2011 IEEE International Workshop on Genomic Signal Processing and Statistics
December 4-6, 2011, San Antonio, Texas, USA

Intervention in General Topology Gene Regulatory Networks

Nidhal Bouaynaya*, Mohammed Rasheed*, Roman Shterenberg†, Dan Schonfeld‡
*Department of Systems Engineering
University of Arkansas at Little Rock, USA
†Department of Mathematics
University of Alabama at Birmingham, USA
‡Department of Electrical and Computer Engineering
University of Illinois at Chicago, USA

Abstract—**We present an optimal intervention framework in general topology gene regulatory networks. In particular, we do not make any assumptions about the structure or the connectivity of the initial gene network. The proposed framework finds an optimal perturbation, which forces the network to converge towards a unique desired steady-state distribution. We cast the intervention as an optimization problem, and we show that it admits at most one optimal solution. The existence of the optimal solution depends on the initial network topology and the desired steady-state distribution. In the case where no optimal solution exists, we construct a sequence of suboptimal perturbations, which converge towards a limiting optimal solution. The general topology intervention framework is applied to the Human melanoma gene regulatory network.**

I. Introduction

The advances in genomic high-throughput technologies marked a great leap in computational methods to analyze genomic data. One of the most important challenges today is to design intervention strategies in order to desirably control the behavior of the cell. An efficient control strategy must take into account all the genes involved in the biological function of interest. The collective gene interactions is modeled as a gene regulatory network having a certain topology (connectivity properties). Various control methods have been proposed in the literature since early 2000 aiming at altering the gene network in order to force the cell to a desirable behavior [1]–[6]. Optimal stochastic control theory [7] has been applied to intervene within gene regulatory networks [2], [8], [9]. Though very useful in engineering systems, optimal stochastic control requires prior knowledge of system parameters such as the target genes to be used as control variables and the cost function to be optimized for the biological system under consideration. Moreover, the finite-horizon optimal control may not alter the long-run behavior of the gene regulatory network. In biology, we are mainly interested in the long-term effects, hence long-run behavior, rather than the transit performance of the system.

The long-run behavior of a gene regulatory network, modeled as a homogeneous Markov chain process [1], is assessed by its stationary distribution(s), also called steady-state distribution(s). Stationary distributions are the fixed points of the Markov chain. That is, if the network states probability distribution is given by the stationary distribution, then the network will follow this distribution forever after. Qian and Dougherty [3] studied the effects of function perturbations on the steady-state distribution of ergodic probabilistic Boolean networks. Their study, however, focuses on rank one perturbations, and its extension to higher rank perturbations is iterative and computationally cumbersome.

Recently, Bouaynaya et al. [4]–[6] formulated the optimal control problem as an inverse perturbation problem, where the aim is to find a (optimal) perturbation that persuade the network to transition towards a desired steady-state distribution. The optimal inverse perturbation problem is casted as a convex optimization problem, thus leading to a globally optimum, non-iterative solution, which can be computed efficiently using standard convex optimization techniques [5], [10]. However, as with all the literature on this topic [2], [3], they consider only ergodic networks. The ergodicity assumption, per contra, may not hold for many gene regulatory networks. Intuitively, one would expect large gene regulatory networks to be non-ergodic, either reducible or periodic.

In this paper, we extend the inverse perturbation framework to general topology networks. In particular, we do not assume that the gene regulatory network is ergodic or has any particular structure. The only assumption is that the considered network can be modeled as a homogenous Markov chain process. In particular, the dynamics of (probabilistic) Boolean networks and Bayesian networks are modeled by homogeneous Markov chains [11].

This paper is organized as follows: In Section II, we introduce the general topology intervention problem. We first investigate the feasibility of the intervention. Subsequently, we derive the optimal intervention strategy as the minimal "energy" perturbation, which forces the network to converge towards the desired distribution. We show that the optimal intervention problem admits at most one solution. In Section III, we apply the proposed framework to the Human melanoma gene regulatory network, where we consider two different desired steady-state distributions. Finally, we summarize the main contributions of this paper in Section IV.

978-1-4673-0490-0/11/$26.00 ©2011 IEEE

II. The General Topology Inverse Perturbation Problem

We consider a gene regulatory network with m genes, where the expression level of each gene is quantized to l values. The dynamics of the network is represented by a finite homogeneous Markov chain, where the transition probability matrix P_0 is of size $n = l^m$. A stationary or steady-state distribution of P_0 is defined as a (column) vector $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi}^t P_0 = \boldsymbol{\pi}^t$. Stationary distributions always exist because the probability transition matrix is stochastic (i.e., its rows sum up to 1). If the Markov chain is ergodic, P_0 has a unique stationary distribution, $\boldsymbol{\pi}_0$, and the Markov chain converges to the matrix whose rows are equal to the stationary distribution, i.e., we have

$$\lim_{n\to\infty} P_0^n = \mathbf{1}\boldsymbol{\pi}_0^t \tag{1}$$

By abuse of terminology, we will say that P_0 converges towards the steady-state distribution $\boldsymbol{\pi}_0$.

In the general case, the Markov chain may have multiple stationary distributions or a unique stationary distribution but fails to converge. For instance, consider the matrix

$$P_0 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{2}$$

P_0 has a unique stationary distribution $\boldsymbol{\pi}_0 = [\,\frac{1}{3},\ \frac{1}{3},\ \frac{1}{3}]^t$. But P_0 does not converge to $\mathbf{1}\boldsymbol{\pi}_0^t$. This is because, even though P_0 is irreducible, it is periodic. For the gene network application, it is not enough to ensure that the perturbation introduces the desired distribution as a stationary distribution of the perturbed network, but we also require convergence of the perturbed network towards the desired distribution.

Let us consider the perturbed matrix

$$P = P_0 + C, \tag{3}$$

where C is zero-row sum perturbation matrix. Denote by $\boldsymbol{\pi}_d > \mathbf{0}$ the desired distribution of the network states. Our goal is to find a (optimal) perturbation matrix C such that the perturbed matrix P converges towards $\boldsymbol{\pi}_d$ as its unique stationary distribution. In particular, no assumptions are made about the structure or the connectivity of P_0.

A. The feasible control

The set of perturbation matrices, C, which force the network to converge towards the desired steady-state distribution $\boldsymbol{\pi}_d > \mathbf{0}$ satisfy the following four constraints:

(i) SLEM $(P_0 + C) < 1$
(ii) $\boldsymbol{\pi}_d^t(P_0 + C) = \boldsymbol{\pi}_d^t$
(iii) $C\mathbf{1} = \mathbf{0}$
(iv) $P_0 + C \geq \mathbf{0}$,

where constraint (iv) denotes an elementwise inequality. SLEM stands for second largest eigenvalue modulus, where the eigenvalues are counted taking into account their multiplicity. In particular, (i) implies that eigenvalue 1 is simple. Along with the positivity of the desired distribution, constraint (i) is equivalent to ergodicity.

Let $\mathcal{F}$ be the feasible set of perturbation matrices, i.e, $\mathcal{F}$ is the set of matrices C satisfying constraints (i) through (iv). $\mathcal{F} \neq \emptyset$ since $(\mathbf{1}\boldsymbol{\pi}_d^t - P_0) \in \mathcal{F}$. In fact SLEM $(\mathbf{1}\boldsymbol{\pi}_d^t) = 0 < 1$, and it is easy to check that $(\mathbf{1}\boldsymbol{\pi}_d^t - P_0)$ satisfies conditions (ii)-(iv). Therefore, the feasible set is not empty, and we can find at least one feasible perturbation matrix, which forces the network to converge towards a desired steady-state distribution.

B. The optimal control

We consider the Frobenius norm as our optimality criterion. The Frobenius norm criterion minimizes the "energy" between the original and perturbed networks. The minimum energy constraint is imposed to limit the structural changes in the network before and after control. The optimal general-topology inverse perturbation control is therefore formulated as follows:

$$\text{Minimize} \quad \|C\|_F^2 \quad \text{subject to} \quad C \in \mathcal{F}, \tag{4}$$

where $\|.\|_F$ denotes the Frobenius norm given by $\|C\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n c_{ij}^2 = \text{Tr}(CC^t)$, and $\mathcal{F}$ is the feasible set defined by constraints (i)-(iv).

A strictly convex function admits at most one minimizer. In general, the optimal solution belongs to the closure of the set, $\bar{\mathcal{F}}$. Consider the closed convex set $\mathcal{D} = \{C \in \mathbb{R}^{n\times n} : \boldsymbol{\pi}_d^t = \boldsymbol{\pi}_d^t(P_0 + C), C\mathbf{1} = \mathbf{0}, P_0 + C \geq \mathbf{0}\}$. That is $\mathcal{D}$ is the set of perturbation matrices satisfying conditions (ii), (iii), and (iv) only. We have $\mathcal{F} \subseteq \mathcal{D}$. Denote by C_* the minimum Frobenius norm perturbation matrix over the set $\mathcal{D}$. We know that C_* exists and is unique because the Frobenius norm is strictly convex and the set $\mathcal{D}$ is convex and closed. Moreover, C_* can be computed efficiently as the solution of a semi-definite programming algorithm [5]. The following proposition shows that C_* is the optimum solution of the optimization problem in (4) if $C_* \in \mathcal{F}$.

Proposition 1. *Let $C_* = argmin_{C\in\mathcal{D}} \|C\|_F^2$. Then, $C_* \in \bar{\mathcal{F}}$. Moreover, if $C_* \in \mathcal{F}$, then it is the unique optimal solution of (4).*

Since $C_* \in \bar{\mathcal{F}}$, there exists a sequence $C_n \in \mathcal{F}$ such that C_n converges towards C_* and $\|C_n\|_F > \|C_*\|_F$. The following proposition provides a construction of such a sequence.

Proposition 2. *Assume that $C_* \notin \mathcal{F}$. Consider the family of matrices described by*

$$C_n = (1 - \epsilon_n)C_* + \epsilon_n(\mathbf{1}\boldsymbol{\pi}_d^t - P_0), \tag{5}$$

where $0 < \epsilon_n \leq 1$ is a sequence converging to zero. Then, we have

1) $C_n \in \mathcal{F}$, $\forall n \in \mathbb{N}$.
2) $C_n \to C_*$
3) $\|C_n\|_F > \|C_*\|_F$, $\forall n \in \mathbb{N}$.

From Proposition 1, if SLEM$(P_0 + C_*) = 1$, then the optimization problem in (4) has no solution. SLEM$(P_0 + C_*)$

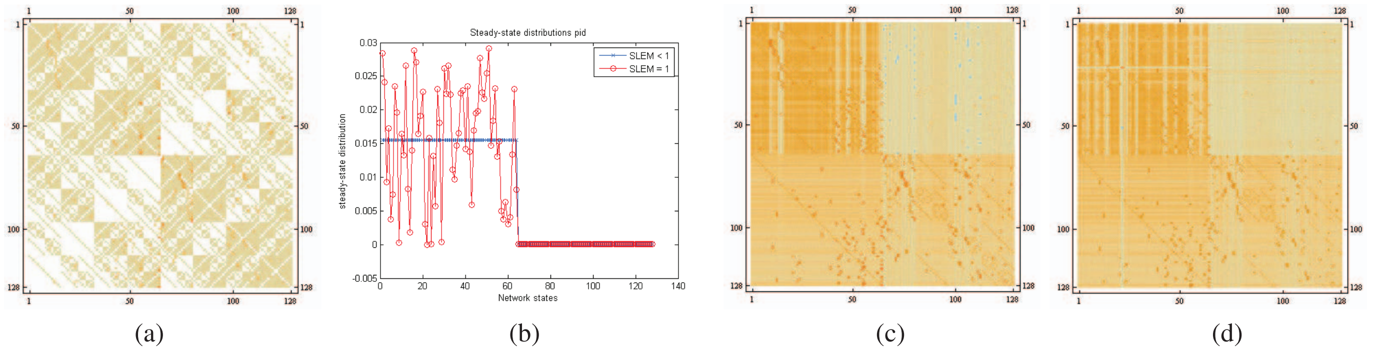

Fig. 1. Optimal inverse perturbation of the Human melanoma gene regulatory network. The matrix plots are obtained using the function *MatrixPlot* in MATHEMATICA. They provide a visual representation of the values of elements in the matrix. The color of entries varies from white to red corresponding to the values of the entries in the range of 0 to 1. (a) The probability transition matrix, P_0, of the melanoma gene regulatory network; (b) Two different steady-state distributions, which correspond to a downregulation of the gene WNT5A; (c) The perturbed matrix P corresponding to the steady-state distribution in blue; (d) The perturbed matrix P corresponding to the steady-state distribution in red;

depends on the probability matrix P_0 and the desired steady-state $\boldsymbol{\pi}_d$.

In summary, for a given initial network topology, there exists at most one minimal energy perturbation matrix, which forces the network to converge towards a unique desired steady-state distribution. The existence of the optimal perturbation depends on the specific values of the transition matrix P_0 and the desired steady-state distribution $\boldsymbol{\pi}_d$. If the minimizer does not exist, Proposition 2 provides a sequence of matrices, which tend towards the limiting optimal minimizer, and therefore can be considered as suboptimal solutions.

III. Simulation Results

We consider the Human melanoma gene regulatory network [12] [2], [3]. The abundance of mRNA for the gene WNT5A was found to be highly discriminating between cells with properties typically associated with high versus low metastatic competence. Furthermore, it was found that an intervention that blocked the Wnt5a protein from activating its receptor, the use of an antibody that binds the Wnt5a protein, could substantially reduce Wnt5A's ability to induce a metastatic phenotype [12]. This suggests a control strategy that reduces WNT5A's action in affecting biological regulation.

A seven-gene probabilistic Boolean network model of the melanoma network containing the genes WNT5A, pirin, S100P, RET1, MART1, HADHB, and STC2 was derived in [13]. The Human melanoma Boolean network consists of $2^7 = 128$ states ranging from $00\cdots0$ to $11\cdots1$, where the states are ordered as WNT5A, pirin, S100P, RET1, MART1, HADHB, and STC2, with WNT5A and STC2 denoted by the most significant bit (MSB) and least significant bit (LSB), respectively.

Since the aim is to downregulate the WNT5A gene, the states from 64 to 127, which correspond to WNT5A upregulated, should have near zero steady-state mass. In our simulations, we consider two different desired steady-state distributions $\boldsymbol{\pi}_d^1$ and $\boldsymbol{\pi}_d^2$, shown in Fig. 1(b). The first distribution, $\boldsymbol{\pi}_d^1$, assigns probability 10^{-4} to the states having WNT5A upregulated and a uniform mass equal to 0.015525 to the other states. The second distribution, $\boldsymbol{\pi}_d^2$, also assigns a uniform mass of 10^{-4} to the undesirable states but assigns random probabilities to the other states such that the total probability mass is equal to 1. The first and second steady-state distributions are plotted in blue and red, respectively, in Fig. 1(b). The corresponding optimal perturbed transition matrices are depicted in Figs. 1(c) and 1(d). Let us denote them by P_1 and P_2, respectively. Observe that the original transition matrix remains unchanged as P_0 (see Fig. 1(a)). We have SLEM$(P_1) < 1$, whereas SLEM$(P_2) = 1$. Therefore, for the desired stationary distribution $\boldsymbol{\pi}_d^1$, there exists an optimal perturbation matrix, which forces the network to converge towards $\boldsymbol{\pi}_d^1$, whereas there exists no optimal solution if the chosen desired steady-state is given by $\boldsymbol{\pi}_d^2$.

We now consider the desired steady-state distribution $\boldsymbol{\pi}_d^2$, which corresponds to a SLEM$(P_0 + C_*) = 1$ and hence non-ergodic perturbed matrix P_2. Proposition 2 states that the corresponding sequence of perturbation matrices C_n, given by Eq. (5), correspond to ergodic perturbed matrices, which converge towards $\boldsymbol{\pi}_d^2$. In Fig. 2, we plotted SLEM$(P_0 + C_n)$ and $\|C_n\|_F$ versus ϵ_n. Observe that the SLEM is a decreasing function, whereas the Frobenius norm increases with ϵ_n. In particular, given a $\delta > 0$, there exists $\epsilon_n > 0$ such that $\|C_n - C_*\| < \delta$, and $C_n \in \mathcal{F}$. Therefore, C_n can be considered as suboptimal solutions to the inverse perturbation problem in (4).

IV. Conclusion

In this paper, we presented a framework to the optimal inverse perturbation problem for general topology networks. In particular, we do not impose the ergodicity assumption on the original probability transition matrix. Intuitively, we expect large gene regulatory networks to be non-ergodic, i.e., not strongly connected or periodic. The general topology inverse perturbation framework addresses the issue of finding a (optimal) perturbation, which forces the network to converge towards a unique desired steady-state distribution. We showed that this problem has at most one solution. The existence of the optimal perturbation depends on the original transition

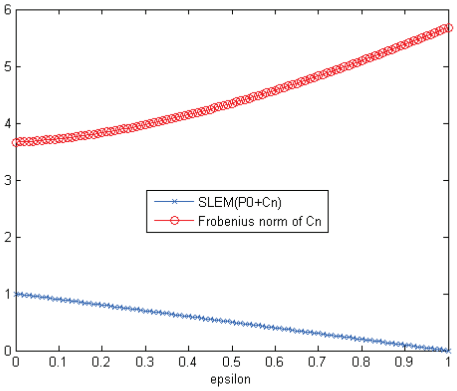

Fig. 2. SLEM $(P_0 + C_n)$ versus ϵ_n (blue), and $\|C_n\|_F$ versus ϵ_n (red), where C_n is given by Eq. (5).

matrix and the desired steady-state distribution. If the optimal perturbation does not exist, we constructed a sequence of suboptimal perturbations, which converge towards a limiting optimal solution.

Appendix

Proof of Proposition 1: Since $P_0 + C_*$ is a stochastic matrix, we have SLEM$(P_0 + C_*) \leq 1$. Therefore, $C_* \in \bar{\mathcal{F}}$. If SLEM$(P_0 + C_*) < 1$, then $C_* \in \mathcal{F}$ and since $\mathcal{F} \subseteq \mathcal{D}$, C_* is also the optimal solution of the optimization problem in (4). ■

Proof of Proposition 2: We will prove each of the three facts separately.

1) We need the following Lemma from [5]

   Lemma 1. *[5] Let*

$$P_n = P_0 + C_n = (1 - \epsilon_n)P_* + \epsilon_n \mathbf{1}\boldsymbol{\pi}_d^t, \qquad (6)$$

   where $P_* = P_0 + C_*$ and $0 < \epsilon_n \leq 1$. Then, we have

$$\textit{SLEM}\ (P_n) = (1 - \epsilon_n)\ \textit{SLEM}\ (P_*). \qquad (7)$$

   The proof of Lemma 1 is provided in [5]. Since P_* is a stochastic matrix, we have SLEM$(P_*) \leq 1$. Therefore, from Lemma 1, we obtain SLEM$(P_n) < 1$ for $0 < \epsilon_n \leq 1$.
2) The fact that $C_n \to C_*$ is obvious given that $\epsilon_n \to 0$.
3) We have $C_n \in \mathcal{F} \subseteq \mathcal{D}$, and C_* is the unique Frobenius norm minimizer over $\mathcal{D}$. Hence, $\|C_n\|_F \geq \|C_*\|$. Moreover, $C_* \notin \mathcal{F}$. In particular, $C_n \neq C_*$, $\forall\ n \in \mathbb{N}$. Therefore, we have strict inequality $\|C_n\|_F > \|C_*\|$, $\forall\ n \in \mathbb{N}$.

■

Acknowledgment

This project is supported by Award Number R01GM096191 from the National Institute Of General Medical Sciences (NIH/NIGMS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health.

The authors would like to extend their gratitude to Dr. Ranadip Pal from Texas Tech University for providing the Human melanoma gene regulatory network dataset.

References

[1] S. Kim, H. Li, E. R. Dougherty, N. Cao, Y. Chen, M. Bittner, and E. B. Suh, “Can Markov chain models mimic biological regulation?” *Journal of Biological Systems*, vol. 10, no. 10, pp. 337–357, 2002.

[2] A. Datta, R. Pal, A. Choudhary, and E. Dougherty, “Control approaches for probabilistic gene regulatory networks,” *IEEE Signal Processing Magazine*, vol. 24, no. 1, pp. 54–63, 2007.

[3] X. Qian and E. R. Dougherty, “Effect of function perturbation on the steady-state distribution of genetic regulatory networks: Optimal structural intervention,” *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 4966–4976, October 2008.

[4] N. Bouaynaya, R. Shterenberg, and D. Schonfeld, “Optimal perturbation control of gene regulatory networks,” in *IEEE International Workshop on Genomic Signal Processing and Statistics*, November 2010.

[5] ——, “Inverse perturbation for optimal intervention in gene regulatory networks,” *Bioinformatics*, vol. 27, no. 1, pp. 103–110, 2011.

[6] ——, “Robustness of inverse perturbation for discrete event control,” in *IEEE International Conference of the Engineering in Medicine and Biology Society*, August 2011.

[7] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 2007.

[8] A. Datta, A. Choudhary, M. L. Bittner, and E. R. Dougherty, “External control in markovian genetic regulatory networks,” *Machine Learning*, vol. 52, pp. 169–191, 2003.

[9] R. Pal, A. Datta, and E. Dougherty, “Optimal infinite horizon control for probabilistic Boolean networks,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2375–2387, 2006.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2003.

[11] H. Lhdesmkia, S. Hautaniemia, I. Shmulevichc, and O. Yli-Harjaa, “Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks,” *Signal Processing*, vol. 86, no. 4, pp. 814–834, 2006.

[12] M. Bittner, P.Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E.Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, and V. Sondak, “Molecular classification of cutaneous malignant melanoma by gene expression profiling,” *Nature*, vol. 406, no. 6795, pp. 536 – 540, 2000.

[13] R. Pal, I. Ivanov, A. Datta, and E. R. Dougherty, “Generating Boolean networks with a prescribed attractor structure,” *Bioinformatics*, vol. 21, pp. 4021 – 4025, 2005.