33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Estimating Parameters in Genetic Regulatory Networks with SUM Logic

Li-Ping Tian, Lizhi Liu, and Fang-Xiang Wu, *Member, IEEE*

Abstract: Many methods for inferring genetic regulatory networks have been proposed. However inferred networks can hardly be used to analyze the dynamics of genetic regulatory networks. Recently nonlinear differential equations are proposed to model genetic regulatory networks. Based on this kind of model, the stability of genetic regulatory networks has been intensively investigated. Because of difficulty in estimating parameters in nonlinear model, inference of genetic regulatory networks with nonlinear model has been paid little attention. In this paper, we present a method for estimating parameters in genetic regulatory networks with SUM regulatory logic. In this kind of genetic regulatory networks, a regulatory function for each gene is a linear combination of Hill form functions, which are nonlinear in parameters and in system states. To investigate the proposed method, the gene toggle switch network is used as an illustrative example. The simulation results show that the proposed method can accurately estimates parameters in genetic regulatory networks with SUM logic.

Keywords: genetic regulatory networks, parameter estimation, toggle genetic regulatory network, SUM logic

I. Introduction

A genetic regulatory network is a complex dynamic system which describes interactions among genes (mRNA) and its products (proteins). Inferring genetic regulatory networks from time series data is a very important step towards understanding and further designing them. Many methods have been proposed to infer genetic regulatory networks, such as Bayesian Networks [1-4], probabilistic graphic model [5, 6], Boolean Networks [7-9]; and differential equations [10-12], and state space models [13]. The inferred networks with these methods can fit time series observation data very well. However, they can hardly be used to analyze the dynamics of genetic regulatory networks.

On the other hand, based on the statistic thermodynamics and biochemical reaction principle [14, 15], a genetic regulatory network can be described by a group of nonlinear differential equations [16-18]. Based on this kind of models, the stability of genetic regulatory networks has been intensively studied [18-23]. Because of their complexity, this kind of model has not been used to infer genetic regulatory networks from time series data.

In this paper, we will present a method to estimate parameters in genetic regulatory networks with SUM logic which are described by nonlinear differential equations. Section II introduces genetic regulatory networks with SUM regulatory logic [24] and discusses the properties of this model for the purpose of parameter estimation. In Section III, we will present a method for estimating the parameters in this kind of model. To investigate the proposed method, the simulation study is conducted on a toggle genetic regulatory network in Section IV. Section V gives our conclusion of this study.

II. Genetic Regulatory Networks with SUM Logic

From [18], genetic regulatory networks with n mRNAs and n proteins can be described by the following equations:

$$\begin{aligned} \dot{m}_i(t) &= -k_{mi} m_i(t) + c_i(p(t)) \\ \dot{p}_i(t) &= -k_{pi} p_i(t) + r_i m_i(t) \end{aligned} \qquad (1)$$

for i=1,2,…,n.

where $m_i(t)$, $p_i(t) \in R_+^n$ represent the concentrations of mRNA *i* and protein *i*, respectively. k_{mi} and k_{pi} are positive real numbers that represent the degradation rates of mRNA i and protein i, respectively. r_i is a positive constant representing the rate of translating mRNA i to protein i. $c_i(p(t)$ is a nonlinear function of $p_1(t), \cdots, p_n(t)$ representing the regulation function of gene i and represents the relative promoter or repressor activity of all possible proteins to gene i as a function of the concentrations $p(t)$ of all possible proteins.

The top equation in model (1) describes the transcriptional process. The bottom equation in model (1) describes the translational process. The term $r_i m_i(t)$ reflects the fact that one protein is translated from only one mRNA molecule. On the other hand, one gene or mRNA is generally activated or repressed by multiple proteins in the transcriptional process indicated in the definition of $c_i(p(t))$. In this paper, we take $c_i(p(t)) = \sum_{j=1}^{n} c_{ij}(p_j(t))$, which called the "SUM" logic [24]. The regulation function $c_{ij}(p_j(t))$ is a function of the Hill form [14, 24] as follows:

$$c_{ij}(p_j(t)) = a_{ij} \frac{1}{1 + (p_j(t)/b_j)^{h_j}}$$

if transcription factor j is a repressor of gene i, or

$$c_{ij}(p_j(t)) = a_{ij} \frac{(p_j(t)/b_j)^{h_j}}{1 + (p_j(t)/b_j)^{h_j}}$$

if transcription factor j is an activator of gene i, where b_j are positive constants, h_j are the Hill coefficient

Manuscript submitted on April 12 2010. This research is supported by Base Fund of Beijing Wuzi University and Fund for Beijing Excellent Team for Teaching Mathematics and Natural Sciences and Engineering Research Council of Canada (NSERC).

L.P. Tian is with School of Information, Beijing Wuzi University, , Beijing, P.R. China.(Email: bmitian@163.com)

L.Z. Liu is with Department of Mechanical Engineering and in the University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada.

F. X. Wu is with Department of Mechanical Engineering and Division of Biomedical Engineering in the University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada. (phone: 1(306)966-5280; fax: 1(306)966-5427; email: faw341@mail.usask.ca)

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

representing the degree of cooperativity. a_{ij} are nonnegative constants. Note that

$$\frac{1}{1+(p_j(t)/b_j)^{h_j}} = 1 - \frac{(p_j(t)/b_j)^{h_j}}{1+(p_j(t)/b_j)^{h_j}}$$

Then system (1) can be rewritten as follows

$$\begin{aligned} \dot{m}_i(t) &= -k_{m_i} m_i(t) + \sum_{j}^{n} f_{ij} g_j(p_j(t)) + l_i \\ \dot{p}_i(t) &= -k_{p_i} p_i(t) + r_i m_i(t) \end{aligned} \qquad (2)$$

for i=1,2,…,n.

where $F = (f_{ij})$ is an n×n matrix representing regulatory relationships of the network, which is defined as: $f_{ij} = 0$ if transcription factor j does not regulate gene i; $f_{ij} = a_{ij}$ if transcription factor j activates gene i; and $f_{ij} = -a_{ij}$ if transcription factor j represses gene i. l_i is a constant and is defined as $l_i = \sum_{j\in \mathrm{Re}} a_{ij}$, where *Re* is the set of repressors of gene *i*. Note that

$$g_j(u) = \frac{(u/b_j)^{h_j}}{1+(u/b_j)^{h_j}} = \frac{u^{h_j}}{b_j^{h_j} + u^{h_j}} \qquad (3)$$

Model (2) can be rewritten in the vector-matrix format as follows:

$$\begin{aligned} \dot{m}(t) &= -K_m m(t) + Fg(p(t,B)) + L \\ \dot{p}(t) &= -K_p p(t) + Rm(t) \end{aligned} \qquad (4)$$

where $m(t) = [m_1(t),\cdots,m_n(t)]^T$, $p(t) = [p_1(t),\cdots,p_n(t)]^T$, $L = [l_1,\cdots,l_n)]^T$, $g(p(t)) = [g_1(p_1(t)),\cdots,g_n(p_n(t))]^T$, $B = [b_1, \cdots,b_n)]^T$, $K_m = diag(k_{m_1},\cdots,k_{m_n})$, $K_p = diag(k_{p_1},\cdots,k_{p_n})$, and $R = diag(r_1,\ \cdots,r_n)$.

Model (2) or (4) has parameters k_{m_i}, k_{p_i}, r_i, a_{ij} and b_j (i, j=1,2,…, n), which in total are (n^2+4n). In reality, many a_{ij}'s are zeros. In this study, it is assumed that F is nonsingular and its structure is known (that is, know that which elements of F are zeros). In model (2) or (4) parameters k_{m_i}, k_{p_i}, r_i, and a_{ij} (i, j=1,2,…, n) are linear. However, b_j (j=1,2,…, n) are nonlinear in n functions of expression (3).

III. Parameter Estimation Method

In this study, assume that Hill coefficients h_j (j=1,2,..,n) are known. The task of parameter estimation for model (2) or (4) is: given data $m(t_s)$ and $p(t_s)$ measured at time point t_s (s=1,2, …, S), estimate parameters k_{m_i}, k_{p_i}, r_i, a_{ij} and b_j (i, j=1,2,…, n).

A. Parameter estimation in the translational process

To estimate parameters k_{p_i} and r_i in the bottom equation of model (2) or (4), the following squares error function (cost function) can be formed:

$$J_{p_i} = \sum_{s=1}^{S} \{\dot{\tilde{p}}_i(t_s) - [-k_{p_i} p_i(t_s) + r_i m_i(t_s)]\}^2 \qquad (5)$$

for i=1,2,…, n, where $\dot{\tilde{p}}_i$ is the estimated derivative of $\dot{p}_i$.

Given a time-series data $x(t_s)$ (s=1,2,…,S), the estimated derivative $\dot{\tilde{x}}(t_s)$ can be calculated by the following formula:

$$\dot{\tilde{x}}(t_s) = \frac{1}{12\Delta t}[x(t_{s-2}) - 8x(t_{s-1}) + 8x(t_{s+1}) - x(t_{s+2})] \qquad (6)$$

for s=3, …, S-2.

Minimizing the cost function (5) gives the estimates of parameter k_{p_i} and r_i, which can be easily done by ordinary least squares method.

B. Parameter estimation in the transcriptional process

To estimate the parameters in the top equation of model (2) or (4), the following squares error function (cost function) can be formed:

$$J = \sum_{s=1}^{S} \left\| \dot{\tilde{m}}(t_s) - [-K_m m(t_s) + Fg(p(t_s),B) + L] \right\|^2 \qquad (7)$$

Minimizing the cost function (7) can give the estimates of parameters k_{m_i}, a_{ij} and b_j (i, j=1,2,…, n). However, it is much complex to minimize the cost function (7) as parameters b_j (j=1,2,…, n) are nonlinear in the model (2) or (4). In this study, we propose a method to estimate the parameters in the top equation of model (2) or (4). From optimization principle

$$\min_{K_m,A,B} J = \min_{B} \min_{K_m,A} J \qquad (8)$$

Therefore, we propose the following iterative algorithm to estimate the parameters k_{m_i}, a_{ij} and b_j (i, j=1,2,…, n) in the top equation of model (2) or (4).

Step 1. Choose the initial guess $\hat{B}^0$ and set *k*=0

Step 2. Substitute $\hat{B}^k$ to the cost function (7) and minimize the cost function (7) with respect to k_{m_i} and a_{ij} by ordinary least squares method to get the solutions. Actually, the cost function (7) can be rewritten as $J = \sum_{i=1}^{n} J_{m_i}$ and

$$J_{m_i} = \sum_{s=1}^{S} \left\| \dot{\tilde{m}}_i(t_s) - [-k_{m_i} m_i(t_s) + \sum_{j=1}^{n} f_{ij} g_j(p(t_s),B^k) + l_i] \right\|^2 \qquad (9)$$

for i=1,2,...,n. Applying ordinary least squares method to each J_{m_i} gives the estimates of k_{m_i} and a_{ij} (j=1,2,..,n) for i=1,2,…, n . Collect all the estimates in the matrix or vector format as $\hat{K}_m^{k+1}$, $\hat{L}^{k+1}$, $\hat{F}^{k+1}$ which have the same structure as K_m, L, and F in model (4), respectively.

Step3. Substitute $\hat{K}_m^{k+1}$, L^{k+1}, F^{k+1} to the cost function (7) and minimize the cost function with respect to B.

$$\begin{aligned} J &= \sum_{s=1}^{S} \left\| \dot{\tilde{m}}(t_s) - [-\hat{K}_m^{k+1} m(t_s) + \hat{F}^{k+1} g(p(t_s,B)) + \hat{L}^{k+1}] \right\|^2 \\ &= \sum_{s=1}^{S} \left\| \hat{F}^{k+1} \{ (\hat{F}^{k+1})^{-1} [\dot{\tilde{m}}(t_s) + \hat{K}_m^{k+1} m(t_s) - \hat{L}^{k+1}] - g(p(t_s),B) \} \right\|^2 \\ &\le \left\| \hat{F}^{k+1} \right\|^2 \sum_{j=1}^{n} \bar{J}_{m_i} \end{aligned}$$

where

$$\bar{J}_{m_j} = \sum_{s=1}^{S} \left\| X_j^{k+1}(t_s) - g_j(p_j(t_s), b_i) \right\|^2 = \sum_{s=1}^{S} \left\| X_j^{k+1}(t_s) - \frac{p_j^{h_j}(t_s)}{b_j^{h_j} + p_j^{h_j}(t_s)} \right\|^2 \qquad (10)$$

and $X_j^{k+1}(t_s)$ is the j$^{\text{th}}$ component of the vector

$$(\hat{F}^{k+1})^{-1}[\dot{\tilde{m}}(t_s) + \hat{K}_m^{k+1} m(t_s) - \hat{L}^{k+1}]$$

From the above derivation, for given $\hat{K}_m^{k+1}$, L^{k+1}, F^{k+1}, the cost function (7) with respect to B can be reduced to minimizing the cost function $\bar{J}_{m_j}$ defined in (10) with respect to $b_j^{h_j}$, which is the cost function for parameter estimation of the linear fractional model. Denote the estimation of b_j by $\hat{b}_j^{k+1}$ (j=1, 2,…, n) which can be calculated as [25]

$$b_j^{k+1} = \left(\frac{1}{S} \sum_{s=1}^{S} \left[\frac{1}{X_j^{k+1}(t_s)} - 1 \right] p_j^{h_j}(t_s) \right)^{1/h_j} \qquad (11)$$

Collect all the estimates in a vector $\hat{B}^{k+1}$ which have the same structure as *B* in model (4).

Step 4. Let k=k+1 and repeat Steps 2 and 3 until a stop criterion is met. In this paper the stopping criteria is set as,

$$(J^k - J^{k+1}) / J^k \le \varepsilon$$

where $J^k = \sum_{s=1}^{S} \left\| \dot{\tilde{m}}(t_s) - [-\widetilde{K}_m^k m(t_s) + \hat{F}^k g(p(t_s), \hat{B}^k) + L^k] \right\|^2$ and ε is a preset small positive number, for example 10^{-5}.

IV. Illustrative Examples

To illustrate the performance of the presented method, we consider gene toggle switch network shown in Figure 1. In this network, two genes are repressed by each other and activated by their own protein [26].

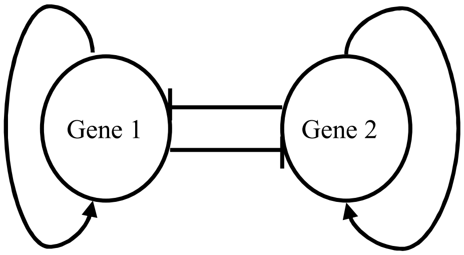

Figure 1. Structure of gene toggle switch network

The mathematical model of gene toggle switch networks can be described as follows [23, 26]:

$$\begin{aligned}
\dot{m}_1(t) &= -k_{m_1} m_1(t) + \frac{a_{11} p_1(t)/b_1}{1 + p_1(t)/b_1} + \frac{a_{12}}{1 + (p_2(t)/b_2)^2} \\
\dot{m}_2(t) &= -k_{m_2} m_2(t) + \frac{a_{21}}{1 + p_1(t)/b_1} + \frac{a_{22}(p_2(t)/b_2)^2}{1 + (p_2(t)/b_2)^2} \qquad (12) \\
\dot{p}_1(t) &= -k_{p_1} p_1(t) + r_1 m_1(t) \\
\dot{p}_2(t) &= -k_{p_2} p_2(t) + r_2 m_2(t)
\end{aligned}$$

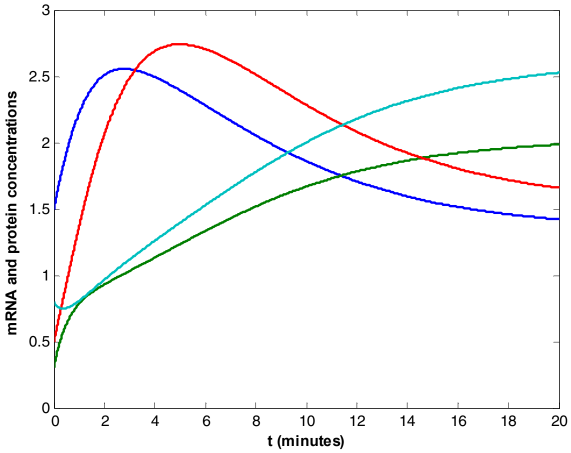

Figure 2. Trajectories of states of system (13)

In this study, the true parameter values in gene toggle network (12) are set as in Column two of Table 1. The time series data from network (13) is generated and shown in Figure 2. Network (12) is stable after 20 minutes. Therefore, we don't use the simulated data after 20 minutes. There is no noise added on the artificial data in the simulation, so they can be considered as noise-free measurements. Nevertheless, noises can be introduced in numerically calculating the derivatives by finite difference formulas (6). In general, the higher the sampling frequency and more data points are used, the more accurate the numerical derivatives are. On the other hand, in practice we may not obtain data with high frequency because of experimental limitations. In this study, the sampling frequency is 1.67Hz (100 data points per minute).

Table 1. The true parameter values in system (12), estimated parameter value, and relative estimation errors

Parameter	True value	Estimated value	REE (%)
k_{m1}	1.2	1.1533	3.89
k_{m2}	1.2	1.1287	5.94
a_{11}	2	1.9388	3.06
a_{12}	3	2.9295	2.35
a_{21}	1	0.9342	6.58
a_{22}	2.5	2.3047	7.81
b_1	2	1.8264	8.68
b_2	1.5	1.4136	5.76
k_{p1}	0.7	0.6973	0.38
k_{p2}	0.7	0.6957	0.62
r_1	0.8	0.7974	0.33
r_2	0.9	0.8944	0.62

We can apply the proposed method to estimating parameters in model (12). The relative estimation error (REE) is employed to measure the accuracy of estimation and is defined as follows:

$$REE = \frac{|estimate - true\_value|}{true\_value} \times 100\%$$

Columns 3 and 4 of Table 1 list the estimated parameter values and the relative estimation errors, respectively. From

Table 1, the relative estimation errors are less than 10% for all parameters, which indicates that the proposed method can accurately estimate parameters in this model.

V. Conclusion

In this paper, we have presented a new method for estimating parameters in genetic regulatory networks with SUM logic. In nature, the estimation of parameters in translational process of this kind of genetic regulatory networks is a nonlinear optimization problem, which cannot be easily solved. In the presented methods, the solution is given by 1) independently solving n linear least squares problems and 2) independently solving n linear fractional model parameter estimation problems. Both kinds of problems can be easily solved. The result from the illustrative example has shown that the proposed method can accurately estimate the parameters in genetic regulatory networks with SUM logic. This study assumed that matrix F in model (4) was nonsingular, which may be not true for some genetic regulatory networks. In the future, we will develop a more general method for Step 3 in Section III.

References

[1] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian Networks to Analyze Expression Data," J. Computational Biology, vol. 7, pp. 601-620, 2000.

[2] C.S. Kim, "Bayesian Orthogonal Least Squares (BOLS) Algorithm for Reverse Engineering of Gene Regulatory Networks," BMC Bioinformatics, vol. 8, article no. 251, 2007.

[3] W. Luo, K.D. Hankenson, and P.J. Woolf, "Learning Transcriptional Regulatory Networks from High Throughput Gene Expression Data Using Continuous Three-Way Mutual Information," BMC Bioinformatics, vol. 9, article no. 467, 2008.

[4] Z. Bar-Joseph, G.K. Gerber, T.I. Lee, N.J. Rinaldi, J.Y. Yoo, F. Robert, D.B. Gordon, E. Fraenkel, T.S. Jaakkola, R.A. Young, and D.K. Gifford, "Computational Discovery of Gene Modules and Regulatory Networks," Nature Biotechnology, vol. 21, no. 11, pp. 1337-1342, 2003.

[5] N. Friedman, "Inferring Cellular Network Using Probabilistic Graphical Models," Science, vol. 33, pp. 799-805, 2004.

[6] R. Ram, M.A. Chetty, "A Markov-blanket-based model for gene regulatory network inference," *IEEE/ACM Trans Comput Biol Bioinform*, 8(2): 353-67, 2011.

[7] R. Somogyi, and C. A. Sniegoski, "Modeling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation" Complexity 1: 45-63, 1996.

[8] S. Liang, S. Fuhrman, and R. Somogyi, "REVEAL, A general reverse engi-neering algorithm for inference of genetic network architectures," *Pacific Symposium on Biocomputing*, Vol. 3, pp. 18-29, 1998.

[9] T. Akutsu, S. Miyano and S. Kuhara, "Identification of gene networks from a small number of gene expression patterns under the Boolean network model," *Pacific Symposium on Biocomputing* 4: 17-28, 1999.

[10] T. Chen, H.L. He, and G. M. Church, "Modeling Gene Expression with Differential Equations" *Pacific Symposium on Biocomputing* 4: 29-40, 1999.

[11] M.J. de Hoon, S. Imoto, K. Kobayashi, N. Ogasawara, S. Miyano, "Inferring Gene Regulatory Networks from Time-Ordered Gene Expression Data of Bacillus Subtilis Using Differential Equations" *Pacific Symposium on Biocomputing* 8: 17-28, 2003.

[12] P. D'haeseleer, X Wen, S. Fuhrman and R. Somogyi, "Linear Modeling of mRNA Expression Levels During CNS Development and Injury" *Pacific Symposium on Biocomputing* 4: 41-52, 1999.

[13] FX Wu, "Gene regulatory network modelling: A state-space approach," *International Journal of Data Mining and Bioinformatics* 2(1): 1-14, 2008.

[14] J. Nielsen, J. Villadsen, and G. Liden, "Bioreaction Engineering Principles," 2nd edition, New York: Kluwer Academic/Plenum Publishers, 2003.

[15] D.M. Wolf and F.H. Eeckman, "On the relationship between genomic regulatory element organization and gene regulatory dynamics," *Journal of Theoretical Biology*, vol. 195, pp.167-186, 1998.

[16] M.B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 20, pp. 335–338, 2000.

[17] T.S. Gardner, C.R. Cantor, and J.J. Collins, "Construction of a genetic toggle switch in *Escherichia Coli*," *Nature*, vol. 403, no. 20, pp. 339–342, 2000.

[18] L. Chen and K. Aihara, "Stability of genetic regulatory networks with time delay," *IEEE Trans. Circuits Syst. I*, vol. 49, No.5, pp. 602-608, 2002

[19] F.X. Wu, "Delay-independent stability of genetic regulatory networks with time delays", Advances in Complex Systems, vol.12,no.1, pp3-19, 2009

[20] C. Li, L. Chen, and K. Aihara, "Stability of genetic networks with SUM regulatory logic: Lur's systems and LMI approach", *IEEE Trans. Circuits Syst. I*, vol. 53, No.11, pp. 2452-2458, 2006

[21] FX Wu, "Stability analysis of genetic regulatory networks with multiple time delays, *IEEE EMBC2007*: 1387-1390, 2007.

[22] F. Ren and J Cao, "Asymptotic and robust stability of genetic regulatory networks with time-varying delays", Neurocomputing 71 (2008) 834-842

[23] F.X. Wu, "Global and Robust Stability Analysis of Genetic Regulatory Networks with Time-Varying Delays and Parameter Uncertainties," *IEEE Transactions on Biomedical Circuits and Systems,* , accepted, 2011

[24] S. Kaliar, S. Mangan, U. Alon, "A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli", *Molecular Systems Biology*, 2005, doi: 10.1038/msb4100010.

[25] FX Wu, L. Mu, and ZK Shi, "Estimation of Parameters in Rational Reaction Rates of Molecular Biological Systems via Weighted Least Squares," International Journal of Systems Science, 40(1):73-80, 2010

[26] J.L. Cherry and F.R. Adler: "How to make a Biological Switch," Journal of Theoretical Biology, vol. 203, no. 2, pp. 117-133, 2000.