work was supported by the Advanced Biomass R&D Center (ABC) of Global Frontier Project funded by the Ministry of Science, ICT and Future Planning (ABC-2011-0031354).

References

August, E., Papachristodoulou, A., 2009. Efficient, sparse biological network determination. BMC Syst. Biol. 3, 25.
Boyd, S.P., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.
Boys, R.J., Wilkinson, D.J., Kirkwood, T.B., 2008. Bayesian inference for a discretely observed stochastic kinetic model. Stat. Comput. 18, 125–135.
Conradi, C., Saez-Rodriguez, J., Gilles, E.-D., Raisch, J., 2005. Using chemical reaction network theory to discard a kinetic mechanism hypothesis. IEE Proc. Syst. Biol. 152, 243–248.
Craciun, G., Pantea, C., 2008. Identifiability of chemical reaction networks. J. Math. Chem. 44, 244–259.
Daigle, B.J., Roh, M.K., Petzold, L.R., Niemi, J., 2012. Accelerated maximum likelihood parameter estimation for stochastic biochemical systems. BMC Bioinform. 13, 68.
Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. Science 297 (5584), 1183–1186.
Feinberg, M. (1979). Lectures on chemical reaction networks. Notes of lectures given at the Mathematics Research Center of the University of Wisconsin in 1979.
Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., Palsson, B.Ø., 2008. Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7, 129–143.
Fichthorn, K.A., Weinberg, W.H., 1991. Theoretical foundations of dynamical Monte Carlo simulations. J. Chem. Phys. 95, 1090–1096.
Gadgil, C., Lee, C.H., Othmer, H.G., 2005. A stochastic analysis of first-order reaction networks. Bull. Math. Biol. 67, 901–946.
García-Parajó, M.F., Veerman, J.A., Bouwhuis, R., Vallée, R., vanHulst, N.F., 2001. Optical probing of single fluorescent molecules and proteins. ChemPhysChem 2, 347–360.
Gardner, T.S., Di Bernardo, D., Lorenz, D., Collins, J.J., 2003. Inferring genetic networks and identifying compound mode of action via expression profiling. Science 301 (5629), 102–105.
Gennemark, P., Wedelin, D., 2009. Benchmarks for identification of ordinary differential equations from time series data. Bioinformatics 25, 780–786.
Gibson, M.A., Bruck, J., 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A 104, 1876–1889.
Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.
Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. 58, 35–55.
Golightly, A., Wilkinson, D.J., 2011. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. Interface Focus 1, 807–820.
Hayter, A., 2012. Probability and Statistics for Engineers and Scientists. Cengage Learning.
Hesterberg, T., Choi, N.H., Meier, L., Fraley, C., 2008. Least angle and $\ell 1$ penalized regression: a review. Stat. Surv. 2, 61–93.
Hyduke, D.R., Palsson, B.Ø., 2010. Towards genome-scale signalling-network reconstructions. Nat. Rev. Genet. 11, 297–307.
Jahnke, T., Huisinga, W., 2007. Solving the chemical master equation for monomolecular reaction systems analytically. J. Math. Biol. 54, 1–26.
Jefferys, W.H., Berger, J.O., 1991. Sharpening Ockham's Razor on a Bayesian strop. Dept. Statistics, Purdue Univ., West Lafayette, Indiana, Tech. Rep.
Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.-L., 2000. The large-scale organization of metabolic networks. Nature 407 (6804), 651–654.
Julius, A., Zavlanos, M., Boyd, S., Pappas, G.J., 2009. Genetic network identification using convex programming. IET Syst. Biol. 3, 155–166.
Kremling, A., Fischer, S., Gadkar, K., Doyle, F.J., Sauter, T., Bullinger, E., Allgöwer, F., Gilles, E.D., 2004. A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. Genome Res. 14, 1773–1785.
Lehmann, E.L., Romano, J.P., 2006. Testing Statistical Hypotheses. Springer.
Li, G.-W., Xie, X.S., 2011. Central dogma at the single-molecule level in living cells. Nature 475, 308–315.
Lillacci, G., Khammash, M., 2010. Parameter estimation and model selection in computational biology. PLoS Comput. Biol. 6, e1000696.
Lillacci, G., Khammash, M., 2013. The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations. Bioinformatics 29, 2311–2319.
MacNamara, S., Bersani, A.M., Burrage, K., Sidje, R.B., 2008. Stochastic chemical kinetics and the total quasi-steady-state assumption: application to the stochastic simulation algorithm and chemical master equation. J. Chem. Phys. 129, 095105.
Munsky, B., Khammash, M., 2006. The finite state projection algorithm for the solution of the chemical master equation. J. Chem. Phys. 124, 044104.
Munsky, B., Trinh, B., Khammash, M., 2009. Listening to the noise: random fluctuations reveal gene network parameters. Mol. Syst. Biol. 5, 318.
Munsky, B., Neuert, G., van Oudenaarden, A., 2012. Using gene expression noise to understand gene regulation. Science 336 (6078), 183–187.
Poovathingal, S.K., Gunawan, R., 2010. Global parameter estimation methods for stochastic biochemical systems. BMC Bioinform. 11, 414.
Radulescu, O., Gorban, A.N., Zinovyev, A., Noel, V., 2012. Reduction of dynamical biochemical reaction networks in computational biology. Front. Genet. 3, 131.
Raj, A., van Oudenaarden, A., 2009. Single-molecule approaches to stochastic gene expression. Annu. Rev. Biophys. 38, 255–270.
Raj, A., Rifkin, S.A., Andersen, E., van Oudenaarden, A., 2010. Variability in gene expression underlies incomplete penetrance. Nature 463 (7283), 913–918.
Schreiber, S.L., 2000. Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287 (5460), 1964–1969.
Tian, T., Xu, S., Gao, J., Burrage, K., 2007. Simulated maximum likelihood method for estimating kinetic rates in gene expression. Bioinformatics 23, 84–91.
Xie, X.S., Choi, P.J., Li, G.-W., Lee, N.K., Lia, G., 2008. Single-molecule approach to molecular biology in living bacterial cells. Annu. Rev. Biophys. 37, 417–444.
Yu, J., Xiao, J., Ren, X., Lao, K., Xie, X.S., 2006. Probing gene expression in live cells, one protein molecule at a time. Science 311 (5767), 1600–1603.
Zavlanos, M.M., Julius, A.A., Boyd, S.P., Pappas, G.J., 2011. Inferring stable genetic networks from steady-state data. Automatica 47, 1113–1122.
Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., Koeppl, H., 2012. Moment-based inference predicts bimodality in transient gene expression. Proc. Natl. Acad. Sci. U. S. A. 109, 8340–8345.