Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Using Noise Transmission Properties to Identify Stochastic Gene Regulatory Networks

Brian Munsky[1,2] and Mustafa Khammash[1]

Abstract— **Gene regulatory networks often occur at such small scales that their dynamics are controlled by individual molecular events. This discrete behavior causes significant quantities of intrinsic noise. In some cases, mechanisms exist in the system to repress this noise. With different parameters, the same mechanisms may amplify the noise. By examining the properties of how noise is transmitted through the system, one can gather significant information about the system and aid researchers to identify system properties. In this paper, we consider a few simple analytical schemes to identify the parameters of gene transcription and translation processes with feedback regulation. While protein distributions can be measured with fluorescent protein tagging and flow cytometry, it is much more difficult to measure the quantities of messenger RNAs in a single cell. We show that with the right experimental procedures involving measurements of proteins alone, one can identify transcription *and* translation parameters.**

I. Introduction

Molecules exist in integer quantities that change at randomly distributed discrete times. In a chemical process, if the molecular population of a species is large, then a one or two molecule change will make little difference to the process. Furthermore, these changes occur so frequently that the process appears to behave continuously. However, if the population is small, then reactions will be much rarer and will have a much larger effect on the system dynamics. In a cell, the rare and discrete nature of chemical components such as genes, RNA molecules, and proteins, can lead to large amounts of intrinsic noise [1]–[7]. This intrinsic noise in gene regulatory networks has attracted much recent attention, and it is well established that different systems will exhibit different noise transmission properties. In some systems noise can be focussed [8], in some noise may cause or enhance resonant fluctuations [9], some systems may result in stochastic switching [10], [11], and in some systems, noise may be repressed [12].

Noise in systems biology is often viewed as a computational obstacle to be overcome. If one does not include it in the model, then one cannot hope to match the behavior of the actual system. However, in many cases, including noise in a model results in an explosion of computational complexity. Many approaches have been proposed to assist in the modeling of discrete stochastic systems such as kinetic Monte Carlo algorithms [13]–[16], stochastic differential equation approaches [17]–[19], the linear noise approximation and other moment matching techniques [20]–[23] and finite state projection approaches [24], [25]. At present, none of these approaches suffices to handle all systems, and there remains much work to be done to improve our computational capabilities. However, as these tools develop, it becomes more possible to overcome the obstacle of intrinsic noise and gain significant benefits in analytical studies. In this paper, we show how careful consideration of the transmission of noise provides a significant amount of information about the process. This information will, in turn, enable one to better identify properties of the system from experimental data.

In the next section we will present a simple mathematical description of a stochastic gene regulatory system with transcription and translation. Then in Sections III through V we show how the parameters of this model can be identified from various pieces of limited information. Finally, in section VII we make some concluding remarks.

II. Moment Analysis of a Simple Gene Network

Here we consider a simple description of gene transcription and translation. Let x denote the population of mRNA molecules, and let y denote the population of proteins in the system. The system population is assumed to change only through four reactions:

$$\emptyset \rightarrow mRNA$$

$$mRNA \rightarrow \emptyset$$

$$mRNA \rightarrow mRNA + protein$$

$$protein \rightarrow \emptyset$$

for which the propensity functions, $w_i(x, y)$, are

$$w_1(x, y) = k_1 + k_{21} y,$$

$$w_2(x, y) = \gamma_1 x,$$

$$w_3(x, y) = k_2 x, \text{ and}$$

$$w_2(x, y) = \gamma_2 y.$$

Here, the terms k_i and γ_i are production and degradation rates, respectively, and k_{21} corresponds to a feedback effect that the protein is assumed to have on the transcription process. In positive feedback, $k_{21} > 0$, the protein increases

[1]Brian Munsky (IEEE student member) and Mustafa Khammash (IEEE Fellow) are members of the Center for Control, Dynamical Systems and Computation, University of California, Santa Barbara, CA 93106. [2]Brian Munsky is also currently with the Center for Nonlinear Science and Computer, Computational and Statistical Sciences Division at Los Alamos National Laboratory, Los Alamos, NM 87545.

This material is based upon work supported by the National Science Foundation under Grant NSF-ITR CCF-0326576 and the Institute for Collaborative Biotechnologies through Grant DAAD19-03-D-0004 from the U.S. Army Research Office.

Emails: brian.munsky@gmail.com; khammash@engr.ucsb.edu

transcription; in negative feedback, $k_{21} < 0$, the protein inhibits transcription.

For this system, one can write the master equation [20]:

$$\begin{aligned}\dot{P}_{i,j}(t) = &-(k_1{+}k_{21}j + \gamma_1 i + k_2 i + \gamma_2 j)P_{i,j}(t)\\ &+ (k{+}k_{21}j)P_{i-1,j}(t)\\ &+ \gamma(i+1)P_{i+1,j}(t)\\ &+ k_2 i P_{i,j-1}(t)\\ &+ \gamma_2(j+1)P_{i,j+1}(t),\end{aligned} \tag{1}$$

where $P_{i,j}(t)$ is the probability that $(x,y) = (i,j)$ at the time t, conditioned on some initial probability distribution $\mathbf{P}(t_0)$. In this expression, the first negative term corresponds to the probability of transitions that begin at the state $(x,y) = (i,j)$ and leave to another state, and the remaining positive terms correspond to the reactions that begin at some other state $(x,y) \neq (i,j)$ and transition into the state (i,j).

The mean values of x and y can be written as:

$$\begin{aligned}v_1(t) = E\{x\} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i P_{i,j}(t)\\ v_3(t) = E\{y\} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} j P_{i,j}(t).\end{aligned} \tag{2}$$

The derivatives of these mean values are found simply by substituting (1) into (2):[1]

$$\dot{v}_1(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i\dot{P}_{i,j}(t) = k_1{+}k_{21}v_3 - \gamma_1 v_1,$$

and

$$\dot{v}_3 = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} j\dot{P}_{i,j}(t) = k_2 v_1 - \gamma_2 v_3.$$

Similarly, expressions for the second uncentered moments can be written:

$$\begin{aligned}v_2 = E\{xx\} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} ii P_{i,j},\\ v_4 = E\{yy\} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} jj P_{i,j},\\ v_5 = E\{xy\} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} ij P_{i,j},\end{aligned} \tag{3}$$

and evolve according to the set of ordinary differential equations:

$$\begin{aligned}\dot{v}_2 &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i^2\dot{P}_{i,j}(t)\\ &= k_1 + (2k_1+\gamma_1)v_1 - 2\gamma_1 v_2 {+} k_{21}v_3 + 2k_{21}v_5,\\ \dot{v}_4 &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} j^2\dot{P}_{i,j}\\ &= k_2 v_1 + \gamma_2 v_3 - 2\gamma_2 v_4, {+}2k_2 v_5,\end{aligned}$$

[1]Sample derivations for $\frac{d}{dt}v_1$ and $\frac{d}{dt}v_5$ are provided below in the appendix.

$$\begin{aligned}\dot{v}_5 &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} ij\dot{P}_{i,j}\\ &= k_2 v_2 + k_1 v_3 {+} k_{21}E v_4 - (\gamma_1+\gamma_2)v_5.\end{aligned}$$

Altogether the various components of the first two moments, $\mathbf{v}(t) := \left[\begin{array}{ccccc} E\{x\} & E\{x^2\} & E\{y\} & E\{y^2\} & E\{xy\}\end{array}\right]^T$, evolve according to the linear time invariant ODE:

$$\begin{aligned}\dot{\mathbf{v}} &= \left[\begin{array}{ccccc} -\gamma_1 & 0 & k_{21} & 0 & 0\\ \gamma_1+2k_1 & -2\gamma_1 & k_{21} & 0 & 2k_{21}\\ k_2 & 0 & -\gamma_2 & 0 & 0\\ k_2 & 0 & \gamma_2 & -2\gamma_2 & 2k_2\\ 0 & k_2 & k_1 & k_{21} & -\gamma_1-\gamma_2\end{array}\right]\mathbf{v} + \left[\begin{array}{c} k_1\\ k_1\\ 0\\ 0\\ 0\end{array}\right]\\ &= \mathbf{A}\mathbf{v} + \mathbf{b}\end{aligned} \tag{4}$$

Now that we have expressions for the dynamics of the first two moments, we will show in the following sections how these expressions can be used to help identify the various parameters: $[k_1, \gamma_1, k_2, \gamma_2, k_{21}]$ from properly chosen data sets.

III. Identifying Transcription Parameters

We begin by considering a simpler birth-death process of mRNA transcripts, whose populations are denoted by x. The moment equation for this system is:

$$\frac{d}{dt}\left[\begin{array}{c} v_1\\ v_2\end{array}\right] = \left[\begin{array}{cc} -\gamma & 0\\ \gamma+2k & -2\gamma\end{array}\right]\left[\begin{array}{c} v_1\\ v_2\end{array}\right] + \left[\begin{array}{c} k\\ k\end{array}\right],$$

where we have dropped the subscripts on k_1 and γ_1. By applying the nonlinear transformation:

$$\left[\begin{array}{c} \mu\\ \sigma^2-\mu\end{array}\right] = \left[\begin{array}{c} v_1\\ v_2 - v_1^2 - v_1\end{array}\right],$$

where μ and σ^2 refer to the mean and variance of x, respectively, we arrive at the transformed set of equations:

$$\begin{aligned}\frac{d}{dt}\left[\begin{array}{c} \mu\\ \sigma^2-\mu\end{array}\right] &= \left[\begin{array}{c} \dot{v}_1\\ \dot{v}_2 - 2\bar{x}\dot{\bar{x}} - \dot{v}_1\end{array}\right]\\ &= \left[\begin{array}{c} -\gamma_1 v_1 + k\\ (\gamma_1+2k)v_1 - 2\gamma v_2 + k - (2v_1+1)(-\gamma v_1 + k)\end{array}\right]\\ &= \left[\begin{array}{cc} -\gamma & 0\\ 0 & -2\gamma\end{array}\right]\left[\begin{array}{c} \mu\\ \sigma^2-\mu\end{array}\right] + \left[\begin{array}{c} k\\ 0\end{array}\right].\end{aligned} \tag{5}$$

Suppose that μ and σ^2 are known at two instances in time, t_0 and $t_1 = t_0 + \tau$, and denote their values at time t_i as μ_i and σ_i^2, respectively. The relationship between (μ_0, σ_0^2) and (μ_1, σ_1^2) is governed by the solution of (5), which can be written:

$$\left[\begin{array}{c} \mu_1\\ \sigma_1^2-\mu_1\end{array}\right] = \left[\begin{array}{c} \exp(-\gamma\tau)\mu_0\\ \exp(-2\gamma\tau)(\sigma_0^2-\mu_0)\end{array}\right] + \left[\begin{array}{c} \frac{k}{\gamma}(1-\exp(-\gamma\tau))\\ 0\end{array}\right] \tag{6}$$

In this expression there are 2 unknown parameters, γ and k, that we wish to identify from the data $\{\mu_0, \sigma_0^2, \mu_1, \sigma_1^2\}$. If $\mu_0 = \sigma_0^2$, the second equation is trivial, and we are left with only one equation whose solution could be any pair:

$$\left(\gamma,\ k = \gamma\frac{\mu_1 - \exp(-\gamma\tau)\mu_0}{1-\exp(-\gamma\tau)}\right).$$

If for the first measurement $\mu_0 \neq \sigma_0^2$ and for the second measurement $\mu_1 \neq \sigma_1^2$, then we can solve for:

$$\gamma = -\frac{1}{2t}\log\left(\frac{\sigma_1^2 - \mu_1}{\sigma_0^2 - \mu_0}\right)$$
$$k = \gamma\frac{\mu_1 - \exp(-\gamma t)\mu_0}{1 - \exp(-\gamma\tau)}.$$

Note that if μ_1 and σ_1^2 are very close, the sensitivity of γ to small errors in this difference becomes very large. From (6), one can see that as τ becomes very large $(\sigma_1^2 - \mu_1)$ approaches zero, and *steady state measurements do not suffice to uniquely identify both parameters*.

IV. Identifying Transcription and Translation Parameters

The full system in (4) has the solution:

$$\mathbf{v}_1 = \mathrm{e}^{\mathbf{A}\tau}\mathbf{v}_0 + \int_0^\tau \mathrm{e}^{\mathbf{A}(\tau-s)}\mathbf{b}ds, \tag{7}$$

where we have again adopted the notation that $\mathbf{v}_i = \mathbf{v}(t_i)$ and $t_{i+1} = t_i + \tau$. Drawing upon the fact that the parameters $\{k_1, \gamma_1, k_2, \gamma_2\}$ are all positive, one can show that the matrix $\mathbf{A}$ is stable and invertible so long as the following condition holds on the feedback term:

$$k_{21} \leq \frac{\gamma_1\gamma_2}{k_2}.$$

Under this condition, (7) can be written as:

$$\mathbf{v}_1 = \mathrm{e}^{\mathbf{A}\tau}\mathbf{v}_0 - \mathbf{A}^{-1}\left(\mathbf{I} - \mathrm{e}^{\mathbf{A}\tau}\right)\mathbf{b}. \tag{8}$$

Suppose that $\mathbf{v}_j$ has been measured at some equally distributed points in time $\{t_0, t_1, \ldots, t_m\}$, and one wishes to identify the parameters $\overline{\lambda} = \{k_1, \gamma_1, k_2, \gamma_2, k_{21}\}$ that satisfy:

$$\mathbf{J}(\overline{\lambda}) := \sum_{j=1}^{m}\left|\mathbf{v}_j - \mathrm{e}^{\mathbf{A}\tau}\mathbf{v}_{j-1} + \mathbf{A}^{-1}\left(\mathbf{I} - \mathrm{e}^{\mathbf{A}\tau}\right)\mathbf{b}\right| = \mathbf{0}.$$

The following subsections provide a few possible approaches to identify these parameters.

A. Looking at the invariant distribution

If the probability distribution dynamics described in (4) has an invariant distribution, then the steady state moments,

$$\mathbf{v}_\infty = \lim_{t\to\infty}[v_1, v_2, v_3, v_4, v_5]^T,$$

must satisfy:

$$\mathbf{A}\mathbf{v}_\infty - \mathbf{b} = \mathbf{0}.$$

This equation can be rewritten in terms of the unknown parameters as:

$$\mathbf{\Psi}_\infty\overline{\lambda} = \lim_{t\to\infty}\mathbf{\Psi}(t)\overline{\lambda} = \mathbf{0},$$

where

$$\mathbf{\Psi}(t) = \begin{bmatrix} 1 & -v_1 & 0 & 0 & v_3 \\ 1{+}2v_1 & v_1 - 2v_2 & 0 & 0 & v_3{+}2v_5 \\ 0 & 0 & v_1 & -v_3 & 0 \\ 0 & 0 & v_1{+}2v_5 & v_3{-}2v_4 & 0 \\ v_3 & -v_5 & v_2 & -v_5 & v_4 \end{bmatrix}.$$

From this expression, it is obvious that there are two possible cases: (1) the rank of the matrix is full and we are left with the trivial solution $\overline{\lambda} = \mathbf{0}$, or (2) the matrix has a null-space spanned by $\{\phi_1, \ldots, \phi_p\}$ and there are an infinite number of parameter sets that will result in the same invariant distribution:

$$\overline{\lambda} = \sum_{i=1}^{p}\alpha_i\phi_i, \text{ for any } [\alpha_1, \ldots, \alpha_p] \in \mathbb{R}^p.$$

So long as the parameters enter linearly into the propensity functions $w(\mathbf{x}) = \sum_{\mu=1}^{M} c_\mu f(\mathbf{x})$, then one can extend this argument for any finite number of n moments of the stationary distribution. This tells us that *the steady state distribution cannot provide enough information* to uniquely identify the set of system parameters. Additional information is needed. For example, if the rank of the null space is one, then the knowledge of any one parameter from the set $\overline{\lambda}$ can provide an additional linearly independent equation, and can enable the unique determination of the parameters. If the rank of the null space is p, then at least p additional, linearly independent, pieces of information will be required.

B. Identifying parameters with full state and derivative information

Suppose that it is possible to measure both the moments and their time derivatives at specific instances in time. In this case, we have the same expressions as above but at a finite time where the time derivatives are non-zero:

$$\mathbf{\Psi}(t)\overline{\lambda} = \dot{\mathbf{v}}(t)$$

Depending on the values of $\mathbf{v}(t)$, the matrix $\mathbf{\Psi}(t)$ may or may not have full rank. In particular, if the system is at an invariant distribution as above, then $\mathbf{\Psi}(t)$ will not be invertible. As another example, if the measurements are taken when $y = 0$ then $E\{y\} = E\{y^2\} = E\{xy\} = 0$ and the 4th and 5th columns of $\mathbf{\Psi}(t)$ will be zero, and the rank will be at most 3. In this case, the parameters γ_2 and k_{12} will not be identifiable. If $\mathbf{v}(t)$ can be specified such that $\mathbf{\Psi}(t)$ is invertible, then the parameters can be identified directly from the measurement of $\mathbf{v}(t)$ and its derivative, $\dot{\mathbf{v}}(t)$.

C. Identification without derivative knowledge

In most cases it is not feasible to measure the time derivative of the moments. More likely, one will only be able to measure the moments at discrete instances in time. In this case one must perform the identification analysis in discrete time according to (8), which can be rewritten as:

$$\mathbf{v}_j = \mathbf{G}\mathbf{v}_{j-1} + \psi.$$

Here, the matrix $\mathbf{G}$ and the vector ψ are the unknown quantities that we wish to identify. These matrices will be subject to some nonlinear constraints of the form

$$\mathbf{G} = \exp\left(\mathbf{A}\tau\right), \text{ and}$$
$$\psi = -\mathbf{A}^{-1}\left(\mathbf{I} - \mathrm{e}^{\mathbf{A}(t_1 - t_0)}\right)\mathbf{b}, \tag{9}$$

where $\mathbf{A} = \mathbf{A}(\lambda)$ and $\mathbf{b} = \mathbf{b}(\lambda)$ are given as above in (4).

The relation between $\mathbf{v}_i$ and $\mathbf{v}_{i-1}$ in (8) can be rearranged as:

$$\mathbf{v}_i = [\mathbf{G}, \psi] \begin{bmatrix} \mathbf{v}_{i-1} \\ 1 \end{bmatrix}.$$

For now, ignore the constraints in (9) and suppose that we want to solve for the 5×6 matrix $[\mathbf{G}, \psi]$. With measurements of $\mathbf{v}_0$ and $\mathbf{v}_1$, we would have only five equations but 30 unknown values (25 in $\mathbf{G}$ and 5 in ψ). This is not yet enough. However, if we take measurements at seven equally distributed points in time $\{\mathbf{v}(t_i)\}$, we can write:

$$\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_6 \end{bmatrix} = [\mathbf{G}, \psi] \begin{bmatrix} \mathbf{v}_0 & \dots & \mathbf{v}_5 \\ 1 & \dots & 1 \end{bmatrix}$$
$$\mathbf{V}_f = \hat{\mathbf{G}}\mathbf{V}_i, \tag{10}$$

where $\hat{\mathbf{G}} = [\mathbf{G}, \psi]$ is the matrix of unknown values. Now we have thirty equations with which we can find the thirty unknown values provided that the equations are linearly independent–a fact that can be checked by examining the rank of the matrix $\mathbf{V}_i$. As long as $\mathbf{V}_i$ has full rank, then the solution for $\hat{\mathbf{G}}$ is given by:

$$\hat{\mathbf{G}} = \mathbf{V}_f \mathbf{V}_i^{-1}.$$

In the case of measurement noise it is often advantageous to have more than the minimum number of measurements in (10). In this case $\hat{\mathbf{G}}$ should be chosen as the argument that minimizes $\mathbf{V}_f - \hat{\mathbf{G}}\mathbf{V}_i$ in the least squares sense:

$$\hat{\mathbf{G}} = \mathbf{V}_f \mathbf{V}_i^{-R}.$$

Once we have extracted $\mathbf{G}$ from $\hat{\mathbf{G}}$, we can diagonalize it:

$$\mathbf{G} = \mathrm{e}^{\mathbf{A}\tau} = \mathbf{S}^{-1}\mathrm{e}^{\mathbf{\Lambda}\tau}\mathbf{S},$$

and solve for the matrix $\mathbf{A}$:

$$\mathbf{A} = \mathbf{S}^{-1}\mathbf{\Lambda}\mathbf{S} = \frac{1}{\tau}\mathbf{S}^{-1}\log(\mathbf{S}\mathbf{G}\mathbf{S}^{-1})\mathbf{S},$$

where $\log(\mathbf{S}\mathbf{G}\mathbf{S}^{-1})$ corresponds to the natural logarithm of the elements of diagonal matrix $\mathbf{S}\mathbf{G}\mathbf{S}^{-1}$. Finally, we also have

$$\psi = -\mathbf{A}^{-1}(\mathbf{I} - \mathbf{G})\,\mathbf{b},$$

which gives:

$$\mathbf{b} = -(\mathbf{I} - \mathbf{G})^{-1}\mathbf{A}\mathbf{v}.$$

Now, it is relatively easy to solve for the parameters: $\{k_1, \gamma_1, k_2, \gamma_2, k_{21}\}$ from the definition of $\mathbf{A}$ in (4).

V. Non-linear optimization based identification

In the previous section, we did not apply the nonlinear constraints (9) on the unknown values of $\mathbf{G}$ and ψ. As a result, we were left with thirty unknowns for which we required thirty linearly independent equations. The advantage of such an approach is that the parameters are easily identified from the data by performing a few simple matrix operations. However, to get these equations, we are forced to measure $\mathbf{v}_i$ at seven different points in time. Since $\mathbf{G}$ and ψ are defined by non-linear equations of only five variables, it is reasonable to expect that these parameters should be recoverable with far fewer measurements. However, in this case it is no longer easy to find closed analytical expressions to determine the parameters from the measurements. Instead we must seek to find the argument that minimizes

$$J(\overline{\lambda}) = \left|\left|\mathbf{V}_f - \hat{\mathbf{G}}\mathbf{V}_i\right|\right|_F,$$

where the $||.||_F$ refers to the Frobenious norm (sum of squares of all elements). In the examples below, this minimization is done numerically under the constraints in (9), and the definitions of $\mathbf{A}$ and $\mathbf{b}$ in (4).

A. Identifying parameters with protein distributions only

While it is not currently possible to measure the cell by cell distribution of mRNAs, it is possible to get this information for protein distributions. To do this, one can attach florescent tags, such as green florescent protein (GFP), to the protein of interest and then measure the expression of that protein using flow cytometry or fluorescence activated cell sorting (FACS). Such an approach will yield a histogram of the number of cells containing different levels of the protein. In this section, we present an identification approach with which this protein distribution information is sufficient to identify rates for transcription and translation.

Supposing that it is only possible to measure the first and second moment of the protein distribution, then these measurements are of the form: $\mathbf{q}_i = \mathbf{C}\mathbf{v}_i$, where

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

In the previous cases, it has been assumed that the initial distribution is known or measurable, but in this case the five initial values of $\mathbf{v}_0$ must now also be estimated in the identification procedure. The identification problem is now to find the set of parameters $\overline{\lambda} = [k_1, \gamma_1, k_2, \gamma_2, k_{21}] \bigcup \hat{\mathbf{v}}_0 \in \mathbb{R}^{10}$, all positive except k_{21} that minimizes

$$J(\overline{\lambda}) = \sum_{i=0}^{m} \left|\mathbf{q}_i - \mathbf{C}\hat{\mathbf{v}}_i\right|_2,$$

where $\mathbf{q}_i$ is the measurement at the i^{th} time point, and $\hat{\mathbf{v}}_i$ is the corresponding estimate of $\mathbf{v}_i$. Substituting the expression (8) for $\hat{\mathbf{v}}$ yields

$$J(\overline{\lambda}) = \left|\mathbf{q}_0 - \mathbf{C}\hat{\mathbf{v}}_0\right|_2 + \sum_{i=1}^{m} \left|\mathbf{q}_i - \mathbf{C}\left(\mathbf{G}^i\hat{\mathbf{v}}_0 + \sum_{j=0}^{i-1}\mathbf{G}^j\psi\right)\right|_2$$

where $\mathbf{G}$ and ψ are functions of $(k_1, \gamma_1, k_2, \gamma_2, k_{21})$ subject to the constraints in (9), and the definitions of $\mathbf{A}$ and $\mathbf{b}$ in (4).

In order to fit the ten unknown quantities in $\overline{\lambda}$, we require at least ten independent equations and ten data points. In the case where the protein first and second moments are measured, this requires measurements at five different time points. With full state measurement, $\mathbf{C} = \mathbf{I}$, as few as two time points will be sufficient, provided that those measurements are rich in all transient dynamics.

VI. Examples

To examine the utility of the above identification techniques, we have numerically generated a set of over 2200 gene regulatory networks in which the parameters are randomly chosen:

$$k_1 = \mathrm{U}(0, 0.2),\ k_2 = \mathrm{U}(0, 0.2),\ k_{21} = \mathrm{U}(-0.0002, 0)$$
$$\gamma_1 = \mathrm{U}(0, 0.002),\ \gamma_2 = \mathrm{U}(0, 0.002)$$

where we have used the notation $\mathrm{U}(a, b)$ to denote a uniform random number between a and b. The initial distributions are also chosen randomly according to:[2]

$$\begin{aligned} v_1^0 &= E\{x(0)\} = \mathrm{U}(0, 10), \\ v_2^0 &= E\{x^2(0)\} = (v_1^0)^2\mathrm{U}(1, 2), \\ v_3^0 &= E\{y(0)\} = \mathrm{U}(0, 100), \\ v_4^0 &= E\{y^2(0)\} = (v_3^0)^2\mathrm{U}(1, 2), \text{ and} \\ v_5^0 &= E\{x(0)y(0)\} = v_3^0 v_1^0. \end{aligned}$$

We seek to identify these parameters and initial conditions through three approaches.

1. Using full state knowledge without non-linear constraints (FL, Section IV-C).

2. Using full state knowledge with non-linear constraints (FNL, Section V).

3. Using partial state knowledge with non-linear constraints (PNL, Section V-A).

Each identification is conducted under the assumption that there is no measurement noise contained in the identification data. For the non-linear optimization approaches (FNL and PNL), the initial guess for each parameter is randomly chosen to be within one degree of magnitude of its true value. All non-linear optimizations use MATLAB's standard optimization routine "fminsearch". In cases when the optimization terminates with a loss function that is greater than ε, the optimization routine makes a new random initial guess and reattempts the optimization. Three cases are possible: (*i*) If the optimization does not converge within twenty attempts, then identification is deemed inconclusive. (*ii*) When the loss function converges to less than ε, and the corresponding parameters, $\hat{\lambda}_i$, satisfy

$$\sum_i \left(\frac{\hat{\lambda}_i - \overline{\lambda}_i}{\overline{\lambda}_i} \right)^2 \leq \delta^2,$$

then that identification is considered to have been successful. (*iii*) Finally, if the optimization routine converges within ε, but the parameters are not satisfactorily close to the true values, the optimization is considered to have yielded a false positive. For our analyses, we have chosen values $\varepsilon = 10^{-7}$ and $\delta = 0.01$.

In every case the FL optimization procedure successfully identified all of the unknown parameters. Also, because this procedure relies only upon a few relatively simple matrix operations and not a numerical optimization, this approach is by far the most efficient. However, this identification approach requires a total of thirty-five measurement quantities for each system (five states at seven time points). In practice such experimental results may be prohibitively expensive or otherwise impossible to obtain.

[2]The initial distributions are chosen in this manner to guarantee that the variance is non-negative, and the covariance of x and y is zero.

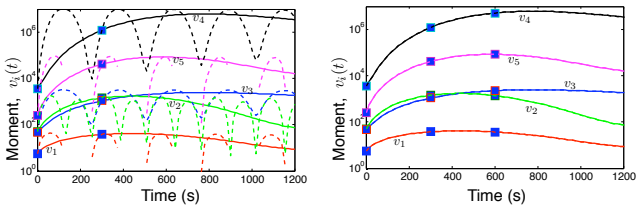

Fig. 1. Comparison of the dynamics of the true (solid lines) and estimated (dashed lines) system moments for a random set of parameters and initial distribution. Here the FNL estimation uses all five elements of the first two moments (v_1 through v_5). (left) Estimation based upon the measurements at two time points shown in squares. (right) Estimation based upon the measurements at three time points.

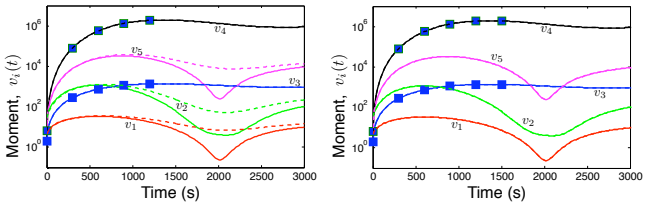

Fig. 2. Comparison of the dynamics of the true (solid lines) and estimated (dashed lines) system moments for a random set of parameters and initial distribution. Here the PNL estimation uses only data about the protein mean and second moments only (v_3 and v_4). (left) Estimation based upon the measurements at five time points shown in squares. (right) Estimation based upon the measurements at six time points.

The FNL routine has been applied for measurements of all five states in $\mathbf{v}(t)$, but at only two points in time. The numerical optimization converged in every case but two, but falsely identified the system parameters for about 8.5% of the systems. By increasing the number of measurements to three time points (less than half the number of measurements required for the FL method), the success rate of the FNL on the same systems and same initial conditions rose to 100%. Fig. 1 illustrates one case in which the FNL identification failed dramatically for a data set of two time points, but succeeded with one additional time point.

For the PNL identification, we have sought to find the parameters using only the protein information at five separate equally distributed points in time. This more computationally intensive approach identified the parameters for about 66.5% of the systems. However this approach failed to converge for 16.1% of the systems and provided false identifications for 17.4% of the systems. Once again, the addition of more time points confers a large advantage (See for example Fig. 1). With protein measurements at 6 time points, the false identification rate dropped to less than 0.2%.

VII. Conclusions

Due to the inherently discrete nature of gene regulatory networks, intrinsic noise is an important concern for many systems biologists. Although the inclusion of noise makes

the analysis of these networks more difficult to conduct, there is much benefit to including noise in one's model. In this paper, we have illustrated how stochastic analysis of gene transcription and translation enables one to identify model parameters. We have presented several approaches with which to conduct this identification. Should it become possible to accurately measure the joint distributions of mRNA and protein molecules at many points in time, then all parameters of the system can be identified through a few relatively simple matrix operations. However, in many cases, it will not be possible to obtain such a wealth of information. By utilizing the structure of the system of moment dynamics, one can often identify the parameters from a much smaller data subset. In particular, we have shown that transcription and translation parameters can simultaneously be identified solely from protein data. In all cases, the identification procedure relies upon precise measurement of transient data. If the initial condition is too close to an invariant manifold, or if the time between measurements is too long, then the parameters will not be uniquely identifiable. In these case additional experiments or alternate initial conditions must be examined. The current work focusses on a simple toy model of a single gene, mRNA, protein triplet. In more realistic problems, there will be many more chemical players. Such studies remain to be done, but it is envisioned that the conclusions will be the same: more information about transient noise transmission will allow one to better identify any system.

REFERENCES

[1] M. McAdams and A. Arkin, "Its a noisy business!" *Tren. Gen.*, vol. 15, no. 2, pp. 65–69, 1999.

[2] M. Elowitz, A. Levine, E. Siggia, and P. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.

[3] M. Thattai and A. van Oudenaarden, "Intrinsic noise in gene regulatory networks," *Proc. Natl. Acad. Sci.*, vol. 98, pp. 8614–8619, 2001.

[4] J. Hasty, J. Pradines, M. Dolnik, and J. Collins, "Noise-based switches and amplifiers for gene expression," *PNAS*, vol. 97, pp. 2075–2080, 2000.

[5] E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman, and A. van Oudenaarden, "Regulation of noise in the expression of a single gene," *Nature Genetics*, vol. 31, pp. 69–73, 2002.

[6] N. Federoff and W. Fontana, "Small numbers of big molecules," *Science*, vol. 297, no. 5584, pp. 1129–1131, 2002.

[7] T. Kepler and T. Elston, "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations." *Biophys. J.*, vol. 81, pp. 3116–3136, 2001.

[8] J. Paulsson, O. Berg, and M. Ehrenberg, "Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation," *PNAS*, vol. 97, no. 13, pp. 7148–7153, 2000.

[9] H. Li, Z. Hou, and H. Xin, "Internal noise stochastic resonance for intracellular calcium oscillations in a cell system," *Phys. Rev. E*, vol. 71, no. 061916, 2005.

[10] A. Arkin, J. Ross, and M. H., "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[11] T. Tian and K. Burrage, "Stochastic models for regulatory networks of the genetic toggle switch," *PNAS*, vol. 103, no. 22, pp. 8372–8377, May 2006.

[12] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano, "Noise in transcription negative feedback loops: simulation and experimental analysis," *Mol. Syst. Biol.*, vol. 2, no. 41, 2006.

[13] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2360, May 1977.

[14] Y. Cao, D. Gillespie, and L. Petzold, "The slow-scale stochastic simulation algorithm," *J. Chem. Phys.*, vol. 122, no. 014116, Jan. 2005.

[15] E. Haseltine and J. Rawlings, "Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics," *J. Chem. Phys.*, vol. 117, no. 15, pp. 6959–6969, Jul. 2002.

[16] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 115, no. 4, pp. 1716–1733, Jul. 2001.

[17] ——, "The chemical langevin equation," *J. Chem. Phys.*, vol. 113, no. 1, pp. 297–306, Jul. 2000.

[18] ——, "The chemical langevin and fokker-plank equations for the reversible isomerization reaction," *J. Phys. Chem.*, vol. 106, pp. 5063–5071, 2002.

[19] J. Aparicio and H. Solari, "Population dynamics: Poisson approximation and its relation to the langevin process," *Physical Review Letters*, vol. 86, no. 18, pp. 4183–4186, April 2001.

[20] N. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. Elsevier, 2001.

[21] J. Elf and M. Ehrenberg, "Fast evaluations of fluctuations in biochemical networks with the linear noise approximation," *Genome Research*, vol. 13, pp. 2475–2484, 2003.

[22] C. Gmez-Uribe and G. Verghese, "Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations," *JCP*, vol. 126, no. 024109, Jan. 2007.

[23] A. Singh and J. Hespanha, "A derivative matching approach to moment closure for the stochastic logistic model," *Bulletin of Mathematical Biology*, vol. 69, pp. 1909–1925, 2007.

[24] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *J. Chem. Phys.*, vol. 124, no. 044104, 2006.

[25] ——, "The finite state projection approach for the analysis of stochastic noise in gene networks," *IEEE Trans. Automat. Contr./IEEE Trans. Circuits and Systems: Part 1*, vol. 52, no. 1, pp. 201–214, Jan. 2008.

APPENDIX

For the reader's convenience, this appendix provides the detailed derivation of the moment dynamics of $E\{xy\}$. The derivations for the other uncentered moments are very similar. In each case, simply substitute the master equation (1) into the definition of the moment of interest and carry out some algebraic manipulations:

$$
\begin{aligned}
\dot{v}_5(t) &= \frac{d}{dt}E\{xy\} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} ij\dot{P}_{i,j} \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} ij\,\{-(k_1+k_{21}j+\gamma_1 i+k_2 i+\gamma_2 j)P_{i,j}(t) \\
&\qquad +(k_1+k_{21}j)P_{i-1,j}(t)+\gamma_1(i+1)P_{i+1,j}(t) \\
&\qquad +k_2 i P_{i,j-1}(t)+\gamma_2(j+1)P_{i,j+1}(t)\} \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} -(k_1 ij+k_{21}ij^2+\gamma_1 i^2 j+k_2 i^2 j+\gamma_2 ij^2)P_{i,j}(t) \\
&\qquad +(k_1 ij+k_{21}ij^2)P_{i-1,j}(t)+\gamma_1(i+1)ijP_{i+1,j}(t) \\
&\qquad +k_2 i^2 jP_{i,j-1}(t)+\gamma_2(j+1)ijP_{i,j+1}(t) \\
&= -k_1E\{xy\}-k_{21}E\{xy^2\}-\gamma_1E\{x^2y\}-k_2E\{x^2y\} \\
&\qquad -\gamma_2E\{xy^2\}+k_1E\{(x+1)y\}+k_{21}E\{(x+1)y^2\} \\
&\qquad +\gamma_1E\{x(x-1)y\}+k_2E\{x^2(y+1)\}\} \\
&\qquad +\gamma_2E\{yx(y-1), \\
&= k_2E\{x^2\}+k_1E\{y\}+k_{21}E\{y^2\}-(\gamma_1+\gamma_2)E\{xy\}, \\
&= k_2v_2+k_1v_3+k_{21}v_4-(\gamma_1+\gamma_2)v_5.
\end{aligned}
$$