Submitted, 2006 Conference on Decision and Control (CDC)
http://www.cds.caltech.edu/~murray/papers/2006a_dm06-cdc.html

Towards Biological System Identification: Fast and Accurate Estimates of Parameters in Genetic Regulatory Networks

Mary J. Dunlop and Richard M. Murray

Abstract— **System identification tools are useful for developing mathematical models of dynamical systems based on experimental observations, but many standard techniques are not applicable to biological problems where nonlinear effects are the norm and output measurements are limited. We focus on parameter identification in genetic regulatory networks as an example of a class of interesting biological system identification problems. We compare the performance of two methods, the extended Kalman filter and a nonlinear least squares fit, as they estimate the parameters in a model of a bistable genetic circuit. The extended Kalman filter does dramatically better than the nonlinear fit in predicting parameters when sensor noise is high. The settling time of the parameter estimate is also measured and it is shown that by choosing inputs appropriately, the convergence time of the parameter estimates can be reduced. We present a method for choosing an approximation of the optimal input for parameter estimation. Some challenges that are unique to biological system identification problems are discussed.**

I. INTRODUCTION

Biological circuit design, also referred to as synthetic biology, is an emerging field of research in which one "builds" a system by inserting synthesized DNA into a biological host (usually a bacterium or yeast) where the inserted DNA encodes a desired genetic "circuit" [1], [2]. Simple circuits make use of the modularity of biology to construct feedback circuits that implement binary switches [3], inverters [4] and simple oscillators [5]. Building these circuits provides two clear mechanisms for understanding biological processes: by building a circuit that exploits biological principles we can better explore our understanding of biological function and (longer term) by building biological devices, we can develop new "instrumentation" for measuring biological signals that yield more information about biological processes.

In analyzing and eventually designing these systems one would like to make use of models that capture the behavior of the circuit and allow an analysis of the trade-offs in stability, performance, and robustness. This is particularly challenging in biological systems because the underlying physics of the system is very complicated and in many cases is not yet well understood. Furthermore, while there is good understanding of how some genetic regulatory mechanisms function, there is still large variability in the parameters that govern the underlying processes. Rate constants, binding affinities, and environmental factors are only known approximately and models for these systems are often very simple and may miss important features that become relevant as we seek to move beyond demonstrations of synthetic circuits to the use of these circuits for scientific and engineering purposes.

One path toward providing more useful models is to develop techniques for biological system identification. Existing techniques in system identification provide a rigorous framework for constructing models from input/output data in the presence of both noise and uncertainty (see, eg, [6]). However, it is not clear that these techniques will be directly applicable to biological systems, where strong nonlinearities are present and the effects of noise are particularly severe.

The existing literature on nonlinear system identification is substantial, although little is available that appears to address the specific types of features that we see in biological systems. Traditional nonlinear system identification makes use of linear relationships between the parameters of the system and nonlinear functions of the state or by separating out nonlinear elements which can be explicitly mapped [7], [6], [8]. Additional techniques have been developed using integral representations, such as Volterra expansions [9]. From a practical perspective, a common technique is to add parameters as states and perform system identification using an extended Kalman filter.

In this paper, we explore the applications of some existing techniques for nonlinear system identification as well as propose some new approaches that are tuned for use in biological circuit design. A central theme in our approach is how to choose the forcing function that is used to generate the data for the system identification process. While there is a substantial literature in design of experiments, there appears to be relatively little work available in choosing inputs for nonlinear systems that provide the best data for system identification. The main contribution of this paper is to provide some insights into how such inputs might be chosen and to assess several different choices in the system identification process in terms of their applicability to biological circuits.

II. BISTABLE SWITCH

We focus our analysis on one genetic circuit, the bistable switch. Simple models show that the switch can have two stable equilibrium points, a promising characteristic for building genetic circuits with "logic" since the switch can encode both high and low states like an electronic circuit. A synthetic bistable switch was successfully constructed and shown to work experimentally in *E. coli* in [3].

Research supported in part by ARO Institute for Collaborative Biotechnology. M.J.D. is supported by the Department of Energy Computational Science Graduate Fellowship.

Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, 91125 mjdunlop@caltech.edu, murray@cds.caltech.edu.

Cells make use of these genetic toggle switches in remarkably diverse situations because switching between states is relatively robust, even in noisy environments. Recently, it has been shown that bistable switches can be used in combination to further decrease sensitivity to noise [10].

Fig. 1a shows the "wiring" diagram for the bistable switch. Protein A represses the production of protein B and protein B represses production of protein A. Addition of an inducer molecule, I, can change the activity of protein B, making it less likely to block production of protein A.

As in [3], the bistable switch can be approximately described by the dimensionless model

$$\dot{p}_A = \frac{\alpha}{1+(u\ p_B)^n} - p_A \quad (1)$$

$$\dot{p}_B = \frac{\alpha}{1+(p_A)^n} - p_B, \quad (2)$$

where p_A and p_B are the concentrations of proteins A and B, u is the input (a function of I, described below), α is the effective protein production rate, and n is the cooperativity. The production rate describes how quickly a protein accumulates compared to the rate at which it degrades. The cooperativity is a measure of how many protein molecules must act cooperatively before gene expression can be turned off.

Fig. 1b shows the effect of the parameters α and n in a single switch. The nonlinear term

$$\frac{\alpha}{1+(p_A)^n} \quad (3)$$

is known as the Hill function. The production rate α is the steady state concentration of protein B when there is no protein A present. As the concentration of A increases, it represses the production of B. The cooperativity n determines how step-like B's switch from high to low concentrations is.

The bistable switch can be controlled by adding an inducer, I, which represses the activity of protein B. This is modeled using the function

$$u = \frac{\beta}{1+\left(\frac{I}{K}\right)^m}. \quad (4)$$

The parameters β, K, and m are specific to the type of inducer added and are assumed to be known in advance.

These equations make several important assumptions. First, for simplicity we consider the symmetric case where the production rate and cooperativity associated with A repressing B (Eqn. 2) are identical to B repressing A (Eqn. 1). This model also assumes that concentrations of chemical species are large enough that they can be expressed with continuous dynamics. Other limitations associated with biological system identification are addressed in Section VII.

III. PARAMETER IDENTIFICATION WITH THE EXTENDED KALMAN FILTER

One standard technique for system identification is the use of an extended Kalman filter. Here we apply the extended Kalman filter to the bistable switch system and use it to estimate the model parameters α and n.

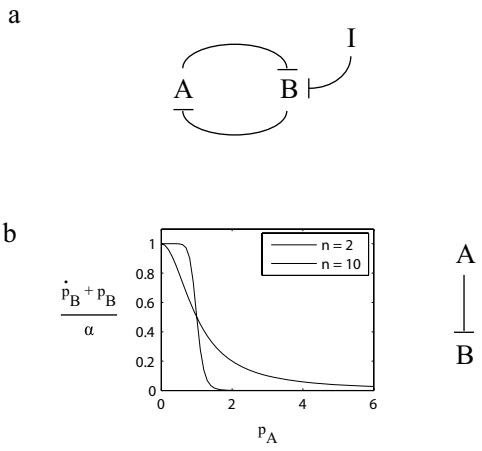

Fig. 1. a) The bistable switch design. T-shaped arrows indicate repression. Protein A represses the production of protein B and visa versa. Forcing is introduced through the inducer I, which represses the activity of protein B. b) Parameters α and n in the Hill function (the nonlinear term in Eqns. 1 and 2). The effects of α and n are shown for a single switch (A represses B).

We design an observer for the following system:

$$\frac{d}{dt}\begin{bmatrix} p_A \\ p_B \\ \alpha \\ n \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{1+(u\ p_B)^n} - p_A \\ \frac{\alpha}{1+(p_A)^n} - p_B \\ 0 \\ 0 \end{bmatrix} \quad (5)$$

$$\dot{x} = f(x,u), \quad (6)$$

where the state $x = (p_A, p_B, \alpha, n)$ includes the two parameters we wish to identify.

For the purposes of this paper we ignore process noise, but include sensor noise. Biologists are only beginning to quantify how process noise affects genetic regulatory mechanisms [11], [12]. Cellular process noise is particularly relevant in studying biological systems and is worth further exploration, but is omitted here for simplicity.

In order to observe all states we first verify that the system is linearly observable. Eqn. 6 is linearized to find

$$A = \left.\frac{\partial f}{\partial x}\right|_{x_{eq},u_{eq}}.$$

The system is linearly observable if both protein concentrations p_A and p_B are measured, but is not linearly observable if only one is measured. This result assumes biologically realistic parameter values, namely $\alpha > 0$, $n > 0$. We consider the case where p_A and p_B can be measured directly and define

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

The output is given by the equation

$$y = Cx + w, \quad (7)$$

where w is the sensor noise. Since p_A and p_B will be measured using the same experimental methods, it is expected

that the sensor noise associated with both states will have the same statistical properties. Thus, we define

$$R = \rho \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

as the covariance matrix $E\{ww^T\}$.

Simulation data are generated by integrating Eqn. 6 to find x, which is then used to find the output

$$y = Cx + \sqrt{\rho}\ \nu, \tag{8}$$

where ν is a vector of normally distributed random numbers with mean 0 and standard deviation 1. The dimension of ν is the same as the dimension of the y (2×1 for the bistable switch example).

The extended Kalman filter is run about a nominal trajectory, so we define the following transformations:

$$\begin{aligned} q &= x - x_n \\ r &= y - y_n \\ s &= u - u_n, \end{aligned}$$

where x_n, y_n, u_n are the state, output, and input for a nominal system that is simulated using the dynamics in Eqn. 6. Note that x_n and y_n are functions of time. The nominal input is chosen to be $u_n = 1$ and parameter values and initial conditions of the nominal system are selected from probability distributions, as described below.

Since

$$\begin{aligned} \dot{q} &= \dot{x} - \dot{x}_n \\ &= f(x,u) - f(x_n,u_n) \\ &= f(q + x_n, s + u_n) - f(x_n,u_n), \end{aligned}$$

the nonlinear observer is

$$\dot{\hat{q}} = f(\hat{q} + x_n, s + u_n) - f(x_n,u_n) + L(r - C\hat{q}), \tag{9}$$

where L is the matrix of observer gains given by

$$L = PC^T R^{-1}$$

and P is found by integrating the Riccati ODE

$$\begin{aligned} \dot{P} &= \tilde{A}P + P\tilde{A}^T - PC^T R^{-1} CP \\ P(0) &= E\{q(t_0)q(t_0)^T\}. \end{aligned}$$

The matrix $\tilde{A}$ is

$$\tilde{A} = \left.\frac{\partial f}{\partial q}\right|_{\hat{q},x_n,s,u_n}.$$

Parameter values α and n and initial conditions $x(0)$ and $x_n(0)$ are selected randomly from normal distributions with the mean and standard deviations given in Table I. Initial state estimates $\hat{q}(0)$ are set to 0. Since negative parameter and state values are not realistic, any randomly generated value that is less than 0 is discarded and regenerated.

Fig. 2 shows the performance of the extended Kalman filter in one test run. In this example both state and parameter estimates converge to the actual values.

TABLE I

MEAN AND STANDARD DEVIATION USED FOR SIMULATIONS

	Mean	Standard Deviation
$p_A(0)$	0.5	0.5
$p_B(0)$	0.5	0.5
α	150	100
n	2	1

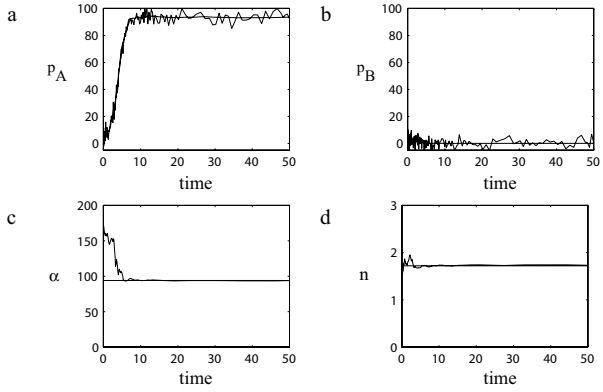

Fig. 2. Extended Kalman filter performance. a) Estimated state $\hat{p}_A$ (red line) compared with measured state p_A (blue line). b) Estimated state $\hat{p}_B$ (red) and measured state p_B (blue). c) Parameter estimate $\hat{\alpha}$ (red) and true parameter value α (blue). d) Parameter estimate $\hat{n}$ (red) and true parameter value n (blue). The following values were used to generate this plot: $u = 1$, $\rho = 10$.

IV. PARAMETER IDENTIFICATION WITH NONLINEAR LEAST SQUARES REGRESSION

A straightforward alternative to the extended Kalman filter is a nonlinear least squares regression to estimate parameters. This approach is prone to problems when the data are noisy.

If $\dot{p}_B + p_B$ is plotted as a function of p_A, as in Fig. 1a, the nonlinear term is isolated and parameters α and n can be found by fitting the Hill function (Eqn. 3) to the data.

Simulation data are generated by integrating Eqns. 1–2 and the output is found as in Eqn. 8. The simulation parameters are picked randomly from normal distributions with mean and standard deviations given in Table I. Initial parameter guesses for the nonlinear fit are selected from the same probability distributions used to pick the parameters. The Matlab function `nlinfit` is used to do the nonlinear regression.

Fig. 3 shows that this technique works well for low noise, but the quality of the parameter estimates decreases as noise levels increase. Fig. 3a shows an example of the isolated nonlinear function. There are two sources of noise in this plot: the sensor noise and noise from numerically calculating the derivative of p_B. These problems compound as sensor noise increases. Figs. 3b and c show the error in the parameter estimate as a function of noise ρ. The nonlinear least squares fit estimates α with low error for most levels of noise. In contrast, the error in the estimate of n is quite sensitive to noise. Estimates are off by a factor of 5 at $\rho = 10^3$.

Molecular biology experiments, particularly at the single

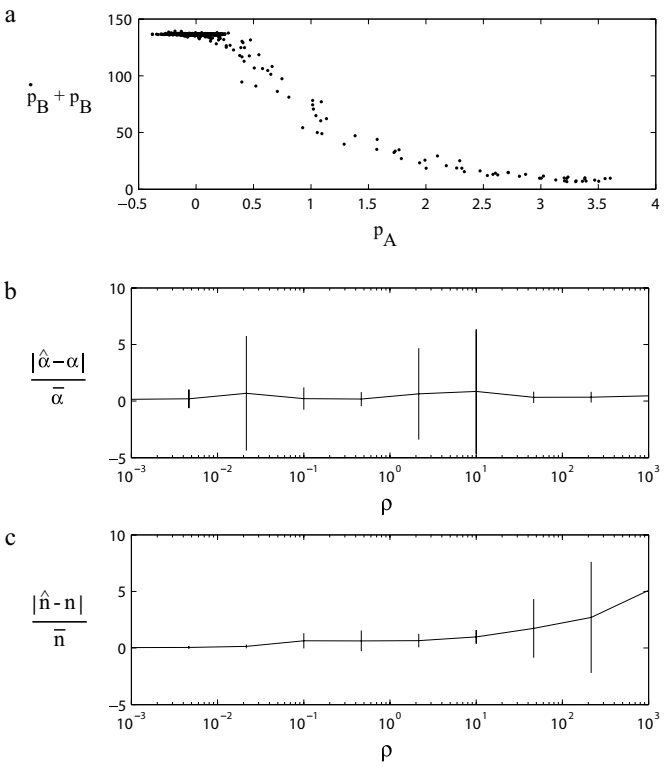

Fig. 3. Nonlinear least squares regression to find parameter estimates. a) Simulation data plotted to isolate the nonlinear function. $\rho = 0.02$ for these data. b) Normalized error in the estimate of α as a function of noise. $\hat{\alpha}$ is the estimate returned by `nlinfit`, α is the actual value used to generate the data, $\bar{\alpha}$ is the mean value used in the probability distribution given in Table I. c) Normalized error in the estimate of n as a function of noise. Notation is the same as in b). In b) and c) the figures show the mean data, averaged over 100 runs. The error bars are one standard deviation above and below the mean. The constant input $u = 1$ is used to generate all the data in these plots.

cell level, must deal with very high levels of measurement noise, so decreasing estimation error when the signal to noise ratio is low is particularly important.

V. OPTIMAL INPUT FOR PARAMETER IDENTIFICATION

A smart choice of input to the bistable switch can help in the process of parameter identification. We aim to find the optimal input for parameter estimation – the input that causes the parameter estimate to converge to the actual state quickly and accurately.

Consider two systems, a nominal system that satisfies dynamics

$$\dot{x} = F(x, u, \lambda), \tag{10}$$

where x is the state, u is the input, and λ is a vector of parameters, and a slightly perturbed system

$$\dot{x}' = F(x', u, \lambda'), \tag{11}$$

where λ' is a slightly perturbed version of λ. The forcing function, u, used to drive the perturbed system is the same as is used in the nominal system.

The input u that drives the two systems furthest apart is hypothesized to be a good choice for parameter estimation. Thus, the general optimization problem is:

$$\max_{u(t)} \int_0^{t_f} (x - x')^2 \, dt \tag{12}$$

subject to constraints on the states, parameters, and input.

To apply this optimization technique to the bistable switch system, we first establish the constraints that the optimization problem must satisfy. To be biologically realistic, protein concentrations must be positive. They must also satisfy the dynamics given in Eqns. 1-2. Inducer concentrations must also be positive (and real). By rearranging Eqn. 4, we obtain

$$I = K(\frac{\beta}{u} - 1)^{\frac{1}{m}},$$

and thus we require

$$0 < u \leq \beta.$$

Therefore, the optimization problem that we aim to solve to find optimal inputs for the parameter identification problem in the bistable switch is:

$$\max_{u(t)} \int_0^{t_f} (p_A - p_A')^2 + (p_B - p_B')^2 \, dt \tag{13}$$

subject to

$$\begin{aligned}
\dot{p}_A &= \frac{\alpha}{1 + (u\ p_B)^n} - p_A \\
\dot{p}_B &= \frac{\alpha}{1 + (p_A)^n} - p_B \\
\dot{p_A}' &= \frac{\alpha'}{1 + (u\ p_B')^{n'}} - p_A' \\
\dot{p_B}' &= \frac{\alpha'}{1 + (p_A')^{n'}} - p_B'
\end{aligned}$$

and

$$\begin{aligned}
p_A, p_B, p_A', p_B' &\geq 0 \\
0 < u &\leq \beta.
\end{aligned}$$

Define the error at time t as

$$e(t) = (p_A(t) - p_A'(t))^2 + (p_B(t) - p_B'(t))^2. \tag{14}$$

We approximate the solution to the full optimization problem by taking small time steps, and at each step choosing the constant u that maximizes the error. Since experimental measurements can only be taken at discrete times, the error should be maximized at the end of each interval. The solution is a piecewise constant $u(t)$.

For $t = \Delta t$ to t_f find

$$\max_{u \in \mathbb{R}, 0 < u \leq \beta} e(t). \tag{15}$$

The result of this optimization approximation is shown in Fig. 4.

We can also constrain the input to be constant over the entire time period and solve for the constant input that maximizes the final error

$$\max_{u \in \mathbb{R}, 0 < u \leq \beta} e(t_f). \tag{16}$$

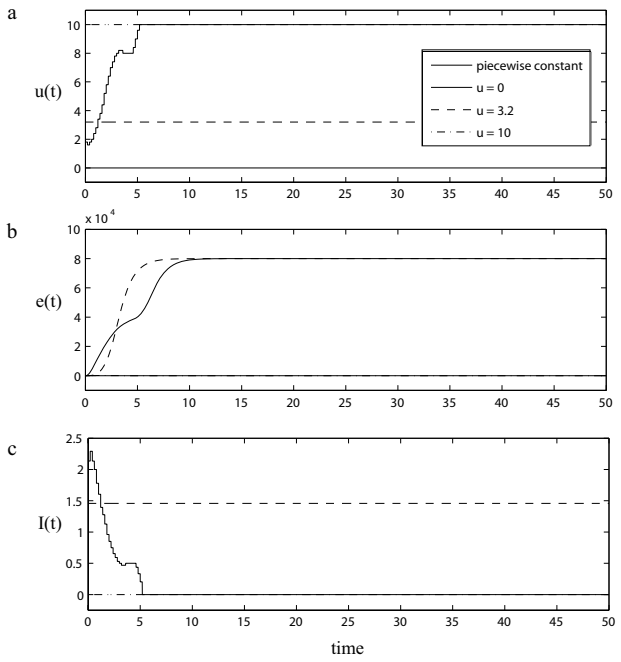

Fig. 4. a) u vs. time for piecewise constant optimization approximation (blue), $u = 0$ (green), the $u = 3.2$ (red) line is the constant input that maximizes the final error, $u = 10$ (cyan). Optimization approximations are found using the parameters $\lambda = [\alpha, n] = [200, 2]$, $\lambda' = [\alpha', n'] = [200, 3]$. b) e vs. time for each of the 4 inputs. Color coding is the same as in a). The green line falls underneath the cyan line in this plot. c) I vs. time for each input, same color coding as a). The $u = 0$ trace is omitted because it corresponds to infinite levels of inducer. In practice, any inducer concentration that saturates the system will achieve this effect. Inducer specific parameters (Eqn. 4) used to generate this figure are $\beta = 10$, $K = 1$, $m = 2$.

Fig. 4 compares the inputs found with both approximations. Fig. 4a shows the inputs, Fig. 4b shows the errors that result from the inputs, and Fig. 4c shows the inducer input required to produce the input signal. The inducer signal is particularly important because most experimental setups do not currently allow for arbitrary input signals. This limitation is discussed in further detail in Section VII.

We also consider two extreme input cases: no input ($u = \beta$) and saturating input ($u = 0$) and compare their performance to the inputs calculated with the optimization procedure.

VI. COMPARISON OF INPUTS FOR PARAMETER IDENTIFICATION

For each of the four inputs, we use the extended Kalman filter to find estimates for both α and n for a broad range of noise values. Fig. 5 summarizes how estimate errors depend upon noise. The error in α is small, even when the measurements are very noisy. This is consistent with the nonlinear fit results in Fig. 3a. The error in the estimate of n is significantly smaller for the extended Kalman filter than the nonlinear fitting method. Even in the high noise case ($\rho = 10^3$), the average estimation error is only off by a factor of 0.5, in contrast to the factor of 5 in the least squares case. For clarity, selected data points from Fig. 5 are summarized in Table II.

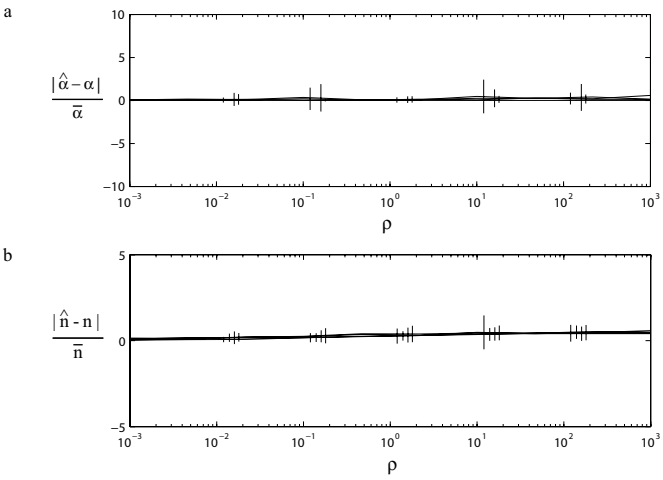

Fig. 5. Parameter identification with the extended Kalman filter for inputs $u =$ piecewise constant optimal approximation (blue), $u = 0$ (green), $u = 3.2$ (red), and $u = 10$ (cyan). a) Normalized error in the estimate of α as a function of noise. $\hat{\alpha}$ is the estimate returned by the extended Kalman filter, α is the actual value used to generate the data, $\bar{\alpha}$ is the mean value used in the probability distribution given in Table I. b) Normalized error in the estimate of n as a function of noise. Notation is the same as in a). Both figures show the mean data, averaged over 100 runs. The error bars are one standard deviation above and below the mean. Parameter estimates are the final values $\hat{\alpha}(t_f)$ and $\hat{n}(t_f)$ returned by the extended Kalman filter. $t_f = 250$ for data in these figures. Table II lists data from these plots at three values of ρ.

TABLE II
SUMMARY OF NORMALIZED ESTIMATE ERRORS

normalized α error	$\rho = 10^{-3}$ mean $\pm$ std	$\rho = 10^{0}$ mean $\pm$ std	$\rho = 10^{3}$ mean $\pm$ std
piecewise const.	0.019 $\pm$ 0.138	0.052 $\pm$ 0.277	0.585 $\pm$ 1.421
u = 0	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.009 $\pm$ 0.006
u = 3.2	0.076 $\pm$ 0.365	0.091 $\pm$ 0.423	0.146 $\pm$ 0.300
u = 10	0.057 $\pm$ 0.193	0.071 $\pm$ 0.223	0.117 $\pm$ 0.272
normalized n error	$\rho = 10^{-3}$ mean $\pm$ std	$\rho = 10^{0}$ mean $\pm$ std	$\rho = 10^{3}$ mean $\pm$ std
piecewise const.	0.027 $\pm$ 0.082	0.254 $\pm$ 0.418	0.584 $\pm$ 0.528
u = 0	0.085 $\pm$ 0.167	0.264 $\pm$ 0.272	0.492 $\pm$ 0.403
u = 3.2	0.054 $\pm$ 0.124	0.342 $\pm$ 0.477	0.406 $\pm$ 0.369
u = 10	0.145 $\pm$ 0.314	0.395 $\pm$ 0.621	0.463 $\pm$ 0.399

Estimation errors in Fig. 5 are calculated with the final values of the parameter estimates, $\hat{\alpha}(t_f)$ and $\hat{n}(t_f)$. The results are consistent for each of the four inputs.

In biological experiments, each measurement that is taken will slightly degrade the signal that is being measured. Thus, parameter estimation methods that converge quickly are better than those that require many measurements. In Fig. 6, the settling time for parameter estimation is compared for the four inputs as a function of noise.

The settling times for $\hat{n}$ are particularly sensitive to the choice of input. The inputs produced with the two optimization procedures described in Section V converge quickly. It is worth exploring other optimization methods and cost functions since there are obvious benefits for choosing inputs well.

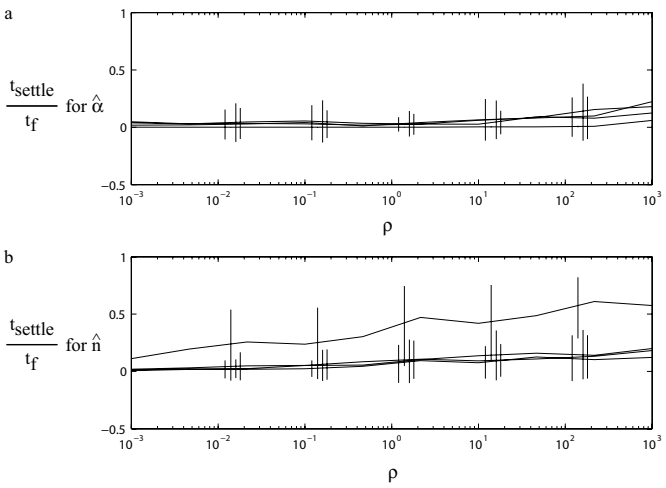

Fig. 6. Normalized settling time for parameter estimation. Colors correspond to the four inputs and are the same as in Figs. 4 and 5 a) Settling time of the estimate $\hat{\alpha}$, normalized by the final time. b) Normalized settling time for $\hat{n}$. Both figures show the mean data, averaged over 100 runs. The error bars are one standard deviation above and below the mean. $t_f = 250$ for these simulations. Settling time is defined as the time it takes the estimate to settle within 0.05μ of the final estimate, where μ is the mean value of the parameter (given in Table I).

VII. CHALLENGES UNIQUE TO BIOLOGICAL SYSTEM IDENTIFICATION

We have explored parameter identification techniques for one biological circuit, but system identification for genetic circuits remains a rich area for study. Challenges specific to biological system identification include indirect and infrequent measurements, noise and variability in the cellular environment, and a very constrained set of allowable input signals.

Measuring the "outputs" of a biological process can be difficult and one is usually limited to indirect measurements of the concentration of a small number of proteins. In this paper we have assumed that protein concentration measurements are directly available (p_A and p_B are known), but it is more common to have proteins either repress or activate the production of a "reporter", typically one that is easy to measure, such as a fluorescent protein. In addition, the amount of data that can be taken in a single experiment is limited. Each time they are measured, fluorescent proteins break down slightly. Thus, experimentalists limit the number of times they make measurements to avoid degradation of their output signal. This paper assumes that output signals can be measured continuously, but in reality data taking is sparse. One might take a data point every 10 minutes in an *E. coli* experiment, where each cell takes roughly 1 hour to divide.

The cellular environment makes system identification tasks particularly challenging. Single cells experience both intrinsic noise, noise within the cell due to the non-homogenous environment, and extrinsic noise, noise from the external environment. The relative size of these noise sources were quantified in [11] and are very important in the understanding of biological systems. Organisms behave predictably despite the stochastic nature of the environment under which they operate, so models for these processes must begin to incorporate these phenomenon. Another cellular environment specific challenge is that the process we wish to measure can sometimes have longer time constants than the cell division period, in which case our "circuit" is being replicated during its operation (this occurs in the repressilator [5], for example). Finally, the genetic circuit that we wish to study is typically not functioning independently from the rest of the cellular environment. We cannot measure a genetic circuit's behavior in isolation, so it is hard to definitively predict how it will behave in a different environment. In vitro experiments conducted outside the cell can help with isolation, but lack the biological realism of in vivo experiments.

A final challenge that must be dealt with as we work towards developing system identification techniques for biological systems is that we have only very limited ability to add inputs to the system. In this paper we have used inducer molecules as the input to the genetic circuit. Inducer addition is typically limited to constant or step-like inputs. It is not currently possible to add a smoothly varying input signal, though microchemostat work may eventually lead to breakthroughs in this area [13], [14]. It is also significantly easier to add inducer than it is to take it away. The reduction of inducer relies upon washing the cells to remove the chemical from the environment. Thus up-steps in inducer are much easier to achieve than down-steps. Simple inducer pulses were used in [4] to show low-pass filtering characteristics in transcriptional cascades, but this is one of the first examples of a non-constant input signal.

There are also a few advantages that biological systems provide, particularly at the cellular level. The first is that we can sometimes run large numbers of experiments in a relatively short period of time, by making use of the fact that cells are constantly dividing and providing new instances of a given circuit. While the cells are genetically identical, the operation of the circuit will vary based on the specific environmental conditions of that cell. This gives a natural way to do parallel experiments to explore variability in the dynamics that describe the circuit's operation. Another advantage of biological systems is that our models may not need to be as precise as in some other disciplines. Biology appears to work with large variability in individual components and hence the level of detail we need in our models may be very different than what we see in other domains.

VIII. CONCLUSION

We have considered parameter identification on a bistable genetic circuit model as an example of an interesting biological system identification problem.

Two methods of parameter identification were discussed: a nonlinear least squares fit and the extended Kalman filter. Both methods were able to accurately predict the protein production rate α for large ranges of sensor noise. The cooperativity n only affects the transient dynamics and is therefore harder to predict, especially when measurements

are noisy. The extended Kalman filter performed better than the nonlinear fit and showed significantly smaller estimation errors when sensor noises were high.

The extended Kalman filter method was also used to explore the settling time of the parameter estimate. Although all inputs that were considered ultimately caused the Kalman filter to converge to an estimate, the amount of time it took to reach this estimate depended upon the input that was used to drive the system.

We have discussed one method for choosing an input for parameter identification where the error between a nominal and perturbed system is maximized. Other cost functions and optimization strategies are worth further exploration.

IX. ACKNOWLEDGMENTS

We thank Michael Epstein, Timothy Chung, and Vijay Gupta for helpful discussions.

REFERENCES

[1] J. Hasty, D. McMillen, J.J. Collins, "Engineered Gene Circuits", *Nature*, vol. 420, 2002, pp 224-230.

[2] R. McDaniel, R. Weiss, "Advances in Synthetic Biology: On the Path from Prototypes to Applications", *Curr. Opn. Biotech.*, vol. 16, 2005, pp 476-483.

[3] T.S. Gardner, C.R. Cantor, J.J. Collins, "Construction of a Genetic Toggle Switch in *Escherichia coli*", *Nature*, vol. 403, 2000, pp 339-342.

[4] S. Hooshangi, S. Thiberge, R. Weiss, "Ultrasensitivity and Noise Propogation in a Synthetic Transcriptional Cascade", *PNAS*, vol. 102, 2005, pp 3581-3586.

[5] M.B. Elowitz, S. Leibler, "A Synthetic Oscillatory Network of Transcriptional Regulators", *Nature*, vol. 403, 2000, pp 335-338.

[6] L. Ljung, "System Identification: Theory for the User", Prentice-Hall, Inc., 1987.

[7] S.A. Billings, "Identification of Nonlinear Systems - A Survey", *IEE Proceedings*, vol. 127, 1980, Part D-Control Theory and Applications, pp 272-285.

[8] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, A. Juditsky, "Nonlinear Black-Box Modeling in System Identification: A Unified Overview", *Automatica*, vol. 31, 1995.

[9] P. Koukoulas, N. Kalouptsidis, "Nonlinear system identification using Gaussian inputs", *IEEE Transactions on Signal Processing*, vol. 43, 1995, pp 1831-1841.

[10] O. Brandman, J.E. Ferrell, R. Li, T. Meyer, "Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions", *Science*, vol. 310, 2005, pp 496-498.

[11] M.B. Elowitz, A.J. Levine, E.D. Siggia, P.S. Swain, "Stochastic Gene Expression in a Single Cell", *Science*, vol. 297, 2002, pp 1183-1186.

[12] P.S. Swain, M.B. Elowitz, E.D. Siggia, "Intrinsic and Extrinsic Contributions to Stochasticity in Gene Expression", *PNAS*, vol. 99, 2002, pp 12795-12800.

[13] F.K. Balagadde, L. You, C.L. Hansen, F.H. Arnold, S.R. Quake, "Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat", *Science*, vol. 309, 2005, pp 137-140.

[14] S. Cookson, N. Ostroff, W.L. Pang, D. Volfson, J. Hasty, "Monitoring Dynamics of Single-Cell Gene Expression Over Multiple Cell Cycles", *Molecular Systems Biology*, 2005, doi:10.10.1038/msb4100032.