STRUCTURAL INTERVENTION OF GENE REGULATORY NETWORKS BY GENERAL RANK-K MATRIX PERTURBATION

Xiaoning Qian[1]*, Byung-Jun Yoon*[2] *and Edward R. Dougherty*[2,3,4]

[1]Dept. of Computer Science & Engineering, University of South Florida, Tampa, FL 33620
[2] Dept. of Electrical & Computer Engineering, Texas A&M University, College Station, TX 77843
[3] Computational Biology Division, Translational Genomics Research Institute, Phoenix, AZ 85004
[4] Dept. of Bioinformatics and Computational Biology,
University of Texas M. D. Anderson Cancer Center, Houston, TX 77030

ABSTRACT

One of the ultimate objectives of studying gene regulatory networks is to derive potential intervention strategies to avoid aberrant cellular behavior. Boolean networks (BNs) and their stochastic extension, probabilistic Boolean networks (PBNs), provide a convenient framework to design different types of intervention strategies. In this paper, we focus on studying structural intervention, in which we perturb regulatory Boolean functions to alter the long-term network dynamics to obtain desirable behavior. Specifically, we extend our previous work that derives optimal structural intervention for rank-1 function perturbations to more general solutions for arbitrary rank-*k* function perturbations. The analytic solution is derived using the Sherman-Morrison-Woodbury (SMW) formula. We apply the derived structural intervention to a mutated mammalian cell cycle network. Our results show that our intervention strategy correctly identifies the main targets to stop uncontrolled cell growth in the mutated cell cycle network.

Index Terms— Genetic regulatory network, Boolean network, probabilistic Boolean network, structural intervention, rank-*k* matrix perturbation, Sherman-Morrison-Woodbury

1. INTRODUCTION

Cellular functions arise as the results of the coordinated interactions among genes and gene products [1, 2]. To design future gene-based therapeutics for complex diseases including cancer, appropriate mathematical models and corresponding computational tools for systematic analysis of these complex interactions are critical. Due to the paucity of the relevant data and the scarcity of information regarding the underlying regulatory mechanisms, coarse models for gene regulatory networks, including Boolean networks (BNs) [1] and probabilistic Boolean networks (PBNs) [3], appear to be a promising avenue to model, simulate, and alter the systematic behavior of biological processes [2, 4, 5].

In PBNs, the underlying network dynamics can be modeled as a finite Markov chain, which enables systematic analysis using the classical Markov chain theory [6, 7]. Current intervention strategies in PBNs can be categorized into two basic types: The first is commonly known as *state perturbation*, in which control policies are derived to force gene expression changes by external control to modulate the dynamics [2, 8–10]. The second type is *structural* or *function perturbation* [2, 11–13], which has a more fundamental impact by altering the underlying rule-based structures of PBNs to change the dynamics permanently so that it alters the long-term abnormal network behavior. In our previous work [13], we have derived an analytical solution to compute the perturbed steady-state distribution (SSD) by function perturbation of regulatory rules. However, the previous solution was limited to rank-1 perturbations with the computationally expensive iterative procedure as the extension for general rank-*k* perturbations.

In this work, we generalize our previous work to derive an analytic solution to compute the long-term dynamic change caused by general function perturbation to PBNs, which can be used to derive structural intervention strategies to achieve desired behavior and reduce the risk of entering into aberrant phenotypes. The derivation for computing perturbed SSDs for general rank-*k* function perturbations is based on a matrix inverse operator using the *Sherman-Morrison-Woodbury (SMW) formula* [14]. With the derived analytic solution, we study a mutated mammalian cell cycle network to show that we may derive structural intervention strategies to identify critical intervention targets.

2. BACKGROUND

Dynamics in PBNs are modeled by regulatory Boolean functions and the related probabilistic parameters. For a binary PBN with n genes, each gene i has its gene expression state quantized to two levels: $x_i \in \{0,1\}$ whose temporal state transition is determined by the values of some other genes via a predictor Boolean function $f_i : \{0,1\}^{K_i} \mapsto \{0,1\}$, where

K_i is the input degree of x_i in the network. The predictor Boolean functions for all the genes determine the network dynamics. There are two commonly accepted transition rules: "majority vote" rule [15], and "strong inhibition" rule [16]. State transitions can either be synchronous or asynchronous [16]. With probabilistic parameters, we can derive the corresponding state transition diagram and analyze the network dynamics via the finite Markov chain theory. The underlying Markov chain of a PBN is irreducible and ergodic. Hence, the network possesses a steady-state distribution (SSD)π such that $\pi^T = \pi^T P$, where P is its transition matrix determined by the network model, and T denotes transpose. The SSD reflects the long-term behavior of a PBN and the change of SSD by different types of perturbations may guide the design of beneficial intervention strategies [2, 8, 13].

We study structural perturbation to model the siRNA interference of regulatory relationships, in which we assume that we can block the regulation between any two genes in the network. Mathematically, the perturbed network will have a perturbed transition matrix $\tilde{P}$ with a new SSD $\tilde{\pi}$. Without loss of generality, for intervention, we partition the network state space into sets U and D of *undesirable* and *desirable states*, often according to the expression states of a given set of genes. We aim to find the optimal structural intervention which gives the minimum $\tilde{\pi}_U = \sum_{\mathbf{x}\in U} \tilde{\pi}_{\mathbf{x}}$, where $\tilde{\pi}_{\mathbf{x}}$ denotes the perturbed steady-state probability at a given state $\mathbf{x}$.

We have recently derived analytical solutions for structural perturbations with simple forms, including 1-bit function perturbations [13]. These perturbations can be mathematically represented by a perturbation matrix E with $\tilde{P} = P + E$. For 1-bit function perturbation, E is a rank-1 matrix: $E = ab^T$. To maintain $\tilde{P}$ as a stochastic matrix, a, b are two arbitrary vectors satisfying $b^T e = 0$ with e as a column vector with all its entries equal to 1. In [13], more general function perturbations, for example, with rank-k matrix E, have to be solved iteratively. Within each iteration, the fundamental matrix $\tilde{Z}$ of the perturbed underlying Markov chain has to be updated together with the SSD $\tilde{\pi}$ based on the following equations:

$$\tilde{\pi}^T = \pi^T + \frac{\pi^T a}{1 - b^T Z a} b^T Z, \tag{1}$$

$$\tilde{Z} = \left[I - \frac{(\pi^T a) e b^T Z}{1 - b^T Z a} \right] \left[Z + \frac{Z a b^T Z}{1 - b^T Z a} \right], \tag{2}$$

where $Z = (I - P + e\pi^T)^{-1}$ is the fundamental matrix corresponding to the original P before perturbation. For each iteration, the time complexity is $O(2^n)$ where n is the number of genes in the network. When the perturbation matrix has a higher rank, the solution will have similar complexity as the power method for computing the perturbed SSD.

3. METHODS

In this work, we derive analytic solutions to efficiently compute the perturbed SSD for more general rank-k function perturbations $E = UV^T$, in which UV^T is a rank-k matrix in the decomposed version: U and V are both $2^n \times k$ matrices satisfying the constraints to guarantee that $\tilde{P}$ remains a stochastic matrix. For the specific forms of U and V matrices from function perturbations to PBNs, we can derive them similarly as in our previous work [13]. Typically, k is much smaller comparing to 2^n for local structural or function perturbations to PBNs. Now, we derive analytic solutions for the perturbed SSD caused by an arbitrary rank-k function perturbation.

As the underlying Markov chain of a PBN is irreducible and ergodic, both the original and perturbed PBN have their SSDs satisfying: $\pi^T = \pi^T P$, and $\tilde{\pi}^T = \tilde{\pi}^T \tilde{P}$. We can write out the change of SSDs caused by perturbation:

$$\begin{aligned} & \tilde{\pi}^T - \pi^T = \tilde{\pi}^T \tilde{P} - \pi^T P = (\tilde{\pi} - \pi)^T P + \tilde{\pi}^T E \\ \Rightarrow \quad & (\tilde{\pi} - \pi)^T (I - P) = \tilde{\pi}^T E, \end{aligned} \tag{3}$$

where I is the identity matrix. Further, we notice that $(\tilde{\pi} - \pi)^T P^\infty = (\tilde{\pi} - \pi)^T e \pi^T = 0$ as for an arbitrary SSD π, we have $\pi^T e = 1$. Therefore, $(\tilde{\pi} - \pi)^T (I - P) = (\tilde{\pi} - \pi)^T (I - P + e\pi^T) = \tilde{\pi}^T E$. Multiplying both sides by the fundamental matrix Z, we have $(\tilde{\pi} - \pi)^T (I - P + P^\infty) Z = \tilde{\pi}^T E Z$, and $(\tilde{\pi} - \pi)^T = \tilde{\pi}^T E Z$. Hence, $\tilde{\pi}^T (I - EZ) = \pi^T$ and

$$\tilde{\pi}^T = \pi^T (I - EZ)^{-1}. \tag{4}$$

Similar results have been given in [7]. For a fully characterized PBN with given π and Z, the task to compute a perturbed SSD by the function perturbation E lies at efficient computation of the inverse matrix $(I - EZ)^{-1}$.

Now, we introduce the SMW formula for computing the inverse matrix with the general form $(A + BC^T)^{-1}$:

$$(A + BC^T)^{-1} = A^{-1} - A^{-1} B (I + C^T A^{-1} B)^{-1} C^T A^{-1}. \tag{5}$$

Recalling that $E = UV^T$, we have $(I - EZ)^{-1} = (I - UV^T Z)^{-1}$. Let $A = I$, $B = U$, and $C = V^T Z$ in (5), we get

$$(I - EZ)^{-1} = I + U (I - V^T Z U)^{-1} V^T Z. \tag{6}$$

Note that $(I - EZ)$ is a $2^n \times 2^n$ matrix. Using the SMW formula, the inverse of this large matrix has been transformed to the inverse of $(I - V^T Z U)$, in which U and V are $2^n \times k$ matrix and therefore $(I - V^T Z U)$ is a $k \times k$ matrix. Hence, for an arbitrary rank-k function perturbation, the determinant factor for the computational complexity to compute $\tilde{\pi}$—the inverse of $(I - EZ)$—has been greatly reduced from $O(2^{3n})$ to $O(k^3)$ when $k \ll 2^n$: $\tilde{\pi}^T = \pi^T + \pi^T U (I - V^T Z U)^{-1} V^T Z$. Our new derived analytic solution for the perturbed SSD based on the SMW formula is much more efficient comparing to either the power method or the iterative method proposed in [13].

4. EXPERIMENTS AND DISCUSSIONS

We implement the derived method to a mutated mammalian cell cycle network. The original cell cycle network [5] has $n = 10$ genes (CycD, Rb, p27, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, and CycB) and the regulatory relationships are given in Figure 1. This network models mammalian cell division that is controlled via extra-cellular signals and coordinates with overall growth, in which Rb will be expressed to coordinate the cell cycle in the absence of the cyclins. Gene

p27 can stop the uncontrolled cell cycle as it blocks the action of CycE or CycA and thereafter Rb can also be expressed, even in the presence of CycE or CycA. In [10], a mutated cell cycle network was derived based on this to demonstrate the effectiveness of mathematically designed intervention using PBNs. In the mutated network, p27 is mutated and is always off. This models the cancerous scenario, in which p27 can never be activated. Under this mutated situation, the cell will cycle in the absence of any growth factor when both CycD and Rb are inactive, which represents a cancerous phenotype. We model this mutated network by a PBN with x_i denoting the expression state for each of the nine genes without p27 in the order of CycD, Rb, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, and CycB. The dashed lines in Figure 1 correspond to inactive interactions due to the mutation of p27. We consider the logic states in which both Rb and CycD are down-regulated as undesirable states: $U = \{\mathbf{x} \mid x_1 = 0, x_2 = 0\}$. To model it as a PBN, we set the perturbation probability for each gene expression state $p = 0.01$ and derive the logic Boolean transition functions based on the "majority vote" rules[15]. In this case, we first define the regulatory relationship between gene i and j as follows:

$$R_{j\to i} = \begin{cases} 1 & j \text{ activates } i \\ 0 & j \text{ does not regulate } i \\ -1 & j \text{ suppresses } i \end{cases}$$

The transition for x_i is determined by:

$$f_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_j R_{j\to i} x_j > 0 \\ x_i & \text{if } \sum_j R_{j\to i} x_j = 0 \\ -1 & \text{if } \sum_j R_{j\to i} x_j < 0, \end{cases} \quad \forall i.$$

We compute the SSD for this mutated cell cycle network as shown in Figure 2(A) and the undesirable mass $\pi_U = \sum_{x_1=0,x_2=0} \pi_{\mathbf{x}} = 0.4591$. The class of structural intervention strategies is to perturb the network to block the suppressing or activating regulatory action between any regulatory pair of genes in the original mutated network (by setting $R_{j\to i} = 0$). Based on the fully specified network model, we can analytically compute the perturbed SSD using the derived solution for the structural perturbation by blocking each single existing interaction. We rank all the existing regulatory interactions based on the reduction of the undesirable SSD mass. Using different colors, Figure 1 illustrates the group of the edges, on which our structural perturbation reduces the undesirable SSD mass. With mutated p27, the suppressing regulatory action from CycE to Rb (in red) indeed leads to the most reduced undesirable mass ($\tilde{\pi}_U = 0.4026$) and hence is the optimal structural intervention target if we block only one regulation. The second and third best perturbations are to block the activating regulation from E2F to CycE, and the suppressing regulation from CycA to Rb, respectively. If we check the top five beneficial structural perturbations illustrated in blue in the figure, it is interesting to note that the critical regulatory interactions for potential intervention targets connect important genes in the network, typically between the cyclins (CycA, CycB, CycD, CycE),

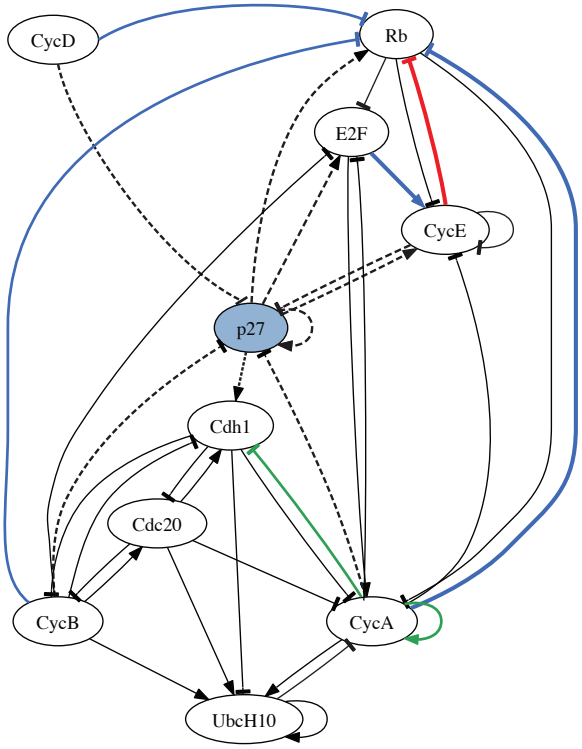

Fig. 1. Mutated mammalian cell cycle network modified from [5]: Normal arrows stand for activations and blunt arrows for suppressing effects. Dashed arrows are the inactive regulations due to the mutation of p27. The red arrow from CycE to p27 is the regulatory relationship that gives the largest reduction of undesirable mass by intervention. The top five regulations according to the undesirable SSD mass reduction are marked in blue. The other regulations leading to beneficial dynamic changes are marked in green.

E2F, and Rb, in the cancerous situation with p27 mutated. This finding is congruent with the recent experimental finding that the Rb and E2F play critical roles for the reverse of the R-point (restriction point), which marks the critical event when a mammalian cell commits to proliferation independent of growth stimulation [17]. The R-point appears to be dysregulated in virtually all cancers and hence the identified intervention targets are biologically significant [17, 18]. Finally, we block the top five regulatory interactions except the suppressing regulation from CycD to Rb for the integrity of the network. The derived structural intervention obtains the undesirable mass $\tilde{\pi}_U^{opt} = 0.3299$ with the perturbed SSD shown in Figure 2(B).

To demonstrate the superiority of the derived solution based on the SMW formula for rank-k function perturbations regarding the time complexity, we have compared its running time with the time spent to compute perturbed SSDs using the power method. Table 1 gives the running time for two methods with the unoptimized code running in MATLAB on a MacPro station with a 2.93GHz CPU and 3GB memory.

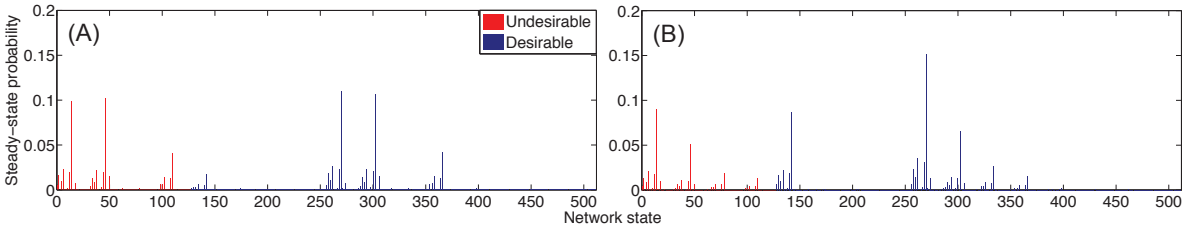

Fig. 2. The steady-state distributions for the mutated mammalian cell cycle network: (A) Original SSD; (B) Perturbed SSD.

Table 1. Running time for computing perturbed SSDs

Method	Power Method	SMW (Proposed)
Time (sec.)	1.89	0.33

These values serve as rough indices that show the reduced time complexity by our new method. For structural intervention to the mutated cell cycle network, the new method leads to more than 80% reduction of the running time. The derived solution is in fact useful for the perturbation analysis of any finite Markov chain model, not just PBNs, and to the best of our knowledge, it is the first analytic solution for general rank-k perturbation. Our preliminary results indicate that our new solutions for perturbed SSDs by general rank-k function perturbations may lead to more efficient methods to derive optimal structural intervention strategies to achieve beneficial dynamic changes.

References

[1] SA Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, pp. 437–467, 1969.

[2] I Shmulevich and ER Dougherty, *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*, SIAM Press, New York, 2010.

[3] I Shmulevich, ER Dougherty, S Kim, and W Zhang, "Probabilistic Boolean networks: A rule-based uncertainty model for gene regulatory networks," *Bioinformatics*, vol. 18, no. 2, pp. 261–274, 2002.

[4] S Huang, "Gene expression profiling, genetic networks, and cellular states: An integrating concept for tumorigenesis and drug discovery," *J Mol Med*, vol. 77, pp. 469–480, 1999.

[5] A Faure, A Naldi, C Chaouiya, and D Theiffry, "Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle," *Bioinformatics*, vol. 22, no. 14, pp. 124–131, 2006.

[6] JG Kemeny and JL Snell, *Finite Markov Chains*, Van Nostrand, New York, 1960.

[7] PJ Schweitzer, "Perturbation theory and finite Markov chains," *J Appl Probab*, vol. 5, pp. 401–413, 1968.

[8] A Datta, R Pal, A Choudhary, and ER Dougherty, "Control approaches for probabilistic gene regulatory networks," *IEEE Signal Processing Mag*, vol. 24, no. 1, pp. 54–63, 2007.

[9] R Pal, A Datta, ML Bittner, and ER Dougherty, "Intervention in context-sensitive probabilistic Boolean networks," *Bioinformatics*, vol. 21, pp. 1211–1218, 2005.

[10] B Faryabi, A Datta, and ER Dougherty, "On approximate stochastic control in genetic regulatory networks," *IET Systems Biology*, vol. 1, no. 6, pp. 361–368, 2007.

[11] I Shmulevich et al., "Control of stationary behavior in probabilistic Boolean networks by means of structural intervention," *Biol Sys*, vol. 10, pp. 431–446, 2002.

[12] Y Xiao and ER Dougherty, "The impact of function perturbations in Boolean networks," *Bioinformatics*, vol. 23, pp. 1265–1273, 2007.

[13] X Qian and ER Dougherty, "Effect of function perturbation on the steady state distribution of genetic regulatory networks: Optimal structural intervention," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4966–4976, 2008.

[14] J Sherman and WJ Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124–127, 1950.

[15] K Lau, S Ganguli, and C Tang, "Function constrains network architecture and dynamics: A case study on the yeast cell cycle Boolean network," *Phys Rev E*, vol. 75, pp. 051907, 2007.

[16] A Garg, A Di Cara, I Xenarios, L Mendoza, and G De Micheli, "Synchronous versus asynchronous modeling of gene regulatory networks," *Bioinformatics*, vol. 24, pp. 1917–1925, 2008.

[17] G Yao, TJ Lee, S Mori, JR Nevins, and L You, "A bistable Myc-Rb-E2F switch: A model for the restriction point," *Nat Cell Biol*, vol. 10, pp. 476–482, 2008.

[18] X Qian, I Ivanov, N Ghaffari, and ER Dougherty, "Intervention in gene regulatory networks via greedy control policies based on long-run behavior," *BMC Systems Biology*, vol. 3, pp. 61, 2009.