# OPTIMAL PERTURBATION CONTROL OF GENE REGULATORY NETWORKS

*Nidhal Bouaynaya*[1], *Roman Shterenberg*[2] *and Dan Schonfeld*[3]

[1]Department of Systems Engineering, University of Arkansas at Little Rock,
[2]Department of Mathematics, University of Alabama at Birmingham,
[3] Department of Electrical and Computer Engineering, University of Illinois at Chicago

**ABSTRACT**

We formulate the control problem in gene regulatory networks as an inverse perturbation problem, which provides the feasible set of perturbations that force the network to transition from an undesirable steady-state distribution to a desirable one. We derive a general characterization of such perturbations in an appropriate basis representation. We subsequently consider the optimal perturbation, which minimizes the overall energy of change between the original and controlled (perturbed) networks. The "energy" of change is characterized by the Euclidean-norm of the perturbation matrix. We cast the optimal control problem as a semi-definite programming (SDP) problem, thus providing a globally optimal solution which can be efficiently computed using standard SDP solvers. We apply the proposed control to the Human melanoma gene regulatory network and show that the steady-state probability mass is shifted from the undesirable high metastatic states to the chosen steady-state probability mass.

## 1. INTRODUCTION

The ultimate goal of gene regulatory network modeling and analysis is to use the network to design effective intervention strategies for affecting the network dynamics in such a way as to avoid undesirable cellular states. As futuristic gene therapeutic interventions, two main control strategies have been proposed to alter gene regulatory network dynamics in a desirable way: (i) introduce external control variables to act upon some control genes, in such a way as to optimize a given cost function [1], and (ii) alter the underlying rule-based structure of the network in order to shift the steady-state mass of the network from undesirable to desirable states.

The first strategy produces a recurrent control policy, over a possibly infinite time horizon interval [1]. Clinically, such an infinite-horizon intervention can be viewed as connecting the patient to an infinitely recurrent feedback control loop. If the control is applied over a finite time horizon and then stopped, the network may not converge towards the desired steady-state distribution. The second strategy, on the other hand, aims at altering the steady-state distribution of the network. A simulation-based study was first conducted in [2], where a procedure to alter the steady-state probability of certain states was implemented using genetic algorithms. Xiao et al. [3] considered an analytical study, where they explored the impact of function perturbations on network attractors. However, their algorithms are rather cumbersome as they need to closely investigate the state changes before and after perturbations. An analytical characterization of the effect on the steady-state distribution caused by perturbation of the regulatory network appears in [4]. The authors relied on the general perturbation theory for finite Markov chains to compute the perturbed steady-state distribution in a sequential manner. However, their intervention is restricted to rank-one perturbations. The extension to higher-rank perturbations is iterative and computationally expensive.

In this paper, we propose a "one-time" (i.e., non-iterative) general intervention strategy that forces the network to converge towards a desired steady-state distribution. Specifically, we formulate the control problem as an inverse perturbation problem by addressing the following question: Given an initial ergodic network, and given a desired steady-state distribution, find an (optimal) perturbation, which forces the network to converge to the desired distribution. The criteria adopted for optimality depends on the particular application and could vary among several possible limitations: potential adverse effects on the patient, length of treatment for the patient, and complexity in the design of bio-molecular control agents. In this paper, we consider minimizing the overall energy of change between the original and perturbed transition matrices criteria for minimal-perturbation control. The "energy" of change is characterized by the Euclidean-norm of the perturbation matrix. The analytical solution to this inverse problem will provide a minimally-perturbed Markov chain characterized by a unique attractor corresponding to the desired distribution.

## 2. MATHEMATICAL NOTATION

In this paper, $\mathbf{1}$ denotes a vector all of whose components are equal to one, and $\mathbf{I}$ stands for the identity matrix. The notation $\mathbf{x} = (\mathbf{y}, \mathbf{z})$ is a shorthand for $\mathbf{x}$ is a linear combination of $\mathbf{y}$ and $\mathbf{z}$. If the inner product $< \mathbf{x}, \mathbf{y} >= 0$, we write $\mathbf{x} \perp \mathbf{y}$. Let $\mathbf{P} \in \mathbb{R}^{n \times n}$, then vec$(\mathbf{P})$ transforms $\mathbf{P}$ into an

$n^2$-dimensional vector by stacking the columns. The curled inequality symbols, $\preceq, \prec, \succeq, \succ$, denote generalized matrix inequalities associated with the positive semi-definite cone. That is, if $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, then $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive-semi-definite.

## 3. THE FEASIBILITY PROBLEM

We consider a gene regulatory network with $m$ genes, where the expression level of each gene is quantized to $l$ values. The dynamic behavior of this network can be represented as a finite-state Markov chain described by a probability transition matrix $\mathbf{P}_0$ of size $n = l^m$. We assume that $\mathbf{P}_0$ is ergodic, i.e., irreducible and aperiodic [5].

A row probability vector $\mu^t = (\mu_1, \cdots, \mu_n)$ is called a stationary distribution, or a steady-state distribution, for $\mathbf{P}_0$ if $\mu^t \mathbf{P}_0 = \mu^t$. Because $\mathbf{P}_0$ is stochastic (i.e., its rows sum up to 1), the existence of stationary distributions is guaranteed [5].

Let $\pi_0$ denote the undesirable steady-state distribution of $\mathbf{P}_0$. We wish to alter this distribution by linearly perturbing the probability transition matrix $\mathbf{P}_0$. Specifically, we consider the perturbed stochastic matrix

$$\mathbf{P} = \mathbf{P}_0 + \mathbf{C}, \tag{1}$$

where $\mathbf{C}$ is a zero row-sum perturbation matrix. The zero row-sum condition is necessary to ensure that the perturbed matrix $\mathbf{P}$ is stochastic. Let us denote by $\pi_d$ the desired stationary distribution. We seek to find a zero row-sum perturbation matrix $\mathbf{C}$ such that the perturbed matrix $\mathbf{P}$ is ergodic and converges to the desired steady-state distribution $\pi_d$.

The genes' dynamics can be obtained from the probability transition matrix as the marginal distribution of the state transition probabilities. It is easy to see that if the probability transition matrix $\mathbf{P}_0$ is perturbed linearly with a zero-row sum matrix, then the conditional probability of each gene will also be perturbed linearly. Thus, "small" perturbations of the probability transition matrix lead to "small" perturbations of the genes' dynamics.

The set of perturbation matrices $\mathbf{C}$, which force the long-term dynamics of the network to transition from $\pi_0$ to $\pi_d$ is given by the polyhedra $\mathcal{D}$:

$$\mathcal{D} = \{\mathbf{C} \in \mathbb{R}^{n\times n} : \pi_d^t = \pi_d^t(\mathbf{P}_0+\mathbf{C}), \mathbf{C}\mathbf{1} = \mathbf{0}, \mathbf{P}_0+\mathbf{C} \geq 0\}. \tag{2}$$

The following proposition provides a characterization of the feasible set of perturbation matrices.

**Proposition 1** *Given an ergodic probability transition matrix $\mathbf{P}_0$ with a stationary distribution $\pi_0$, and given a desired probability vector $\pi_d$, consider the basis $\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ formed by the vectors $\mathbf{x}_1 = \mathbf{1}$, $\mathbf{x}_2 = (\mathbf{1}, \pi_d) \perp \pi_d$, i.e., $\mathbf{x}_2$ is a linear combination of $\mathbf{1}$ and $\pi_d$ that is orthogonal to $\pi_d$, and $\mathbf{x}_i \perp (\mathbf{1}, \pi_d)$ for $i \geq 3$. Let $\mathbf{A}$ be the representation matrix of basis $\mathcal{X}$ in the canonical basis. Then, the perturbed matrix $\mathbf{P} = \mathbf{P}_0 + \mathbf{C}$ has the following block form representation in basis $\mathcal{X}$:*

$$\mathbf{P}_{\mathcal{X}} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & * & * & * \\ \vdots & * & * & * \\ 0 & * & * & * \end{pmatrix}, \tag{3}$$

*where $*$ denotes any real value such that $\mathbf{P} = \mathbf{A}\mathbf{P}_{\mathcal{X}}\mathbf{A}^{-1} \geq 0$.*

For instance, a feasible solution is obtained when all values $*$ are equal to zero. In this case, $\mathbf{P} = \mathbf{A}\mathbf{P}_{\mathcal{X}}\mathbf{A}^{-1} = \mathbf{1}\pi_d^t$ (see Fig. 1(d)). In particular, proposition 1 shows that there are infinitely many perturbation matrices $\mathbf{C}$, which can force the network to transition from its original undesirable steady-state to a desirable one. All such perturbations, in principle, constitute plausible control strategies and can therefore be used to drive the network from one steady-state to another. We impose the minimum-energy constraint in order to limit the structural changes in the network and reduce the transient dynamics after perturbation.

## 4. THE OPTIMAL CONTROL PROBLEM

### *4.0.1. Minimal-perturbation energy control*

The minimal perturbation energy control is defined by minimization of the Euclidean-norm of the perturbation matrix. It corresponds, biologically, to the control which minimizes the overall "energy" of change between the perturbed and unperturbed gene regulatory networks. The Euclidean- or spectral-norm of $C$ is defined as

$$\|\mathbf{C}\|_2 \quad = \quad \sqrt{\lambda_{\max}(\mathbf{C}^t\mathbf{C})} = \max_{\mathbf{x}:\|\mathbf{x}\|=1} < \mathbf{C}^t\mathbf{C}\mathbf{x}, \mathbf{x} >, \tag{4}$$

where $\lambda_{\max}(\mathbf{C}^t\mathbf{C}) \geq 0$ is the highest eigenvalue of the positive-semi-definite matrix $\mathbf{C}^t\mathbf{C}$. The minimum perturbation energy control can be formulated as the following optimization problem:

Minimal-perturbation energy control:

$$\text{Minimize} \quad \|\mathbf{C}\|_2 \quad \text{subject to} \quad \mathbf{C} \in \mathcal{D}, \tag{5}$$

where $\mathcal{D}$ is the feasible set in Eq. (2).

The optimization problem formulated in Eq. (5) is a convex optimization problem, which can be efficiently computed using standard toolboxes for convex optimization [6]. A fundamental property of convex optimization problems is that any locally optimal point is also globally optimal. Moreover, because the Euclidean-norm is strictly convex, the optimal solution is unique.

Next, we express the convex optimization problem as a semi-definite programming (SDP) problem: Using the fact that [6]

$$\|\mathbf{C}\|_2 \leq t \iff \mathbf{C}^t\mathbf{C} \preceq t^2\mathbf{I}, \quad t \geq 0, \tag{6}$$

we can express the problem in Eq. (5) in the following form

$$\begin{aligned} &\text{Minimize} \quad t \\ &\text{subject to} \quad \mathbf{C}^t\mathbf{C} \preceq t^2 I, \quad \mathbf{C} \in \mathcal{D}, \end{aligned} \tag{7}$$

with variables $t \in \mathbb{R}$ and $\mathbf{C} \in \mathbb{R}^{n\times n}$. The problem (7) is readily transformed to a SDP standard form, in which a linear function is minimized, subject to a linear matrix inequality and linear equality constraints. We first observe that, from the Schur complement, we have

$$\mathbf{C}^t\mathbf{C} \preceq t^2 I \text{ (and } t \geq 0) \Longleftrightarrow \begin{pmatrix} t\mathbf{I} & \mathbf{C} \\ \mathbf{C}^t & t\mathbf{I} \end{pmatrix} \succeq 0. \tag{8}$$

The inequalities in (7) can be expressed as a single linear matrix inequality by using the fact that a block diagonal matrix is positive-semi-definite if and only if its blocks are positive semi-definite.

$$\begin{aligned} &\text{Minimize} \quad t \\ &\text{subject to} \quad \begin{pmatrix} t\mathbf{I} & \mathbf{C} & 0 \\ \mathbf{C}^t & t\mathbf{I} & 0 \\ 0 & 0 & \text{vec}(\mathbf{P}_0 + \mathbf{C}) \end{pmatrix} \succeq 0 \\ &\qquad\qquad \pi_d^t(\mathbf{P}_0 + \mathbf{C}) = \pi_d^t, \quad \mathbf{C}\mathbf{1} = 0 \end{aligned} \tag{9}$$

The optimization problem in Eq. (9) is a standard semi-definite programming problem, which can be solved efficiently using standard SDP solvers. A list of 16 SDP solvers can be found at the SDP website maintained by Helmberg [7].

## 5. SIMULATION RESULTS

We apply the proposed inverse perturbation control to a probabilistic Boolean network derived from gene expression data collected in a study of metastatic melanoma [9], [8], [1], [4]. 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 [8]. This suggests a control strategy that reduces the WNT5A genes action in affecting biological regulation.

A seven-gene probabilistic Boolean network (PBN) model of the melanoma network containing the genes WNT5A, pirin, S100P, RET1, MART1, HADHB, and STC2 was derived in [10]. Figure 1(a), derived in [8], illustrates the relationship between genes in the Human melanoma regulatory network. Note that 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.

Using the breadth first search algorithm, we found that the melanoma probabilistic Boolean network is irreducible. Therefore, it has a unique stationary distribution, and we can apply the inverse perturbation control developed in this paper. Because the control objective is to downregulate the WNT5A gene, let us consider as an example a (hypothetical) desired steady-state distribution where the probability of the states having WNT5A upregulated is $10^{-4}$ and the probability of the other states, which correspond to WNT5A downregulated is set equal to $0.015525$ so that the state probabilities sum up to 1 (see Fig. 1(b)). Observe that the states from 0 to 63 have WNT5A downregulated (0) and hence are desirable states, as compared to states 64 to 127 that have WNT5A upregulated (1) and hence are undesirable. The steady-state distribution of the Human melanoma network of the original and perturbed networks are shown in Fig. 1. Observe that the after-control steady-state is identical to the desired steady-state. Therefore, the control has enabled us to shift the steady-state probability mass from the undesirable states to states with lower metastatic competence.

The minimal-energy perturbed matrix, which optimally solves the SDP problem in (9), is $\|C^*\|_2 = 1.20667$. The SDP problem has been implemented in MATLAB and uses the CVX software for convex optimization [11].

# Acknowledgement

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

## 6. APPENDIX

**Proof 1 (Proof of Proposition 1)** *We will distinguish between the matrix* $\mathbf{P}$ *and its corresponding operator* $\tilde{\mathbf{P}}$*. Let us find a finite dimensional operator* $\tilde{\mathbf{P}}$ *that satisfies:*

$$(i)\ \tilde{\mathbf{P}}\mathbf{1} = \mathbf{1}, \quad (ii)\ \tilde{\mathbf{P}}^t\pi_d = \pi_d. \tag{10}$$

*Consider the basis* $\mathcal{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ *defined as follows:* $\mathbf{y}_1 = \pi_d, \mathbf{y}_2 = (\mathbf{1}, \pi_d) \perp \mathbf{1}$ *and* $\mathbf{y}_i = \mathbf{x}_i$*, for* $3 \leq i \leq n$*. We know that the operator* $\tilde{\mathbf{P}}$ *can be written in the following form*

$$\tilde{\mathbf{P}} = \sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_{jk} < \bullet, \mathbf{y}_j > \mathbf{x}_k. \tag{11}$$

*Applying the operator to the vector* $\mathbf{x}_1$*, we obtain*

$$\tilde{\mathbf{P}}\mathbf{x}_1 = \sum_{k=1}^{n} \alpha_{1k}\mathbf{x}_k = \alpha_{11}\mathbf{x}_1 + \sum_{k=2}^{n} \alpha_{1k}\mathbf{x}_k. \tag{12}$$

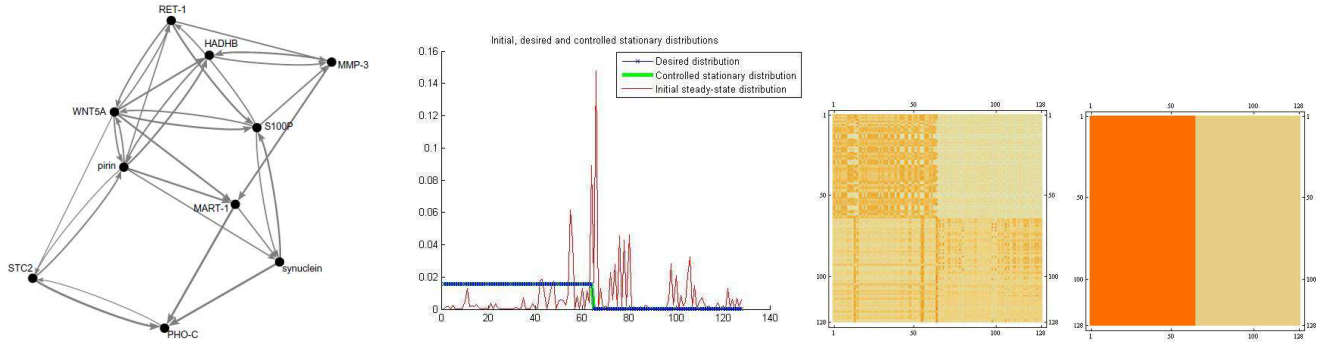


**Fig. 1**. Intervention in the Human melanoma gene regulatory network: (a) An abstract diagram of the melanoma gene regulatory network [8]; (b) The original (red line), desired (blue line), and minimal-perturbation energy controlled (green line) steady-state distributions of the Human melanoma gene regulatory network. The $x$-axis represents the 128 states of the network, and the $y$-axis indicates the probability of each state; (c) A plot of the minimum-energy perturbed matrix $\mathbf{P} = \mathbf{P}_0 + \mathbf{C}^*$; (d) A plot of the feasible perturbed matrix $\mathbf{P} = \mathbf{1}\pi_d^t$. Both perturbed matrices converge towards the desired steady-state $\pi_d$.

*Since* $\mathbf{x}_1 = \mathbf{1}$ *by construction, the condition* $\tilde{\mathbf{P}}\mathbf{1} = \mathbf{1}$ *is equivalent to* $\alpha_{11} = 1$ *and* $\alpha_{1k} = 0$*, for* $k = 2, \cdots, n$*. Similarly, the condition* $\tilde{\mathbf{P}}^t\pi_d = \pi_d$ *is equivalent to* $\alpha_{11} = 1$ *and* $\alpha_{j1} = 0$*, for* $j = 2, \cdots, n$*. Finally, the operator* $\tilde{\mathbf{P}}$ *satisfying the two conditions in (10) can be written as:*

$$\tilde{\mathbf{P}} = < \bullet, \mathbf{y}_1 > \mathbf{x}_1 + \sum_{j=2}^{n}\sum_{k=2}^{n} \alpha_{jk} < \bullet, \mathbf{y}_j > \mathbf{x}_k. \qquad (13)$$

*Applying the operator* $\tilde{\mathbf{P}}$ *in Eq. (13) to the basis* $\mathcal{X}$*, we obtain the desired block form representation* $\mathbf{P}_{\mathcal{X}}$*. The condition* $\mathbf{P} \geq 0$ *is satisfied if and only if* $< \tilde{\mathbf{P}}\mathbf{e}_i, \mathbf{e}_j > \geq 0$ *for all* $1 \leq i, j \leq n$ *or equivalently* $\mathbf{A}\mathbf{P}_{\mathcal{X}}\mathbf{A}^{-1} \geq 0$*, where* $\{\mathbf{e}_i\}_{i=1}^n$ *denotes the canonical basis, i.e., the standard basis for the Euclidean space* $\mathbb{R}^n$*.*

## 7. REFERENCES


[1] A. Datta, R. Pal, A. Choudhary, and E.R. Dougherty, "Control approaches for probabilistic gene regulatory networks - what approaches have been developed for addreassinig the issue of intervention?," *IEEE Signal Processing Magazine*, vol. 24, no. 1, pp. 54–63, January 2007.

[2] I Shmulevich, E R Dougherty, and W Zhang, "Control of stationary behavior in probabilistic Boolean networks by means of structural intervention," *Journal of Biological Systems*, vol. 10, no. 4, pp. 431–445, 2002.

[3] Y. Xiao and E.R. Dougherty, "The impact of function perturbations in Boolean networks," *Bioinformatics*, vol. 23, no. 10, pp. 1265–1273, 2007.

[4] Xiaoning Qian and Edward 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.

[5] E. Seneta, *Non-negative Matrices and Markov Chain*, Springer, 2nd edition, 2006.

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

[7] C. Helmberg, "Semidefinite programming page," http://www-user.tu-chemnitz.de/ helmberg/semidef.html, 2003.

[8] 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.

[9] 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.

[10] 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.

[11] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," `http://cvxr.com/cvx`, May 2010.