2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Efficient Parameter Identification for Stochastic Biochemical Networks Using a Reduced-order Realization

Yutaka Hori, Mustafa H. Khammash and Shinji Hara

Abstract— **In this paper, we propose a parameter identification method for stochastic biochemical reaction networks using flow cytometry data. A distinctive feature of the proposed method is that it is computationally efficient compared to existing works, thus it is applicable to complex biochemical networks. To this end, we first show that it is possible to construct a significantly small-order realization of the stochastic biochemical system using flow cytometry measurements. Then, the small-order realization is utilized for the development of the efficient identification method. Finally, the proposed method is demonstrated with an existing biological example.**

I. Introduction

Recent progress of high-throughput experimental technique has enabled us to develop reliable dynamical models of biochemical processes based on a large amount of experimental data. Consequently, both stochastic and deterministic modeling frameworks were established and validated during the last decade. In the current systems biology, however, the determination of the rate parameters in such models still remains a challenging task.

The difficulty of the parameter identification mainly arises from the stochastic variability of the cellular processes and the limitation of available measurements. In fact, genetically identical cells can exhibit different dynamical behaviors both qualitatively and quantitatively [1]–[3]. This suggests that the time evolution of RNA and protein abundance can be more appropriately captured by the statistical distribution over many cells rather than a single trajectory of a particular cell. Such distribution measurements are now available by using flow cytometry, for example. However, the parameter identification methods based on the distribution measurements are not yet established to a satisfactory level, even though system identification is one of the actively studied areas in engineering. Moreover, it is often the case that the number of available temporal snapshots of the distribution is limited, and one is asked to estimate the parameters only from the transient or sparse time series data.

To date, most parameter identification methods of stochastic biochemical networks were developed based on statistical inference. One of the earliest attempts took a Bayesian approach for an approximated dynamics of stochastic chemical reactions [4]. This work was later extended for the exact stochastic dynamics [5], and more recently the Bayesian inference combined with the sequential Monte carlo method [6] was also shown to be useful. In another line of research, maximum likelihood estimation was explored in [7] and [8]. However, tuning and evaluation of these statistical-inference-based algorithms are not necessarily easy, since the structure of the dynamics does not explicitly appear in the derivation of the algorithms.

A more structured approach was taken in Munsky and Khammash [9], where the identification problem of stochastic biochemical reactions was formulated as an output error minimization of a linear time-invariant system. A key idea was to employ the finite state projection [10], with which the dynamics was written in the form of a finite dimensional linear time-invariant system. Thus, the identification problem became more tractable from a control theoretic viewpoint.

Despite these great efforts, a major drawback of the existing methods is the high computational cost, which mainly comes from the extensive simulations of the system for various parameters. In fact, these identification methods require several hours to days even for simple biochemical networks, which restricts their applicability. This, in turn, motivates us to develop a computationally efficient identificaiton framework.

The goal of this paper is to propose a parameter identification method of stochastic biochemical reactions using the distribution measurements obtained by flow cytometry. A distinctive feature of the proposed method is that it requires little computational time compared to the existing works. The key idea is to utilize a small-order linear system that can be computed from the measurements. Specifically, we show that the eigensystem realization algorithm [11], which is a generalization of the celebrated Ho-Kalman's algorithm [12], [13], can produce a small-order realization, while preserving essential information of the system. This realization is then shown to satisfy a certain Lyapunov equation. Consequently, the parameter identification problem is formulated as the root finding of the Lyapunov equation.

The organization of this paper is as follows. In section II, we introduce the dynamics of stochastic biochemical reactions and formulate the identification problem. Then, a rough idea of the algorithm is shown in Section III. In Section IV, we describe the realization algorithm and present key properties of the realization. The identification problem is then formulated as a root finding problem in Section V. In Section VI, the proposed method is demonstrated with a

This work was supported in part by Grant-in-Aid for JSPS Fellows of Japan Society for the Promotion of Science (JSPS) under grant No. 23-9203, Human Frontier Science Program under grant RGP0061/2011 and the National Science Foundation (NSF) under grant ECCS-0835847. Y. Hori and S. Hara are with the Department of Information Physics and Computing, The University of Tokyo, Tokyo 113-8656, Japan. {Yutaka_Hori, Shinji_Hara}@ipc.i.u-tokyo.ac.jp. M. H. Khammash is with Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland. mustafa.khammash@bsse.ethz.ch

genetic toggle switch [2]. Finally, Section VII concludes the paper.

II. Dynamics of stochastic chemical reactions and Problem statement

In this section, we first introduce the dynamical model of stochastic biochemical reactions, then we formulate the parameter identification problem considered in this paper.

A. Chemical master equation (CME)

Consider a set of chemical reactions that involves M molecular species $\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_M$ interacting via R reaction channels $\mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_R$. Let $\boldsymbol{x} \in \mathbb{N}_0^N$ denote the state vector defined by $\boldsymbol{x} := [x_1, x_2, \cdots, x_N]$, where $x_i \in \mathbb{N}_0$ is the number of the molecule $\mathcal{M}_i$ $(i = 1, 2, \cdots, M)$. In stochastic chemical reactions, the state vector $\boldsymbol{x}$ is the random variable, and the evolution of $\boldsymbol{x}$ follows the Markov process characterized by the chemical master equation (CME) [14]

$$\frac{\partial \mathbb{P}(\boldsymbol{x}, t)}{\partial t} = \sum_{i=1}^{R} w_i(\boldsymbol{x} - \boldsymbol{s}_i)\mathbb{P}(\boldsymbol{x} - \boldsymbol{s}_i, t) - w_i(\boldsymbol{x})\mathbb{P}(\boldsymbol{x}, t), \tag{1}$$

where $\mathbb{P}(\boldsymbol{x}, t)$ denotes the probability of the system being the state $\boldsymbol{x}$ at time t [1]. The vector $\boldsymbol{s}_i \in \mathbb{Z}^M$ is the stoichiometry of the reaction $\mathcal{R}_i$, and $w_i(\cdot) : \mathbb{N}_0^n \to \mathbb{R}_+ (i = 1, 2, \cdots, R)$ is the corresponding propensity function, which determines the rate of the reaction $\mathcal{R}_i$ $(i = 1, 2, \cdots, R)$. In other words, the number of molecules is changed by $\boldsymbol{s}_i$ by the reaction $\mathcal{R}_i$, and $\mathcal{R}_i$ fires with the probability $w_i(\boldsymbol{x})dt$ for the infinitesimal time dt.

B. Dynamics of the probability distribution

Let a scalar value $p_{\boldsymbol{x}}(t) \in [0, 1]$ indexed by each state $\boldsymbol{x}$ be defined by $p_{\boldsymbol{x}}(t) := \mathbb{P}(\boldsymbol{x}, t)$. Then, the probability distribution $\mathbb{P}(\boldsymbol{x}, t)$ can be equivalently expressed by the vector form as $\boldsymbol{p}_\infty(t) := [p_{\boldsymbol{x}_1}(t), p_{\boldsymbol{x}_2}(t), \cdots]$. The vector $\boldsymbol{p}_\infty(t)$ is (countably) infinite dimension since the index $\boldsymbol{x}$ represents the number of molecules that counts up to infinity. The chemical master equation (1) is then written as

$$\dot{\boldsymbol{p}}_\infty(t) := \mathcal{F}(\boldsymbol{\theta})\boldsymbol{p}_\infty(t), \tag{2}$$

where the infinite dimensional matrix $\mathcal{F}(\boldsymbol{\theta})$ is the infinitesimal generator that embodies the propensity functions $w_i(\cdot)$, and $\boldsymbol{\theta} := [\theta_1, \theta_2, \cdots \theta_L] \in \mathbb{R}_+^L$ is the reaction rate constants in the propensity functions.

In most cases, the exact solution of (2) is hard to obtain due to the infinite dimensionality. However, we can approximately solve (2) by employing the finite state projection (FSP) [10]. Let $\boldsymbol{p}(t) \in [0, 1]^n$ denote a n dimensional subvector of $\boldsymbol{p}_\infty(t)$ that is constructed by truncating all the other entries of $\boldsymbol{p}_\infty(t)$ except the n entries. Accordingly, the n by n matrix $F(\boldsymbol{\theta}) \in \mathbb{R}^{n \times n}$ is defined as the submatrix of $\mathcal{F}(\boldsymbol{\theta})$. Then, the solution $\boldsymbol{p}(t)$ of the ordinary differential equation

$$\dot{\boldsymbol{p}}(t) = F(\boldsymbol{\theta})\boldsymbol{p}(t) \tag{3}$$

approximates $\boldsymbol{p}_\infty(t)$.

The idea behind FSP is that the number of molecules x_i $(i = 1, 2, \cdots, N)$ inside a cell is normally small, thus truncating the subspace of $\boldsymbol{p}_\infty(t)$ with large x_i's merely affects the overall dynamics. In general, the approximation error can be arbitrarily small by increasing the order of the system (3), and prescribed approximation accuracy can always be satisfied. In particular, the upper bound of the approximation error was analytically obtained in [10].

In what follows, we assume that the order of the system (3), n, is sufficiently large such that the approximation error is negligible for the time interval of our interest, then we derive an identification algorithm based on the finite dimensional model (3).

C. Problem statement

In this paper, we propose a method that identifies the parameters $\boldsymbol{\theta}$ from measured distribution of the molecular copy numbers, *i.e.,* $\boldsymbol{p}(t)$. The distribution data can be obtained by flow cytometry as fluorescence intensity of single cells. In other words, a snapshot of $\boldsymbol{p}(t)$ can be measured for each scan of flow cytometer.

We assume that the measurements are obtained at $N + 1$ discrete time points with a constant sampling interval T_s. It should be noted that, in practice, N is much smaller than the order of the system n due to technical limitation. Consequently, we can construct the data matrix

$$P_{0:N} = [\boldsymbol{p}[0], \boldsymbol{p}[1], \cdots, \boldsymbol{p}[N]] \in \mathbb{R}^{n \times (N+1)}, \tag{4}$$

where $\boldsymbol{p}[k] := \boldsymbol{p}(kT_s)$ $(k = 0, 1, 2, \cdots)$ and $P_{i:j} := [\boldsymbol{p}[i], \boldsymbol{p}[i+1], \cdots, \boldsymbol{p}[j]] \in \mathbb{R}^{n \times (j-i+1)}$ $(i \leq j)$.

The parameter identification problem is then stated as follows.

Problem. *Given the measurements $P_{0:N}$ and the chemical master equation (3), identify the parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_L] \in \mathbb{R}_+^L$.*

For later convenience, we define the discrete-time system associated with (3) as

$$\boldsymbol{p}[k+1] = A(\boldsymbol{\theta})\boldsymbol{p}[k], \tag{5}$$

where $A(\boldsymbol{\theta}) \in \mathbb{R}^{n \times n}$

$$A(\boldsymbol{\theta}) := e^{F(\boldsymbol{\theta})T_s} \tag{6}$$

It should be noted that experimentally measured distributions are not necessarily identical to $\boldsymbol{p}[k]$ in practical experiments due to the noise from the physical environment and the difference between the empirical and the ideal distributions. In what follows, we first develop the identification method, assuming the ideal case, where the measurements are not subject to extrinsic noise. The effect of extrinsic noise is then examined in Section VI.

[1] More precisely, $\mathbb{P}(\boldsymbol{x}, t)$ should be defined as $\mathbb{P}(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0)$ with the initial state $\boldsymbol{x}_0$ and time t_0. In this paper, however, we omit the condition and simply write $\mathbb{P}(\boldsymbol{x}, t)$ to avoid notational complexity.

Algorithm 1 Overview of the proposed method

Input: Measurements $P_{0:N}$ and $F(\boldsymbol{\theta})$

Output: Identified parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_L]$

1) Compute the r-th order realization (7) and the projection matrices (Ψ, Φ) from $P_{0:N}$ (see Section IV).
2) Identify $\boldsymbol{\theta}$ by searching the solution of the Lyapunov equation $\mathcal{L}(\boldsymbol{\theta})$. (see Section V).

III. Overview of the proposed method

The major difficulty of the identification problem is twofold. First, the order of the system (5) can be very large, thus simulations and mathematical operations of the model require large computational cost. Suppose, for example, there are $M = 3$ species of molecules, and the number of molecules of each species is far smaller than 49 for the time interval of our interest. Then, the order of the model (5) would be $n = 50^3$. Secondly, the number of measured distribution, N, is often small, thus the identification needs to be done based on limited observations.

These difficulties motivate us to develop a novel parameter identification framework that uses only a small-order system instead of (5) and a small number of measurements. The rest of this section is devoted to describing the rough idea.

The proposed algorithm consists of two parts: (i) construction of a small-order realization and (ii) parameter identification based on the realization. In the first part, we construct a r-th order realization whose output matches the measurement $P_{0:N-1}$. We denote the r-th order realization by

$$\begin{aligned} \boldsymbol{q}[k+1] &= A_r \boldsymbol{q}[k], \\ \tilde{\boldsymbol{p}}[k] &= C_r \boldsymbol{q}[k], \end{aligned} \tag{7}$$

where $\boldsymbol{q}[k] \in \mathbb{R}^r, A_r \in \mathbb{R}^{r\times r}$ and $C_r \in \mathbb{R}^{n\times r}$. Then, this realization satisfies $\tilde{\boldsymbol{p}}[k] = \boldsymbol{p}[k]$ $(k = 0, 1, 2, \cdots, N-1)$ and $r \leq N (\ll n)$. Thus, the small-order realization (7) contains essential information of the dynamics (5). In particular, it is shown in Section IV that the systems (7) can be related to (5) by the coordinate projection $\Psi \in \mathbb{R}^{r\times n}$ and $\Phi \in \mathbb{R}^{n\times r}$ as

$$A_r = \Psi A(\boldsymbol{\theta}^\star)\Phi^T, C_r = \Phi^T, \boldsymbol{q}[0] = \Psi \boldsymbol{p}[0], \tag{8}$$

where $\boldsymbol{\theta}^\star$ stands for the actual parameter of the system.

As a result, it is possible to identify the parameters based on (7). Specifically, we first derive an algebraic equation that the system (7) should satisfy. Then, the parameter identification problem can be reduced to finding $\boldsymbol{\theta}$ that satisfies the algebraic equation.

The overall algorithm is summarized in Algorithm 1. A distinctive feature of the proposed algorithm is that it does not require complex mathematical operations of the original large order system (5), thus it is computationally efficient. This feature will be explained with an illustrative example in Section VI. In the following sections, we describe the details of the algorithm.

IV. Realization algorithm

In this section, we first describe the realization algorithm, then we show two distinctive properties of the r-th order realization.

Given the measurement data matrix $P_{0:N}$, the r-th order realization (7) is constructed by the following Algorithm 2.

Algorithm 2 Realization algorithm

Input: $P_{0:N}$

Output: The r-th order realization (7), $\boldsymbol{q}[0], \Psi$ and Φ

1) Compute the singular value decomposition (SVD) of $P_{0:N-1} \in \mathbb{R}^{n\times N}$

$$P_{0:N-1} = [U,\ \bar{U}] \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} \begin{bmatrix} V^T \\ \bar{V}^T \end{bmatrix}, \tag{9}$$

where $\Sigma := \mathrm{diag}(\sigma_1, \sigma_2, \cdots \sigma_r) \in \mathbb{R}^{r\times r}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$, and $U, \bar{U}, V$ and $\bar{V}$ are defined accordingly.

2) Define $\Psi \in \mathbb{R}^{r\times n}$ and $\Phi^T \in \mathbb{R}^{n\times r}$ as

$$\Psi := \Sigma^{-\frac{1}{2}} U^T \in \mathbb{R}^{r\times n}, \quad \Phi^T := U\Sigma^{\frac{1}{2}} \in \mathbb{R}^{n\times r}. \tag{10}$$

3) Obtain $A_r \in \mathbb{R}^{r\times r}, C_r \in \mathbb{R}^{n\times r}$ and $\boldsymbol{q}[0] \in \mathbb{R}^r$ by

$$\begin{aligned} A_r &:= (\Sigma^{-\frac{1}{2}} U^T P_{1:N})(P_{0:N-1}^T U \Sigma^{-\frac{3}{2}}) \\ C_r &:= \Phi^T,\ \boldsymbol{q}[0] := \Psi \boldsymbol{p}[0]. \end{aligned} \tag{11}$$

Note that the number of measurement time points N is assumed to be much smaller than the order of the system (5), *i.e.,* $N \ll n$, thus the data matrix $P_{0:N}$ has a tall thin shape.

Algorithm 2 is a special case of the eigensystem realization algorithm (ERA) [11], which is a generalization of the celebrated Ho and Kalman's realization method [12]. It is known that these algorithms produce the minimum realization if the data matrix $P_{0:N-1}$ has sufficiently large number of columns that spans the entire controllability subspace of (5), which is satisfied if $N \geq n$ [11]. In this paper, however, we are interested in the case where $N \ll n$, for which the properties of the realization (7) is not necessarily clear. Hence, we hereafter study the properties of the realization (7) under the assumption that N is small such that the column space of $P_{0:N-1}$ does not span the entire controllability subspace of (5).

Remark 1. In practice, the singular value of $P_{0:N-1}$ does not necessarily decay to zero due to numerical error and extrinsic noise added to $\boldsymbol{p}[k]$ $(k = 0, 1, 2, \cdots)$. In such cases, we would choose r so that the truncated singular values are sufficiently small compared to those in Σ. In particular, when the additive extrinsic noise is white, the singular values present a characteristic decay pattern, and we can reasonably determine r as shown in Section 6.7 of [13]. □

A distinctive feature of Algorithm 2 is that it produces a significantly small-order realization, because r is determined

from the SVD of (9), where $r \leq N(\ll n)$ holds. In addition, we can verify from $P_{1:N} = A(\boldsymbol{\theta}^\star)P_{0:N-1}$ and (9) that the matrices Ψ and Φ computed in the algorithm characterize the relation between the two realizations (5) and (7) as shown in (8). These features allow us to derive a computationally efficient algorithm for parameter identification as seen below.

When $N \ll n$, it is possible to show that the realization (7) has the following properties: (i) the output of (7) matches the measurement up to the first N time points and (ii) the time-limited gramian is given by Σ in (9). In what follows, we provide mathematical statements of these properties.

Proposition 1. *Consider the r-th order realization (7) obtained by Algorithm 2. Then, $\tilde{\boldsymbol{p}}[i] = \boldsymbol{p}[i]$ holds for $i = 0, 1, 2, \cdots, N-1$.*

Unlike the standard ERA [11], $\tilde{\boldsymbol{p}}[i]$ does not match $\boldsymbol{p}[i]$ for all i, but it does match for the finite time points. Moreover, the realization satisfies the following Lyapunov equation.

Proposition 2. *Consider the r-th order realization (7) obtained by Algorithm 2. Then, $X = \Sigma$ is a solution of the Lyapunov equation*

$$A_r X A_r^T - X - \boldsymbol{q}[N]\boldsymbol{q}^T[N] + \boldsymbol{q}[0]\boldsymbol{q}^T[0] = 0, \qquad (12)$$

where Σ is defined in (9).

Proposition 2 implies that the Lyapunov equation associated with the realization has the diagonal solution Σ, which is the singular value of $P_{0:N-1}$ computed in (9).

Remark 2. We can verify that the solution $X = \Sigma$ is the time limited gramian $X = \sum_{k=0}^{N-1} \boldsymbol{q}[k]\boldsymbol{q}^T[k]$ of (7) (see also Section 7.6.3 of [15]). This implies that the error between $\tilde{\boldsymbol{p}}[i]$ and $\boldsymbol{p}[i]$ depends on the sum of the truncated singular values, as is the case with the standard balanced truncation [13]. In particular, the error between $\tilde{\boldsymbol{p}}[i]$ and $\boldsymbol{p}[i]$, which is the output of the model obtained by FSP [10], can be arbitrarily small by increasing the dimension of the reduced order model r to $N(\ll n)$. □

V. Parameter identification algorithm based on the small-order realization

In this section, we describe a parameter estimation method that utilizes the realization obtained in the previous section.

Let $\mathcal{L}(\boldsymbol{\theta}) \in \mathbb{R}^{r\times r}$ be defined as

$$\mathcal{L}(\boldsymbol{\theta}) := (\Psi A(\boldsymbol{\theta})\Phi^T)\Sigma(\Psi A(\boldsymbol{\theta})\Phi^T)^T - \Sigma + Q, \qquad (13)$$

where $Q := -\boldsymbol{q}[N]\boldsymbol{q}^T[N] + \boldsymbol{q}[0]\boldsymbol{q}^T[0]$. Then, we see from (8) that the Lyapunov equation (12) can be equivalently written as $\mathcal{L}(\boldsymbol{\theta}^\star) = 0$. This implies that $\boldsymbol{\theta}^\star$ is a solution of the equation $\mathcal{L}(\boldsymbol{\theta}) = 0$. Therefore, the parameter identification problem can be recast as the root finding of $\mathcal{L}(\boldsymbol{\theta}) = 0$ for given Ψ, Φ, Σ and Q, which are obtained from Algorithm 2.

We here formulate the root finding problem as an optimization problem. Although a direct approach would be to solve $\min_{\boldsymbol{\theta}} \|\mathcal{L}(\boldsymbol{\theta})\|$, the computation of $A(\boldsymbol{\theta})$ in $\mathcal{L}(\boldsymbol{\theta})$ requires manipulation of the large matrix $F(\boldsymbol{\theta})$, since $A(\boldsymbol{\theta})$ is defined by the matrix exponential of $F(\boldsymbol{\theta})$ as shown in (6). Hence, we here define $K(\boldsymbol{\theta}) \in \mathbb{R}^{r\times r}$ to approximate $\Psi A(\boldsymbol{\theta})\Phi^T$ as

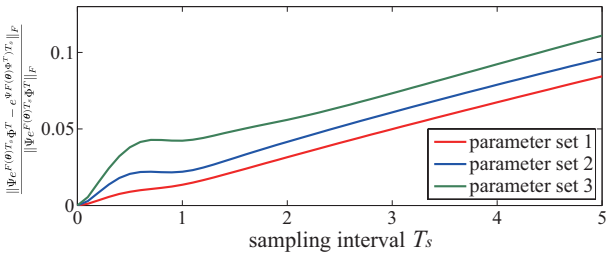

Fig. 1. Approximation error defined by (16) in terms of T_s. The matrix exponentials $\Psi e^{F(\boldsymbol{\theta})T_s}\Phi^T$ and $e^{(\Psi F(\boldsymbol{\theta})\Phi^T)T_s}$ are computed with Ψ and Φ obtained by Algorithm 2 with $T_s = 1$ and $r = 16$.

$$\Psi A(\boldsymbol{\theta})\Phi^T = \Psi e^{F(\boldsymbol{\theta})T_s}\Phi^T \simeq e^{(\Psi F(\boldsymbol{\theta})\Phi^T)T_s} =: K(\boldsymbol{\theta}). \qquad (14)$$

Note that $\Psi F(\boldsymbol{\theta})\Phi^T$ is a r by r matrix, thus the computational cost of the matrix exponential $K(\boldsymbol{\theta}) = e^{(\Psi F(\boldsymbol{\theta})\Phi^T)T_s}$ is much smaller than that of $\Psi e^{F(\boldsymbol{\theta})T_s}\Phi^T$. Then, the optimization problem is obtained as follows.

Optimization problem. *Given $F(\boldsymbol{\theta}), \Psi, \Phi, \Sigma$ and Q,*

$$\min_{\boldsymbol{\theta}} \|K(\boldsymbol{\theta})\Sigma K(\boldsymbol{\theta})^T - \Sigma + Q\|_2 \qquad (15)$$

subject to $\boldsymbol{\theta} \in [\underline{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}]$, where $K(\boldsymbol{\theta}) := e^{(\Psi F(\boldsymbol{\theta})\Phi^T)T_s}$, and $\bar{\boldsymbol{\theta}}$ and $\underline{\boldsymbol{\theta}} \in \mathbb{R}_+^L$ are given upper and lower bounds of the parameters, respectively.

This problem can be solved with nonlinear programming (NLP) solvers.

Remark 3. We note that the heavy computational burden in the previous works [6]–[9] came from the large number of stochastic simulations and/or the computation of the matrix exponential $e^{F(\boldsymbol{\theta})t}$, where the size of $F(\boldsymbol{\theta})$, or n, grows exponentially with respect to the number of molecules M (see also the first paragraph of Section III). On the other hand, the computationally dominant part of the proposed algorithm is the evaluation of the objective function (15) in the optimization. This calculation, however, involves only the small matrices of the size r by r. In particular, it follows that $r < N \ll n$, and the number of measurements N is independent of n. Thus, the proposed algorithm runs orders of magnitudes faster than the previous works. It should be noted that $F(\boldsymbol{\theta})$ is a sparse matrix, thus the multiplication $\Psi F(\boldsymbol{\theta})\Phi^T$ is fast despite the exponential growth of the size of $F(\boldsymbol{\theta})$. □

The approximation (14) poses an additional constraint to the sampling interval T_s. We see that $\Psi A(\boldsymbol{\theta})\Phi^T$ is approximated with small error when the sampling interval T_s is small, since

$$\Psi e^{F(\boldsymbol{\theta})T_s}\Phi^T = I_r + \Psi F(\boldsymbol{\theta})\Phi^T T_s + \sum_{i=2}^{\infty} \frac{1}{i!}\Psi F^i(\boldsymbol{\theta})\Phi^T T_s^i,$$

$$e^{(\Psi F(\boldsymbol{\theta})\Phi^T)T_s} = I_r + \Psi F(\boldsymbol{\theta})\Phi^T T_s + \sum_{i=2}^{\infty} \frac{1}{i!}(\Psi F(\boldsymbol{\theta})\Phi^T)^i T_s^i,$$

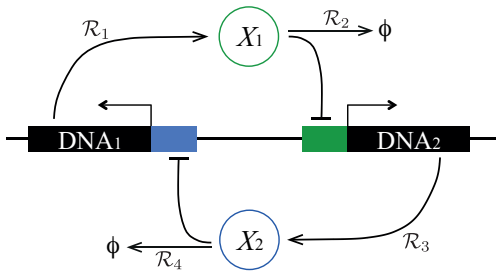

Fig. 2. Schematic diagram of the genetic toggle switch

TABLE I
REACTIONS IN GENETIC TOGGLE SWITCH

	Reaction	Propensity $w_i(\cdot)$
$\mathcal{R}_1$	$\mathrm{DNA}_1 + X_2 \to X_1 + X_2$	$\dfrac{\beta_1}{1+(x_2/K_1)^{\nu_1}}$
$\mathcal{R}_2$	$X_1 \to \phi$	$d_1 x_1$
$\mathcal{R}_3$	$\mathrm{DNA}_2 + X_1 \to X_2 + X_1$	$\dfrac{\beta_2}{1+(x_1/K_2)^{\nu_2}}$
$\mathcal{R}_4$	$X_2 \to \phi$	$d_2 x_2$

and the higher order terms can be negligible when T_s is sufficiently small compared to the magnitude of $F(\boldsymbol{\theta})$. This can be also confirmed from Fig. 1, where

$$\frac{\|\Psi e^{F(\boldsymbol{\theta})T_s}\Phi^T - e^{(\Psi F(\boldsymbol{\theta})\Phi^T)T_s}\|_F}{\|\Psi e^{F(\boldsymbol{\theta})T_s}\Phi^T\|_F}, \tag{16}$$

which is the normalized square sum of the entry-wise difference is plotted for the example described in Section VI. Thus, it is desirable that we select the sampling interval T_s so that it is fast enough for the time constants of the chemical reactions, but not too fast to exceed the technical limit of the flow cytometer.

VI. Example of Genetic Toggle Switch

We here demonstrate the proposed method on the genetic toggle switch [2]. The regulatory network of the genetic toggle switch is illustrated in Fig. 2. The network is composed of the two proteins X_1 and X_2 inhibiting each other, and it consists of the four chemical reactions shown in Table I. In the table, the symbol x_i denotes the copy number of X_i $(i = 1, 2)$, and the positive constants β_i and d_i $(i = 1, 2)$ denote the maximum production rate and the degradation rate of each molecule, respectively. The constant ν_i $(i = 1, 2)$ is the Hill coefficient specifying the cooperativity of the promoter region, and K_i $(i = 1, 2)$ represents the copy number at which the production rate becomes $\beta_i/2$. The goal of the parameter identification is to find $(\beta_1, \beta_2, d_1, d_2, K_1 K_2, \nu_1, \nu_2)$ from given joint distributions of x_i $(i = 1, 2)$.

We here demonstrate the identification results for two different data sets, namely, (i) ideal measurement and (ii) noisy measurement. The ideal measurement was constructed from the simulation of (5), and the noisy measurement was constructed by 50,000 sample paths of the stochastic simulations [16]. The noisy measurement was thus corrupted

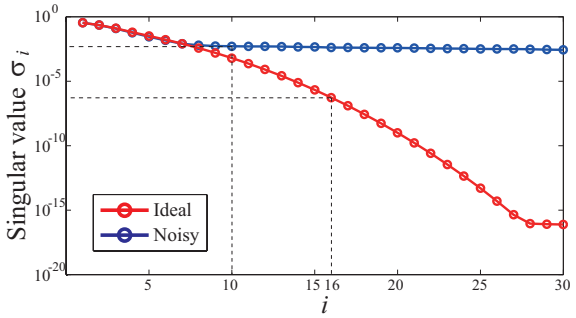

Fig. 3. Singular values of $P_{0:29}$ for the ideal and the noisy measurements

by noise due to the finite number of samples. Note that such high-throughput measurements are available with state-of-the-art flow cytometers. In the simulations, the number of molecules at the initial time, $t = 0$, was assumed to follow $x_i \sim U(0, 9)$ $(i = 1, 2)$, where $U(0, 9)$ stands for the discrete uniform distribution defined on $[0, 9]$. The actual parameters of the system were set as shown in set 1 in Table II.

We assumed that the measurements were obtained at 31 time points with the sampling interval $T_s = 1$ minute. Thus, $N = 30$ and $P_{0:30} = [\boldsymbol{p}[0], \boldsymbol{p}[1], \cdots, \boldsymbol{p}[30]]$. Figure 4 (Left) shows the measured distribution at $t = 30$. In this example, the system reached an equilibrium state around $t = 300$, which implies that we estimated the parameters from the transient response.

Given the measurements $P_{0:30}$, our first task was to construct the realization (7) by Algorithm 2. Let us focus on the identification based on (i) the ideal measurement data for a while. The red line in Fig. 3 shows the singular values of $P_{0:29}$. From this plot, we selected the order of the realization as $r = 16$, with which $\sigma_r \simeq 10^{-6}$. With this realization in hand, we solved the optimization problem derived in Section V. The lower and upper bounds of the parameter search regions were set as $[0.1, 0.5, 0.01, 0.01, 5.0, 5.0, 1.0, 1.0] \le [\beta_1, \beta_2, d_1, d_2, K_1, K_2, \nu_1, \nu_2] \le [5.0, 6.0, 0.1, 0.1, 25, 25, 5.0, 5.0]$. The initial values for the optimization were set as the mean of the lower and the upper bound. The identification result is shown in Table II (set 1, ideal). We see that the parameters were estimated with high accuracy (relative error less than $\pm 10\%$). Moreover, the total computation time for the identification was 23 seconds, which is substantially faster than the existing methods mentioned in Section I.

In order to see the performance of the algorithm under a more realistic situation, we estimated the parameters using (ii) the noisy measurement as well. We see in Fig. 3 that the decay rate of the singular values significantly drops after σ_{10}, which implies that the subspace associated with σ_1 through σ_{10} corresponds to the signal subspace [13]. Hence, we here constructed a realization with $r = 10$. In fact, this realization could replicate the measurement with high accuracy as shown in Fig. 4 (Right). Then, the optimization problem was solved with the same setting as before. The result is shown in Table II (set 1, noisy). The accuracy of

TABLE II

PARAMETER IDENTIFICATION RESULTS

Set		β_1	β_2	d_1	d_2	K_1	K_2	ν_1	ν_2	Time (sec.)
1	Actual	0.950	0.880	0.0255	0.0308	13.0	8.98	2.6	3.1	–
	Ideal	0.941	0.900	0.0254	0.0337	12.4	9.54	2.7	3.0	23
	Noisy	0.736	1.07	0.0193	0.0497	13.5	10.8	2.5	2.9	22
2	Actual	0.600	0.950	0.0190	0.0200	14.1	7.18	2.0	2.8	–
	Ideal	0.605	0.948	0.0192	0.0195	13.8	7.22	2.0	2.8	38
	Noisy	0.572	0.961	0.0196	0.0333	11.4	11.5	2.9	2.1	14
3	Actual	1.07	1.05	0.0214	0.0602	10.2	15.4	1.8	2.6	–
	Ideal	1.10	1.03	0.0233	0.0619	10.1	16.2	1.9	2.7	41
	Noisy	0.784	0.952	0.0227	0.0531	14.8	14.7	2.6	2.5	15

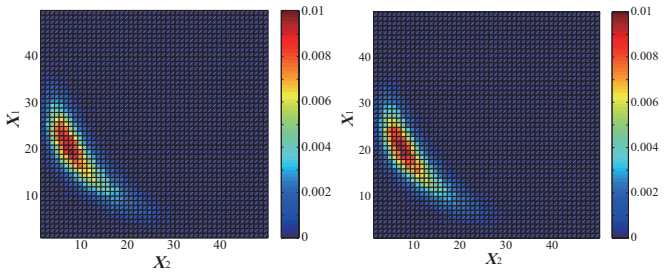

Fig. 4. (Left) Measured distribution and (Right) simulated distributions with the 10th order realization at $t = 30$. The simulated distribution replicates the measurements with high accuracy.

the identification is somewhat degraded due to the extrinsic noise, but the parameters are roughly estimated from the broad search space.

Identification results for two other parameter sets are also shown as parameter sets 2 and 3 in Table II. In most examples, all of the parameters were identified with high accuracy using the ideal measurements. Not surprisingly, the accuracy was degraded for the noisy measurements, but β_i and d_i $(i = 1, 2)$ could be estimated. On the other hand, the identification of K_i and ν_i $(i = 1, 2)$, which appear in the nonlinearity, was relatively difficult under the existence of extrinsic noise.

In these examples, we solved the optimization by 'fmincon' command in MATLAB R2009a with specifying the interior point method as the algorithm. The computation was executed on a standard desktop computer with Intel Core2 Quad processor Q9400 2.66GHz and 4GB RAM.

VII. CONCLUSION

In this paper, we have proposed a computationally efficient parameter identification method for stochastic biochemical networks. We have first shown that the realization obtained by the eigensystem realization algorithm has the diagonal time-limited gramian. This property has then allowed us to formulate the identification problem as the root finding of the Lyapunov equation. Finally, we have demonstrated the proposed method with the genetic toggle switch, and confirmed its distinctive features.

Although we have assumed that the joint distribution is available in this paper, it is preferable that the parameters can be identified from marginal distributions. Our future work will be devoted to such extensions. Noise robustness is another issue to be further considered in the future.

REFERENCES

[1] H. H. McAdams and A. Arkin, "It's a noisy business! genetic regulation at the nanomolar scale," *Trends in Genetics*, vol. 15, no. 2, pp. 65–69, 1999.

[2] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[3] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.

[4] A. Golightly and D. J. Wilkinson, "Bayesian sequential inference for stochastic kinetic biochemical network models," *Journal of Computational Biology*, vol. 13, no. 3, pp. 838–851, 2006.

[5] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, "Bayesian inference for discretely observed stochastic kinetic model," *Statistics and Computing*, vol. 18, no. 2, pp. 125–135, 2008.

[6] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *Journal of the Royal Society Interface*, vol. 6, no. 31, pp. 187–202, 2009.

[7] S. Reinker, R. M. Altman, and J. Timmer, "Parameter estimation in stochastic biochemical reactions," *IEE Proceedings - Systems Biology*, vol. 153, no. 4, pp. 168–178, 2006.

[8] S. K. Poovathingal and R. Gunawan, "Global parameter estimation methods for stochastic biochemical systems," *BMC Bioinformatics*, vol. 11, no. 414, 2010.

[9] B. Munsky and M. Khammash, "Identification from stochastic cell-to-cell variation: a genetic switch case study," *IET Systems Biology*, vol. 4, no. 6, pp. 356–366, 2010.

[10] ——, "The finite state projection algorithm for the solution of the chemical master equation," *Journal of Chemical Physics*, vol. 124, no. 4, p. 044104, 2006.

[11] J. N. Juang and R. S. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," *Journal of Guidance, Control and Dynamics*, vol. 8, no. 5, pp. 620–627, 1985.

[12] B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input-output functions," *Regelungstechnik*, vol. 14, no. 12, pp. 545–548, 1966.

[13] T. Katayama, *Subspace methods for system identification*. Springer, 2005.

[14] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, no. 1–3, pp. 404–425, 1992.

[15] A. C. Antoulas, *Approximation of large-scale dynamical systems*. Society for Industrial and Applied Mathematic, 2005.

[16] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.