EURASIP Journal on Bioinformatics and Systems Biology
a SpringerOpen Journal

RESEARCH Open Access

A comparison study of optimal and suboptimal intervention policies for gene regulatory networks in the presence of uncertainty

Mohammadmahdi R Yousefi[1*] and Edward R Dougherty[2]

Abstract

Perfect knowledge of the underlying state transition probabilities is necessary for designing an optimal intervention strategy for a given Markovian genetic regulatory network. However, in many practical situations, the complex nature of the network and/or identification costs limit the availability of such perfect knowledge. To address this difficulty, we propose to take a Bayesian approach and represent the system of interest as an uncertainty class of several models, each assigned some probability, which reflects our prior knowledge about the system. We define the objective function to be the expected cost relative to the probability distribution over the uncertainty class and formulate an optimal Bayesian robust intervention policy minimizing this cost function. The resulting policy may not be optimal for a fixed element within the uncertainty class, but it is optimal when averaged across the uncertainly class. Furthermore, starting from a prior probability distribution over the uncertainty class and collecting samples from the process over time, one can update the prior distribution to a posterior and find the corresponding optimal Bayesian robust policy relative to the posterior distribution. Therefore, the optimal intervention policy is essentially nonstationary and adaptive.

Keywords: Optimal intervention; Markovian gene regulatory networks; Probabilistic Boolean networks; Uncertainty; Prior knowledge; Bayesian control

Introduction

A fundamental problem of translational genomics is to develop optimal therapeutic methods in the context of genetic regulatory networks (GRNs) [1]. Most previous studies rely on perfect knowledge regarding the state transition rules of the network; however, when dealing with biological systems such as cancer cells, owing to their intrinsic complexity, little is known about how they respond to various stimuli or how they function under certain conditions. Moreover, if there exists any knowledge regarding their functioning, it is usually marginal and insufficient to provide a perfect understanding of the full system. To address uncertainty, one can construct an uncertainty class of models, each representing the system of interest to some extent, and optimize an objective function across the entire uncertainty class. In this way, success in therapeutic applications is fundamentally bound to the degree of *robustness* of the designed intervention method.

Markovian dynamical networks, especially probabilistic Boolean networks (PBNs) [2], have been the main framework in which to study intervention methods due to their ability to model randomness that is intrinsic to the interactions among genes or gene products. The stochastic state transition rules of any PBN can be characterized by a corresponding Markov chain with known transition probability matrix (TPM) [3]. Markov decision processes (MDPs), on the other hand, are a standard framework for characterizing optimal intervention strategies. Many GRN optimization problems have been formulated in the context of MDPs - for instance - infinite-horizon control [4], constrained intervention [5], optimal intervention in asynchronous GRNs [6], optimal intervention when there are random-length responses to drug intervention [7], and optimal intervention to achieve the maximal beneficial

*Correspondence: yousefi@ece.osu.edu
[1]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA
Full list of author information is available at the end of the article

shift in the steady-state distribution [8]. Herein, PBNs will be our choice of reference model for GRNs.

The first efforts to address robustness in the design of intervention policies for PBNs assumed that the errors made during data extraction, discretization, gene selection and network generation introduce a mismatch between the PBN model and the actual GRN [9,10]. Therefore, uncertainties manifest themselves in the entries of the TPM. A *minimax* approach was taken in which robust intervention policies were formulated by minimizing the worst-case performance across the uncertainty class [9]. Thus, the resulting policies were typically conservative. To avoid the detrimental effects of extreme, but rare, states on minimax design and motivated by the results of Bayesian robust filter design [11], the authors in [10] adopted a Bayesian approach whereby the optimal intervention policy depends on the prior probability distribution over the uncertainty class of networks. Constructing a collection of optimal policies, each being optimal for a member of the uncertainty class, the goal was to pick a single policy from this collection that minimizes the average performance relative to the prior distribution. The corresponding policy provides a *model-constrained robust* (MCR) policy. It was noted that this model-constrained policy may not yield the best average performance among all possible policies (we will later define the set of all possible policies for this problem). The authors also considered a class of *globally robust* (GR) policies, which are designed optimally only for a centrality parameter, such as the mean or median, to represent the mass of the uncertainty distribution.

Since [10] was concerned only with stationary policies, it did not consider the possibility of finding nonstationary policies under a Bayesian updating framework, where state transitions observed from the system are used directly to enrich the prior knowledge regarding the uncertainty class. The resulting nonstationary intervention policy, which we refer to it as the *optimal Bayesian robust* (OBR) policy, is our main interest in the present paper. As our main optimization criterion, we use the expected total discounted cost in the long run. This choice is motivated by the practical implications of discounted cost in the context of medical treatment, where the discounting factor emphasizes that obtaining good treatment outcomes at an earlier stage is favored over later stages.

Since the early development of MDPs, it was recognized that when dealing with a real-world problem it seldom happens that the decision maker is provided with the full knowledge of the TPM, but rather some prior information often expressed in a probabilistic manner. Taking a Bayesian approach, an optimal control policy may exist in the expected value sense specifying the best choice of control action in each state. Since the decision maker's state of knowledge about the underlying true process evolves in time as the process continues, the best choice of control action at each state might also evolve. Because the observations are acquired through a controlled process (a control action is taken at every stage of the process), the optimal policy derived through the Bayesian framework may not necessarily ever coincide with a policy that is optimal for the true state of nature. In fact, frequently, the optimal policy is not *self-optimizing* [12]; rather, optimal control will provide the best trade-off between exploration rewards and immediate costs.

Bellman [13] considered a special case of this problem - the two-armed bandit problem with discounted cost - and later used the term *adaptive control* for control processes with incompletely known transition probabilities. He suggested transforming the problem into an equivalent dynamic program with completely known transition laws for which the state now constitutes both the physical state of the process and an *information* state summarizing the past history of the observed state transitions from the process [14]. This new state is referred to as the *hyperstate*. Along this line of research, authors in [15-17] developed the theory of the OBR policy for Markov chains with uncertainty in their transition probabilities, where there is a clear notion of optimality defined with respect to all possible scenarios within the uncertainty class. This approach is in contrast with the MCR methodology because the resulting policy may not be optimal for any member of the uncertainty class but it yields the best performance when averaged over the entire uncertainty class.

Following the methodology proposed in [17] and assuming that the prior probability distribution of a random TPM belongs to a conjugate family of distributions which are closed under consecutive observations, one can formulate a set of functional equations, similar to those of fully known controlled Markov chains, and use a method of successive approximation to find the unique set of solutions to these equations. In this paper, we adopt this approach for the robust intervention of Markovian GRNs and provide a simulation study demonstrating the performance of OBR policies compared with several suboptimal methods, such as MCR and two variations of GR policies, when applied to synthetic PBNs with various structural properties and parameters, as well as to a mutated mammalian cell cycle network.

The paper is organized as follows. First, we give an overview of controlled PBNs and review the nominal MDP problem where the TPMs of the underlying Markov chain are completely known. We then formulate the OBR policy for PBNs with uncertainty in their TPMs and provide the dynamic programming solution to this optimization problem. We demonstrate a conjugate family of probability distributions over the uncertainty class where each row of the random TPM follows a Dirichlet distribution with certain parameters. Assuming that the rows are

independent, the posterior probability distribution will again be a Dirichlet distribution with updated parameters. This provides a compact representation of the dynamic programming equation and facilitates the computations involved in the optimization problem. Several related suboptimal policies are also discussed in detail. Finally, we provide simulation results over both synthetic and real networks, comparing the performance of different design strategies discussed in this paper.

Methods

Controlled PBNs

PBNs constitute a broad class of stochastic models for transcriptional regulatory networks. Their construction takes into account several random factors, including effects of latent variables, involved in the dynamical genetic regulation [3]. The backbone of every PBN is laid upon a collection of Boolean networks (BNs) [18]. A BN is composed of a set of n nodes, $V = \{v^1, v^2, \ldots, v^n\}$ (representing expression level of genes $g^1, g^2, \ldots, g^n$ or their products) and a list of Boolean functions $F = \{f^1, f^2, \ldots, f^n\}$ describing the functional relationships between the nodes. We restrict ourselves to binary BNs, where we assume that each node takes on value of 0, corresponding to an unexpressed (OFF) gene and 1, corresponding to an expressed (ON) gene. This definition extends directly to any finitely discrete-valued nodes. The Boolean function $f^i : \{0,1\}^{j_i} \to \{0,1\}$ determines the value of node i at time $k+1$ given the value of its predictor nodes at time k by $v^i_{k+1} = f^i(v^{i1}_k, v^{i2}_k, \ldots, v^{ij_i}_k)$, where $\{v^{i1}, v^{i2}, \ldots, v^{ij_i}\}$ is the *predictor set* of node v^i. In a BN, all nodes are assumed to update their values synchronously according to F. The dynamics of a BN are completely determined by its state transition diagram composed of 2^n states. Each state corresponds to a vector $\mathbf{v}_k = (v^1_k, v^2_k, \ldots, v^n_k)$ known as the *gene activity profile* (GAP) of the BN at time k. To make our analysis more straightforward, we will replace each GAP, $\mathbf{v}_k$, with its decimal equivalent denoted by $x_k = 1 + \sum_{i=1}^{n} 2^{n-i} v^i_k$, where $x_k \in \mathcal{S} = \{1, \ldots, 2^n\}$ for all k.

A PBN is fully characterized by the same set of n nodes, V, and a set of m constituent BNs, $\mathbf{F} = \{F^1, F^2, \ldots, F^m\}$, called *contexts*, a selection probability vector $R = \{r^1, r^2, \ldots, r^m\}$ over $\mathbf{F}$ ($r^i \geq 0$ for $i = 1, \ldots, m$ and $\sum_{i=1}^{m} r^i = 1$), a network switching probability $q > 0$, and a random gene perturbation probability $p \geq 0$. At any updating epoch, depending on the value of a random variable $\xi \in \{0,1\}$, with $P(\xi = 1) = q$, one of two mutually exclusive events will occur. If $\xi = 0$ then the values of all nodes are updated synchronously according to an operative constituent BN; if $\xi = 1$ then another operative BN, $F^l \in \mathbf{F}$, is randomly selected with probability r^l, and the values of the nodes are updated accordingly. The current BN may be selected consecutively when a switch is called for [1]. PBNs also admit random gene perturbations where the current state of each node in the network can be randomly flipped with probability p.

A PBN is said to be *context-sensitive* if $q < 1$; otherwise, a PBN is called *instantaneously random*. The number of states in a context-sensitive PBN is $m2^n$, whereas the state transition diagram of an instantaneously random PBN is composed of the same 2^n states in $\mathcal{S}$. It is shown in [19] that averaging over the various contexts, relative to R, reduces the transition probabilities of a context-sensitive PBN to an instantaneously random PBN with identical parameters. PBNs with only one constituent BN, i.e., $m = 1$, are called BNs with perturbation and are of particular interest in some applications [8,20]. For the sake of simplicity and reducing the computational time, we will focus only on instantaneously random PBNs.

Since the nature of transitions from one state to another in a PBN is stochastic and has the Markov property, we can model any PBN by an equivalent homogeneous Markov chain, whose states are members of $\mathcal{S}$ and the TPM of this Markov chain can be calculated as described in [19]. We denote the TPM of an instantaneously random PBN by $\mathcal{P}$ and let $\{Z_k \in \mathcal{S}, k = 0, 1, \ldots\}$ be the stochastic process of the state transitions for this PBN. Originating from state $i \in \mathcal{S}$, the successor state $j \in \mathcal{S}$ is selected randomly according to the transition probability $\mathcal{P}_{ij} = P(Z_{k+1} = j \mid Z_k = i)$, the (i,j) element of the TPM. For every $i \in \mathcal{S}$, the transition probability vector $(\mathcal{P}_{i1}, \mathcal{P}_{i2}, \ldots, \mathcal{P}_{i|\mathcal{S}|})$ is a stochastic vector such that $\mathcal{P}_{ij} \geq 0$ and $\sum_{j \in \mathcal{S}} \mathcal{P}_{ij} = 1$ for every $i \in \mathcal{S}$. Random gene perturbation guarantees the ergodicity of the equivalent Markov chain, resulting in a unique invariant measure equal to its limiting distribution.

To model the effect of interventions, we assume that PBNs admit an external control input, A, from a set of possible inputs signals, $\mathcal{A}$, that determines a specific type of intervention on a set of *control genes*. It is common to assume that the control input is binary, i.e., $\mathcal{A} = \{0,1\}$, where $A = 0$ indicates no-intervention and $A = 1$ indicates that the expression level of a single control gene, g^c (or equivalently v^c), for a given $c \in \{1, 2, \ldots, n\}$, should be flipped. For this control scheme, $A = 0$ does not alter the TPM of the original uncontrolled PBN. However, assuming that the network is in state i, the action $A = 1$ replaces the row corresponding to this state by the row that corresponds to the state $\tilde{i}$, where the binary representation of $\tilde{i}$ is the same as i except v^c being flipped. The effect of this binary control scheme on any PBN can be easily generalized to more than one control gene with more than two control actions; in this paper, we only consider the binary control scheme.

Let $\{(Z_k, A_k), Z_k \in \mathcal{S}, A_k \in \mathcal{A}, k = 0, 1, \ldots\}$ denote the stochastic process of a state-action pair. The law of motion

for the controlled network, with binary external control, is represented by a matrix $\mathcal{P}(a)$ with its (i,j) element defined as

$$\mathcal{P}_{ij}(a) = P(Z_{k+1} = j \mid Z_k = i, A_k = a) = \begin{cases} \mathcal{P}_{ij}, & \text{if } a = 0, \\ \mathcal{P}_{ij}, & \text{if } a = 1. \end{cases} \tag{1}$$

$\mathcal{P}_{ij}(a)$ is the probability of going to state $j \in \mathcal{S}$ at time $k+1$ from state $i \in \mathcal{S}$, while taking action $a \in \mathcal{A}$, at time k. By this construction, it is clear that the controlled TPM, $\mathcal{P}(a)$, can be calculated directly from $\mathcal{P}$.

The nominal problem

External intervention in the context of Markovian networks refers to a class of sequential decision making problems in which actions are taken at discrete time units to alter the dynamics of the underlying GRN. It is usually assumed that the decision maker can observe the state evolution of the network at consecutive time epochs $k = 0, 1, \ldots, N$, where the *horizon* N may be finite or infinite. At each k, upon observing the state, the decision maker chooses an action from $\mathcal{A}$ that will subsequently alter the dynamics of the network. Hence, the stochastic movement of the GRN from one state to another is completely characterized based on the current state and action taken at this state by (1).

Associated with each state and action, there is an immediate cost function $g : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ to be accrued until the next decision epoch, which we assume is nonnegative and bounded. This cost may reflect the degree of desirability of different states and/or the cost of intervention that is applied. Whenever the process moves from state i to j under action a, a known cost $g_{ij}(a)$ is incurred. We also assume that $\lambda \in (0,1)$ is the discount factor reflecting the present value of the future cost. An *intervention policy*, denoted by μ, is a prescription for taking actions from the set $\mathcal{A}$ at each point k in time. In general, one can allow a policy for taking an action at time k to be a mapping from the entire history of the process up to time k to the action space. This mapping need not be deterministic; on the contrary, it might involve a random mechanism that is a function of the history. However, for the problem we consider, there exists a deterministic policy that is optimal. We denote the set of all admissible policies by $\mathcal{M}$. The TPM $\mathcal{P}$, initial state $Z_0 = i$, and any given policy $\mu = \{\mu_0, \mu_1, \ldots\}$ in $\mathcal{M}$ determine a unique probability measure, P_i^{μ}, over the space of all trajectories of states and actions, which correspondingly defines the stochastic processes Z_k and A_k of the states and actions for the controlled network [12]. In the nominal optimization problem, we desire an intervention policy $\mu \in \mathcal{M}$ such that the objective function

$$J_{\mathcal{P}}^{\mu}(i) = \lim_{N\to\infty} E_i^{\mu}\left\{\sum_{k=0}^{N-1} \lambda^k g_{Z_k Z_{k+1}}(A_k)\right\}, \tag{2}$$

is minimized, i.e., $J_{\mathcal{P}}^{*}(i) = \min_{\mu\in\mathcal{M}} J_{\mathcal{P}}^{\mu}(i)$ for all $i \in \mathcal{S}$. In the above equation, E_i^{μ} denotes expectation relative to the probability measure P_i^{μ}.

This optimization problem is usually solved by formulating a set of simultaneous functional equations and a mapping $TJ : \mathcal{S} \to \mathbb{R}$, obtained by applying the dynamic programming mapping to any function $J : \mathcal{S} \to \mathbb{R}$, for all $i \in \mathcal{S}$ defined by

$$(TJ)(i) = \min_{a\in\mathcal{A}}\left\{\sum_{j\in\mathcal{S}} \mathcal{P}_{ij}(a) g_{ij}(a) + \lambda \sum_{j\in\mathcal{S}} \mathcal{P}_{ij}(a) J(j)\right\}. \tag{3}$$

The optimal cost function J^* uniquely satisfies the above functional equation, i.e., it is the fixed point of the mapping T. One can determine the optimal policy with the help of convergence, optimality, and uniqueness theorems for the solution, proven in [21]. These results furnish an iterative method for successive approximation of the optimal cost function, which in turn gives the optimal intervention policy. It can be further shown that the optimal intervention policy belongs to the class of *stationary deterministic* policies, meaning that $\mu_k = \mu$ for all k and $\mu : \mathcal{S} \to \mathcal{A}$ is a single-valued mapping from states to actions.

OBR intervention policy

In many real-world intervention scenarios, perfect knowledge regarding $\mathcal{P}$ may be unavailable or very expensive to acquire. Therefore, we resort to a probabilistic characterization of the elements of $\mathcal{P}$ and optimize relative to this uncertainty. Our results in this section are mainly derived from the Bayesian treatment of MDPs by [17]. Let

$$\Omega = \left\{\mathcal{P} : \mathcal{P} \text{ is } |\mathcal{S}| \times |\mathcal{S}|, \mathcal{P}_{ij} \geq 0, \sum_{j\in\mathcal{S}} \mathcal{P}_{ij} = 1 \text{ for all } i,j \in \mathcal{S}\right\}, \tag{4}$$

denote the set of all valid uncontrolled TPMs. The uncertainty about the random matrix $\mathcal{P}$ is characterized by the prior probability density $\pi(\mathcal{P})$ over the set Ω. Given $\pi(\mathcal{P})$ and some initial state i, we define

$$J^{\mu}(i,\pi) = \lim_{N\to\infty} E_{i,\pi}^{\mu}\left\{\sum_{k=0}^{N-1} \lambda^k g_{Z_k Z_{k+1}}(A_k)\right\}, \tag{5}$$

where the expectation is taken not only with respect to the random behavior of the state-action stochastic process but also with respect to the random choice of $\mathcal{P}$ according to its prior distribution, $\pi(\mathcal{P})$. The goal is to find an optimal policy μ^* such that (5) is minimized for any $i \in \mathcal{S}$ and any prior distribution π, i.e., $\mu^* = \text{argmin}_{\mu\in\mathcal{M}} J^{\mu}(i,\pi)$. We denote the optimal cost by $J^*(i,\pi)$.

Suppose that we could find optimal intervention policies for every element of Ω. Letting $J^*_{\mathcal{P}}(i)$ denote the optimal cost for any $\mathcal{P} \in \Omega$ and $i \in \mathcal{S}$ and assuming that the optimal cost $J^*(i,\pi)$ exists, we have $E_\pi[J^*_{\mathcal{P}}(i)] \le J^*(i,\pi)$ for all $i \in \mathcal{S}$ and any π. In other words, $E_\pi[J^*_{\mathcal{P}}(i)]$ is the best that could be achieved if we were to optimize for every element of the uncertainty class for fixed i and π.

Since at every stage of the problem an observation is made immediately after taking an action, we can utilize this additional information and update the prior distribution to a posterior distribution as the process proceeds in time. Therefore, we can treat $\pi(\mathcal{P})$ as an additional state and call (i,π) the hyperstate of the process. From this point of view, we seek an intervention policy that minimizes the total expected discounted cost when the process starts from a hyperstate (i,π). Suppose the true, but unknown, TPM is $\hat{\mathcal{P}}$. At time 0, the initial state z_0 is known and $\mathcal{P}$ is distributed according to π. Based on z_0 and π, the controller chooses an action a_0 according to some intervention policy. Based on $(z_0, a_0, \hat{\mathcal{P}})$ the new state z_1 is realized according to the probability transition rule $\hat{\mathcal{P}}_{z_0 z_1}(a_0)$ and a cost $g_{z_0 z_1}(a_0)$ is incurred. Based on (z_0, π, a_0, z_1), the controller chooses an action a_1 according to some (possibly another) intervention policy and so on [12]. Although the number of states in $\mathcal{S}$ and actions in $\mathcal{A}$ are finite, the space of all possible hyperstates is essentially uncountable. Therefore, finding an optimal intervention policy which provides a mapping from the space of hyperstates to the space of actions in a sense similar to the nominal case is rather difficult. However, as we will see, it is possible to find an optimal action for a fixed initial hyperstate using an equivalent dynamic program.

Dynamic programming solution

We assume that the rows of $\mathcal{P}$ are mutually independent. Note that this assumption might not hold true for a large class of problems; however, the analysis becomes overwhelmingly complicated if one is willing to relax this assumption. The posterior probability density of $\mathcal{P}$, when the process moves from state i to state j under control a, is found via Bayes' rule:

$$\pi'(\mathcal{P}; i, a, j) = \begin{cases} c\mathcal{P}_{ij}\pi(\mathcal{P}), \text{ if } a = 0, \\ c'\mathcal{P}_{ij}\pi(\mathcal{P}), \text{ if } a = 1, \end{cases} \tag{6}$$

where c and c' are normalizing constants depending on i, a, and j. Under the sequence of events described above, Martin [17] showed that the minimum expected discounted cost over an infinite period, $N = \infty$, exists and formulated an equivalent dynamic program with a set of simultaneous functional equations. The dynamic programming operator T, similar to (3) but now with the hyperstate (i,π), takes the following form:

$$(TJ)(i,\pi(\mathcal{P})) = \min_{a\in\mathcal{A}} \left\{ \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) g_{ij}(a) + \lambda \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) J(j, \pi'(\mathcal{P}; i, a, j)) \right\}, \tag{7}$$

for all $i \in \mathcal{S}$, where $\bar{\mathcal{P}}_{ij}(a) = E[\mathcal{P}_{ij}(a)]$ with respect to the prior probability density function π. It is shown in [17] that there exists a unique bounded set of optimal costs J^* satisfying

$$J^*(i,\pi(\mathcal{P})) = \min_{a\in\mathcal{A}} \left\{ \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) g_{ij}(a) + \lambda \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) J^*(j, \pi'(\mathcal{P}; i, a, j)) \right\},$$

which is the fixed point of the operator T. Since the space of all possible hyperstates (i,π) is uncountable, construction of an optimal intervention policy for all (i,π), except for some special cases, may not be feasible. However, given that the process starts at (i,π), the minimization argument in the above equation yields an optimal action to take only for the current hyperstate.

The difficulty in solving (7), which makes it more complicated than (3), is that the total expected discounted cost when different actions are taken now involves the difference in expected immediate costs and the expected difference in future costs due to being in different states at the next period as well as the effect of different information states resulting from these actions [22]. It should be noted that since the decision maker's knowledge regarding the uncertainty about $\mathcal{P}$ evolves with each transition, the intervention policy will also evolve over time. In a sense, the optimal policy will adapt, implying that stationary optimal policies as defined for the nominal problem do not exist. The optimal nonstationary intervention policy derived through the process discussed above is referred to as the OBR policy.

Special case: independent Dirichlet priors

Suppose that both prior and posterior distributions belong to the same family of distributions, i.e., they are conjugate distributions. Then, instead of dealing with prior and posterior at every stage of the problem, we will

only need to keep track of the *hyperparameters* of the prior/posterior distributions. A special case of the families of distributions closed under consecutive observations is the Dirichlet distribution, which is the conjugate prior of the multinomial distribution.

Let the initial state z_0 be known and $\mathbf{z}_n = (z_0, z_1, z_2, \ldots, z_n)$ represent a sample path of n independent transitions recorded from the network under the influence of an intervention policy. Then the posterior probability density of $\mathcal{P}$, $\pi'(\mathcal{P})$, can be found using Bayes' rule:

$$\pi'(\mathcal{P}) \propto \pi(\mathcal{P}) \prod_{i\in\mathcal{S}} \prod_{j\in\mathcal{S}} (\mathcal{P}_{ij})^{\beta_{ij}}, \tag{8}$$

where β_{ij} denotes the number of transitions in $\mathbf{z}_n$ from state i to state j. The right product in (8) is called the *likelihood function* and the constant of proportionality can be found by normalizing the integral of $\pi'(\mathcal{P})$ over Ω to 1. Note that although the transitions made in $\mathbf{z}_n$ result from an intervention policy, we have formulated the likelihood function only in terms of the elements of $\mathcal{P}$ (and not $\mathcal{P}(a)$). This is a consequence of our particular intervention model, where we can substitute for $\mathcal{P}_{ij}(a)$ with $\mathcal{P}_{\bar{i}j}$ whenever $a = 1$ as shown in (1). To be more precise, we have $\beta_{ij} = \beta_{ij}(0) + \beta_{\bar{i}j}(1)$, where $\beta_{ij}(a)$ is the number of transitions in $\mathbf{z}_n$ from state i to state j under control a.

For a fixed state i, a transition to state j is an outcome of a multinomial sampling distribution with parameters $\{\mathcal{P}_{i1}, \mathcal{P}_{i2}, \ldots, \mathcal{P}_{i|\mathcal{S}|}\}$ constituting the standard $(|\mathcal{S}| - 1)$-simplex. As stated in the beginning of this section, the conjugate prior for the multinomial distribution is given by the Dirichlet distribution. By the independence assumption imposed on the rows of $\mathcal{P}$, one can write the prior for $\mathcal{P}$ as

$$\pi(\mathcal{P}) = c(\alpha) \prod_{i\in\mathcal{S}} \prod_{\in\mathcal{S}} (\mathcal{P}_{ij})^{\alpha_{ij}-1}, \tag{9}$$

where $\alpha_{ij} > 0$ and $\alpha = [\alpha_{ij}]$ is the hyperparameter matrix with the rows arranged in the same manner as $\mathcal{P}$. The constant of proportionality is given by

$$c(\alpha) = \prod_{i\in\mathcal{S}} \frac{\Gamma\left(\sum_{j\in\mathcal{S}} \alpha_{ij}\right)}{\prod_{j\in\mathcal{S}} \Gamma(\alpha_{ij})}, \tag{10}$$

where Γ is the gamma function. The uniform prior distribution is obtained if $\alpha_{ij} = 1$ for all $i, j \in \mathcal{S}$. As we increase a specific α_{ij}, it is as if we bias the posterior distribution on the corresponding element of $\mathcal{P}$ with some transition samples before ever observing any samples. It can be verified that

$$E[\mathcal{P}_{ij}] = \frac{\alpha_{ij}}{\sum_{l\in\mathcal{S}} \alpha_{il}} = \bar{\mathcal{P}}_{ij}, \tag{11}$$

and

$$\mathrm{var}[\mathcal{P}_{ij}] = \frac{\bar{\mathcal{P}}_{ij}(1 - \bar{\mathcal{P}}_{ij})}{\sum_{l\in\mathcal{S}} \alpha_{il} + 1}.$$

We also have the following theorem, which is due to Martin [17].

Theorem 1. *Let $\mathcal{P}$ have a probability density function given in (9) and (10) with the hyperparameter matrix α and suppose that a sample with a transition count matrix $\beta = [\beta_{ij}]$ is observed. Then the posterior probability density function of $\mathcal{P}$ will have the same form as in (9) and (10), but with the hyperparameter matrix $\alpha + \beta$.*

Assuming α as the hyperparameter representing $\pi(\mathcal{P})$ and using Theorem 1, one can rewrite Equation 7 as

$$(TJ)(i,\alpha) = \min_{a\in\mathcal{A}} \left\{ \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) g_{ij}(a) + \lambda \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) J(j, \alpha + \gamma) \right\},$$

where γ is a matrix of all zeros except $\gamma_{ij} = 1$ if $a = 0$ or $\gamma_{\bar{i}j} = 1$ if $a = 1$, and

$$\bar{\mathcal{P}}_{ij}(a) = \begin{cases} \bar{\mathcal{P}}_{ij}, \text{ if } a = 0, \\ \bar{\mathcal{P}}_{\bar{i}j}, \text{ if } a = 1. \end{cases}$$

The optimal cost $J^*(i,\alpha)$ is defined by

$$J^*(i,\alpha) = \min_{\mu\in\mathcal{M}} J^{\mu}(i,\alpha),$$

for a given $i \in \mathcal{S}$ and prior hyperparameter α. Taking an approach based on the method of successive approximation, let $J_k(i,\alpha)$ for $k = 0, 1, \ldots$ be defined recursively for all $i \in \mathcal{S}$ and any valid hyperparameter matrix α by

$$J_{k+1}(i,\alpha) = \min_{a\in\mathcal{A}} \left\{ \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) g_{ij}(a) + \lambda \sum_{j\in\mathcal{S}} \bar{\mathcal{P}}_{ij}(a) J_k(j, \alpha + \gamma) \right\}, \tag{12}$$

with $\{J_0(i,\alpha)\}$ as a set of bounded initial functions. Under some mild conditions, the sequence of functions $\{J_k(i,\alpha)\}$ converges monotonically to the optimal solution $J^*(i,\alpha)$ for any $i \in \mathcal{S}$ and uniformly for all valid α [17]. Faster rates of convergence can be achieved for smaller values of λ. Assuming that the method of successive approximation converges in K steps, then for a specific value of (i,α), one needs to evaluate $(|\mathcal{A}| \times |\mathcal{S}|)^K$ terminal values necessary for the computation of $J^*(i,\alpha)$. Therefore, to minimize computational time, we restrict ourselves to

small values for λ and K. Once the successive approximation converges, an action a^* that minimizes the RHS of (12) is optimal.

The intervention policy optimally adapts to the consecutive observations as follows: we start with an initial hyperstate (z_0, α_0), with α_0 reflecting our prior knowledge regarding the unknown network (or equivalently $\mathcal{P}$). We can calculate $\bar{\mathcal{P}}$ using (11) with respect to α_0 and utilize the successive approximation method in (12) for a fixed K to find an optimal action a^*. We then apply the action a^* to the network and let it transition from state z_0, or $\tilde{z}_0$ depending on the optimal action, to a new random state z_1 according to $\hat{\mathcal{P}}$. We incorporate the new observation into our prior knowledge and update the hyperparameter matrix to α_1 by incrementing the entry at (z_0, z_1) or $(\tilde{z}_0, z_1)$ of the hyperparameter matrix α_0 by 1. We repeat the entire optimization procedure, but now with the new hyperstate (z_1, α_1), etc. A schematic diagram of this procedure is demonstrated in Figure 1.

The extreme computational complexity of finding the OBR intervention policy for MDPs with large state-space poses a major obstacle when dealing with real-world problems. It is relatively straightforward to implement the procedure described above for networks with three or four genes. However, for larger networks, one should resort either to clever ways of indexing all possible transitions, such as hash tables or a branch-and-bound algorithm, or to approximation methods, such as reinforcement learning. See [12,22,23] for more details. An alternative approach, as we will demonstrate, is to implement suboptimal methods that, in general, have acceptable performance. Yet another potential approach to circumvent the explosion of the space of all hyperstates is to reduce the size of the uncertainty class. For example, we can assume that some rows of the underlying TPM are perfectly known and uncertainty is only on some other rows, with the implication that the regulatory network is partially known. We will leave the analysis of such approaches to future research.

Suboptimal intervention policies

Besides the OBR policy, three suboptimal policies are of particular interest: *MCR*, *GR*, and *adaptive GR* (AGR). Similar to the previous section, let $\mathcal{P}$ be random, having a probability density $\pi(\mathcal{P})$ over the set of valid TPMs, Ω, defined in (4).

Let $\mathcal{M}_{\mathrm{MCR}}$ denote the set of all policies that are optimal for some element $\mathcal{P} \in \Omega$. Each policy in $\mathcal{M}_{\mathrm{MCR}}$ is stationary and deterministic (each corresponds to a problem with known TPM). Because Ω is uncountable and there exits a finite number of stationary deterministic policies, one might find policies that are optimal for many elements of Ω. Assuming that the initial state Z_0 is randomly distributed according to some probability distribution η, the policy μ_{MCR} yields the minimum cost, which is defined by

$$J_{\mathrm{MCR}}(i) = \min_{\mu \in \mathcal{M}_{\mathrm{MCR}}} E_\pi \left[E_\eta \left[J^\mu_{\mathcal{P}}(Z_0) \right] \right], \tag{13}$$

where $J^\mu_{\mathcal{P}}(Z_0)$ is defined in (2) for any fixed Z_0 and $\mathcal{P}$. Since we are limiting ourselves to policies in $\mathcal{M}_{\mathrm{MCR}}$, it is seldom the case that a single policy minimizes $E_\pi[J^\mu_{\mathcal{P}}(Z_0)]$ for all $Z_0 \in \mathcal{S}$. Hence, we take the expected value of $J^\mu_{\mathcal{P}}(Z_0)$ with respect to η in (13) as a single value representing the expected cost. The resulting MCR intervention policy is therefore fixed for a given prior distribution in the sense that it will not adapt to the observed transitions.

We define the GR policy as the minimizing argument for the optimization problem given by $J_{\mathrm{GR}}(i) = \min_{\mu \in \mathcal{M}} J^\mu_{\bar{\mathcal{P}}}(i)$, for all $i \in \mathcal{S}$, where $\bar{\mathcal{P}} \in \Omega$ is the mean of the uncertainty class Ω with respect to the prior distribution π. The optimization method presented for the nominal problem can be readily applied. Hence, the resulting policy, μ_{GR}, is stationary and deterministic. In the case of independent Dirichlet priors, $\bar{\mathcal{P}}$ is given by Equation 11. Here we are considering the mean as an estimate for unknown $\mathcal{P}$. However, one can use any other estimate of $\mathcal{P}$ and find the optimal policy in a similar fashion. Similar to the MCR policy, this intervention method is also fixed for a given prior distribution and it will not adapt to the observed transitions.

The AGR policy is similar to the GR policy in the sense that it is optimal for the mean of the uncertainty class Ω. However, instead of taking the mean with respect to the prior distribution π and using the same policy for the entire process, we update π to a posterior π', defined in (6), whenever a transition is made and calculate the mean of Ω with respect to π'. Since the posterior evolves as we observe more and more transitions, the AGR policy also evolves - therefore, the name adaptive. We denote the cost and the corresponding policy resulting from this procedure, for any initial hyperstate (i, π), by $J_{\mathrm{AGR}}(i, \pi)$ and μ_{AGR}, respectively. In the case of independent Dirichlet priors, we can simply replace π with α.

Results

In this section, we provide a comparison study on the performance of optimal and suboptimal policies based on simulations on synthetically generated PBNs and a real network. Since we implement the method of successive approximation to calculate μ_{OBR}, we restrict ourselves to synthetic networks with $n = 3$ genes. Given that, as we will show, μ_{AGR} yields very similar performance compared to the optimal policy, we can implement μ_{AGR} for networks of larger size and use it as the baseline for comparison with other suboptimal policies, keeping in mind that the optimal policy should and will outperform any suboptimal method.

Synthetic networks

We first consider randomly generated PBNs with $n = 3$ genes and $m = 3$ equally likely constituent BNs (total number of states being 8) with the maximum number of predictors for each node set to 2 ($j_i \leq 2$ for all $i \in \{1, 2, \ldots, n\}$). The *bias* of a randomly generated PBN is the probability that each of its Boolean regulatory functions takes on the value 1 in its truth table. We assume that the bias is taken randomly from a beta distribution with mean 0.5 and standard deviation 0.01. The gene perturbation probability is assumed to be $p = 0.001$. In the context of gene regulation, there are some genes associated with phenotypes (typically undesirable ones). We refer to these genes as *target genes* and our goal in controlling the network is to push the dynamics of these genes away from undesirable states towards desirable ones. Once the set of target genes is identified, one can partition the state space $\mathcal{S}$ into subsets of desirable and undesirable states, denoted by $\mathcal{D}$ and $\mathcal{U}$, respectively. In our synthetic network simulations, we choose the control and target genes to be the least and most significant bits in the binary representation of states, respectively, and assume that downregulation of the target gene is undesirable. As for the discount factor and the immediate cost function $g_{ij}(a)$, we set $\lambda = 0.2$ and

$$g_{ij}(a) = \begin{cases} 2.1, & \text{if } j \in \mathcal{U} \text{ and } a = 1, \\ 2.0, & \text{if } j \in \mathcal{U} \text{ and } a = 0, \\ 0.1, & \text{if } j \in \mathcal{D} \text{ and } a = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

the interpretation being that a cost will be incurred if the future state in undesirable or there is an intervention in the network.

To design an OBR policy for a given network, we need to assign the prior probability distribution to the set Ω. As discussed earlier, independent Dirichlet priors parameterized by α constitute a natural choice for this application. Therefore, we only need to assign values to α. The choice of prior hyperparameters plays a crucial role in the design of an optimal policy: the tighter the prior around the true, but unknown, TPM $\hat{\mathcal{P}}$, the closer the OBR cost is to that of $\hat{\mathcal{P}}$. Since our synthetic networks are generated randomly and not according to some biologically motivated GRN, it would be difficult to assign prior probabilities for individual networks. Therefore, we use the randomly generated PBNs themselves for this purpose and perturb and scale the elements of the TPMs via the ε-contamination method.

A random PBN, $\hat{\mathcal{P}}$, is first generated. This network will serve as the true, but unknown, PBN. Then a contamination matrix $\mathcal{Q}$ of the same size ($|\mathcal{S}| \times |\mathcal{S}|$) is generated, where each row is sampled uniformly from the $|\mathcal{S} - 1|$-simplex. Note that $\mathcal{Q}$ is a valid TPM. We now define the hyperparameter matrix α by

$$\alpha = \kappa \left((1 - \varepsilon)\hat{\mathcal{P}} + \varepsilon \mathcal{Q} \right), \tag{15}$$

where $\kappa > 0$ controls the tightness of the prior around the true PBN and $\varepsilon \in [0, 1]$ controls the level of contamination. For networks with three genes, we assume that $\varepsilon = 0.1$ and demonstrate the effect of κ on the performance of intervention policies.

We generate 500 random PBNs, denoted by $\{\mathcal{N}^l\}$ for $l = 1$ to 500, for each set of parameters and calculate their TPMs, denoted by $\{\hat{\mathcal{P}}^l\}$. These networks will serve as the ground-truth for our simulation study. For a given pair of κ and ε, we then construct hyperparameter matrices, denoted by $\{\alpha^l\}$, using (15), each corresponding to a random network. To compare the performance of different intervention policies, for each randomly generated network $\mathcal{N}^l$, we take a Monte Carlo approach and generate 500 random TPMs, denoted by $\{\hat{\mathcal{P}}^{l,l'}\}$ for $l' = 1$ to 500, from the α^l-parameterized independent Dirichlet priors. The set $\{\hat{\mathcal{P}}^{l,l'}\}$ will essentially represent Ω and the prior distribution.

To design and evaluate the performance of μ_{MCR} for each random PBN $\mathcal{N}^l$, we proceed as follows: We find the optimal intervention policy for each $\hat{\mathcal{P}}^{l,l'}$, apply this policy to every element in the set $\{\hat{\mathcal{P}}^{l,l'}\}$, and calculate the average over all equally likely initial states, $Z_0 \in \mathcal{S}$, of the infinite-horizon expected discounted cost using (2) for that element. The expected performance of the each

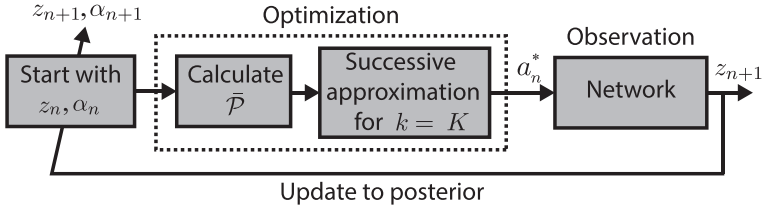

Figure 1 Optimization procedure for an OBR policy. We start with a hyperstate $(z_n, \boldsymbol{\alpha}_n)$. We calculate $\bar{\boldsymbol{\mathcal{P}}}$ using $\boldsymbol{\alpha}_n$ and utilize the successive approximation method for a fixed K to find an optimal action a_n^*. We then apply the action a_n^* to the network and let it transition from state z_n, or $\tilde{z_n}$ depending on the optimal action, to a new random state z_{n+1} according to $\hat{\boldsymbol{\mathcal{P}}}$. We incorporate the new observation into our prior knowledge and update the hyperparameter matrix to $\boldsymbol{\alpha}_{n+1}$ by incrementing the entry at (z_n, z_{n+1}) or $(\tilde{z}_n, z_{n+1})$ of the hyperparameter matrix $\boldsymbol{\alpha}_n$ by 1. We repeat the entire optimization procedure, but now with the new hyperstate $(z_{n+1}, \boldsymbol{\alpha}_{n+1})$.

policy optimal for $\hat{\mathcal{P}}^{l,l'}$, relative to the prior distribution, can be computed by taking the average of the resulting costs over all $\hat{\mathcal{P}}^{l,l'}$. We repeat this procedure for every element of $\{\hat{\mathcal{P}}^{l,l'}\}$ and declare a policy MCR if it yields the minimum expected performance. We denote the expected cost function for a random PBN $\mathcal{N}^l$ obtained via an MCR policy by J^l_{MCR}.

Finding μ_{GR} for each PBN $\mathcal{N}^l$, on the other hand, is easier and it requires only the value of the hyperparameter α^l. Once found, the performance of this policy is evaluated by applying it to all elements of $\{\hat{\mathcal{P}}^{l,l'}\}$ and taking the average of the resulting costs. Similar to the MCR policy, we assume that the initial states are equally likely and calculate the average over all possible initial states. We denote the expected cost function corresponding to the GR policy derived for $\mathcal{N}^l$ by J^l_{GR}.

To quantify the performance of the OBR policy for each random PBN $\mathcal{N}^l$, we directly evaluate the cost function defined in (5) relative to the independent Dirichlet prior distribution, π^l, parameterized by α^l. This is accomplished using the sample set of 500 random TPMs, $\{\hat{\mathcal{P}}^{l,l'}\}$. Starting from a hyperstate and a TPM $\hat{\mathcal{P}}^{l,l'}$, we derive an optimal action from (12) using the method of successive approximations with $K = 5$ and some initial cost function. We then observe a transition according to $\hat{\mathcal{P}}^{l,l'}$ and find the incurred discounted immediate cost according to (14), depending on the new observed state and the optimal action just taken. We update our prior hyperparameter and carry out the optimization problem again, but now with the updated hyperparameter and the recently observed state, and accumulate the newly incurred discounted immediate cost. We iterate this for seven epochs, thus observing seven different hyperstates for a sampling path, and record the total accumulated discounted cost over this period. We then repeat this entire process, for the same $\hat{\mathcal{P}}^{l,l'}$ for 100 iterations (although the same TPM is used, different sampling paths will result due to random transitions), and take the average of all 100 total accumulated discounted cost values. This will represent the cost associated with $\hat{\mathcal{P}}^{l,l'}$ and the initial state. We implement a similar procedure for all initial states (assuming all equally likely) and all elements of $\{\hat{\mathcal{P}}^{l,l'}\}$ and take the average of the resulting costs, yielding the expected optimal cost, $E_\eta[J^*(Z_0, \pi^l)]$, with respect to the uniform probability distribution η over the initial states in $\mathcal{S}$. Since we use the same hyperparameter α^l in our Monte Carlo simulation for a given random PBN $\mathcal{N}^l$, we denote the expected optimal cost obtained from a OBR policy by J^l_{OBR}.

We take a similar approach for evaluating the performance of μ_{AGR}. Instead of using the method of successive approximations at every epoch, we use the current value of the hyperparameter to calculate the mean of Ω and use this to find the optimal action to take at that hyperstate. Every other step of the process is essentially the same to those of the OBR policy. We denote the expected optimal cost obtained from this policy by J^l_{AGR}.

We also evaluate three other cost functions for each PBN $\mathcal{N}^l$: $J^l_{\mathrm{LB}} := E_\pi\left[E_\eta[J^*_{\mathcal{P}}(Z_0)]\right]$, $J^l_{\mathrm{T}} := E_\eta[J^*_{\hat{\mathcal{P}}^l}(Z_0)]$, and $J^l_{\mathrm{ET}} := E_\pi\left[E_\eta[J^l_{\mathcal{P}}(Z_0)]\right]$, where $J^l_{\mathcal{P}}$ is the cost of applying an optimal intervention policy corresponding to $\hat{\mathcal{P}}^l$ to an element $\mathcal{P}$ of Ω. The first cost function, J^l_{LB}, is a lower bound on the performance of the OBR policy, J^l_{BA}. The second cost function, J^l_{T}, corresponds to the cost of applying an optimal intervention policy as if we knew the true network, $\hat{\mathcal{P}}^l$, to the true network itself. The third cost function, J^l_{ET}, is the expected cost, relative to the prior, of applying an intervention policy that is optimal for the true network. We can calculate these cost functions assuming that Ω and the prior distribution π^l are represented by the set $\{\hat{\mathcal{P}}^{l,l'}\}$ corresponding to each PBN $\mathcal{N}^l$.

All cost functions discussed above are defined relative to a given random PBN $\mathcal{N}^l$. Since we have 500 such networks, for each parameter value, we report the average performance across all random networks and provide a statistical comparison on the performance of different intervention policies. The results are presented in Table 1. As seen in the table, the optimal policy performance, in the average sense, is consistently better than all suboptimal policies. The closest performance to the optimal method is achieved by the AGR policy, which is not surprising, since this policy adapts to the process over time by updating the prior distribution to a posterior distribution and optimizes with respect to the mean of the posterior.

As it has been reported in the previous studies [7,8,24], the performance of an optimal policy might not significantly exceed those of suboptimal policies when averaged across random PBNs; nonetheless, there are networks for which the optimal policy notably outperforms the suboptimal ones. To demonstrate this, we use the difference

Table 1 Average costs across all 500 randomly generated PBNs with $n = 3$ genes and $\varepsilon = 0.1$

	$E[J^l_{\mathrm{LB}}]$	$E[J^l_{\mathrm{T}}]$	$E[J^l_{\mathrm{ET}}]$	$E[J^l_{\mathrm{MCR}}]$	$E[J^l_{\mathrm{GR}}]$	$E[J^l_{\mathrm{AGR}}]$	$E[J^l_{\mathrm{OBR}}]$
$\kappa = 0.1$	0.7626	1.0803	1.0998	1.0948	1.0991	1.0816	1.0812
$\kappa = 1.0$	0.8078	1.0296	1.0531	1.0520	1.0526	1.0458	1.0457
$\kappa = 5.0$	0.9417	1.0209	1.0525	1.0518	1.0513	1.0502	1.0501

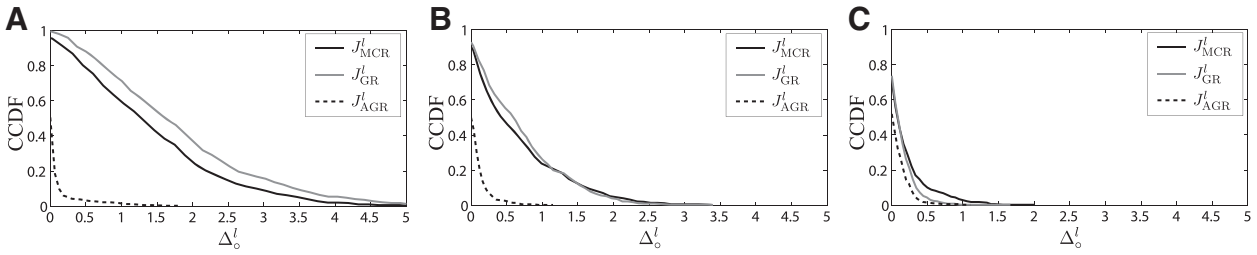

Figure 2 Empirical CCDF of $\Delta_\circ^l$ for different intervention policies across randomly generated PBNs with three genes. (A) $\kappa = 0.1$. **(B)** $\kappa = 1.0$. **(C)** $\kappa = 5.0$.

between the optimal and suboptimal costs to quantify the gain made by implementing an optimal policy. We define *percent decrease* by

$$\Delta_\circ^l = 100 \times \frac{J_\circ^l - J_\bullet^l}{J_\circ^l},$$

where $J_\circ^l$ and $J_\bullet^l$ denote two different intervention policies. Since PBNs are randomly generated, $\Delta_\circ^l$ will also be a random variable with a probability distribution. We estimate the complementary cumulative distribution function (CCDF) of this distribution for different values of $\Delta_\circ^l$ using its empirical distribution function.

For networks with three genes, we assume that $J_\bullet^l = J_{\mathrm{OBR}}^l$ and $J_\circ^l$ is any suboptimal policy. Figure 2 shows the empirical CCDF of $\Delta_\circ^l$ for 500 random PBNs for different values of κ and different intervention policies. The graphs illustrate that as the prior distribution gets tighter around the true TPM by increasing κ, the difference between the optimal and suboptimal policies vanishes. Again, the best performance among the suboptimal policies is achieved by J_{AGR}^l.

As suggested by these results, we may use the suboptimal AGR policy instead of an optimal method for larger networks without a significant lose of optimality. For this purpose, we carry out a similar set of simulations with 500 randomly generated PBNs of size $n = 4$ genes. We assume that each PBN consists of $m = 3$ equally likely constituent BNs with the maximum number of predictors for each node set to 2, the total number of states being 16. The network bias is drawn randomly from a beta distribution with mean 0.5 and standard deviation 0.01. The gene perturbation probability is $p = 0.001$. We generate prior distributions using (15) for each network and different parameter values for κ and ε. To model Ω, we draw 5,000 random TPMs from each prior distribution. We assume that each random sampling path has length 10 and set the discounting factor λ to 0.2. We emulate different sampling paths during the calculation of J_{AGR}^l by repeating the entire process for each randomly generated TPM for 1,000 iterations and take the average. The results averaged over 500 random PBNs are presented in Table 2. The AGR policy yields the best performance relative to other suboptimal policies.

We graph the empirical CCDF of $\Delta_\circ^l$ for these networks in Figure 3 for different values of the pair (κ, ε) and different suboptimal policies. Here, we have that $J_\bullet^l = J_{\mathrm{AGR}}^l$ and $J_\circ^l$ are any other suboptimal policy. Similar to networks with three genes, as the prior distributions get more concentrated around the true parameters, the

Table 2 Average costs across all 500 randomly generated PBNs with $n = 4$ genes

	$E[J^l_{\mathrm{LB}}]$	$E[J^l_{\mathrm{T}}]$	$E[J^l_{\mathrm{ET}}]$	$E[J^l_{\mathrm{MCR}}]$	$E[J^l_{\mathrm{GR}}]$	$E[J^l_{\mathrm{AGR}}]$
$(\kappa, \varepsilon) = (0.1, 0.0)$	0.7559	1.0878	1.0869	1.0856	1.0869	1.0773
$(\kappa, \varepsilon) = (1.0, 0.0)$	0.8702	1.0888	1.0888	1.0918	1.0888	1.0854
$(\kappa, \varepsilon) = (5.0, 0.0)$	0.9510	1.0579	1.0578	1.0612	1.0578	1.0572
$(\kappa, \varepsilon) = (0.1, 0.1)$	0.7711	1.1099	1.1260	1.1248	1.1258	1.1156
$(\kappa, \varepsilon) = (1.0, 0.1)$	0.8722	1.1106	1.1278	1.1314	1.1276	1.1236
$(\kappa, \varepsilon) = (5.0, 0.1)$	0.9714	1.0826	1.1011	1.1049	1.1009	1.1002
$(\kappa, \varepsilon) = (0.1, 0.25)$	0.7177	1.0796	1.1289	1.1234	1.1248	1.1133
$(\kappa, \varepsilon) = (1.0, 0.25)$	0.8307	1.0853	1.1348	1.1325	1.1305	1.1257
$(\kappa, \varepsilon) = (5.0, 0.25)$	0.9729	1.0629	1.1178	1.1157	1.1137	1.1130

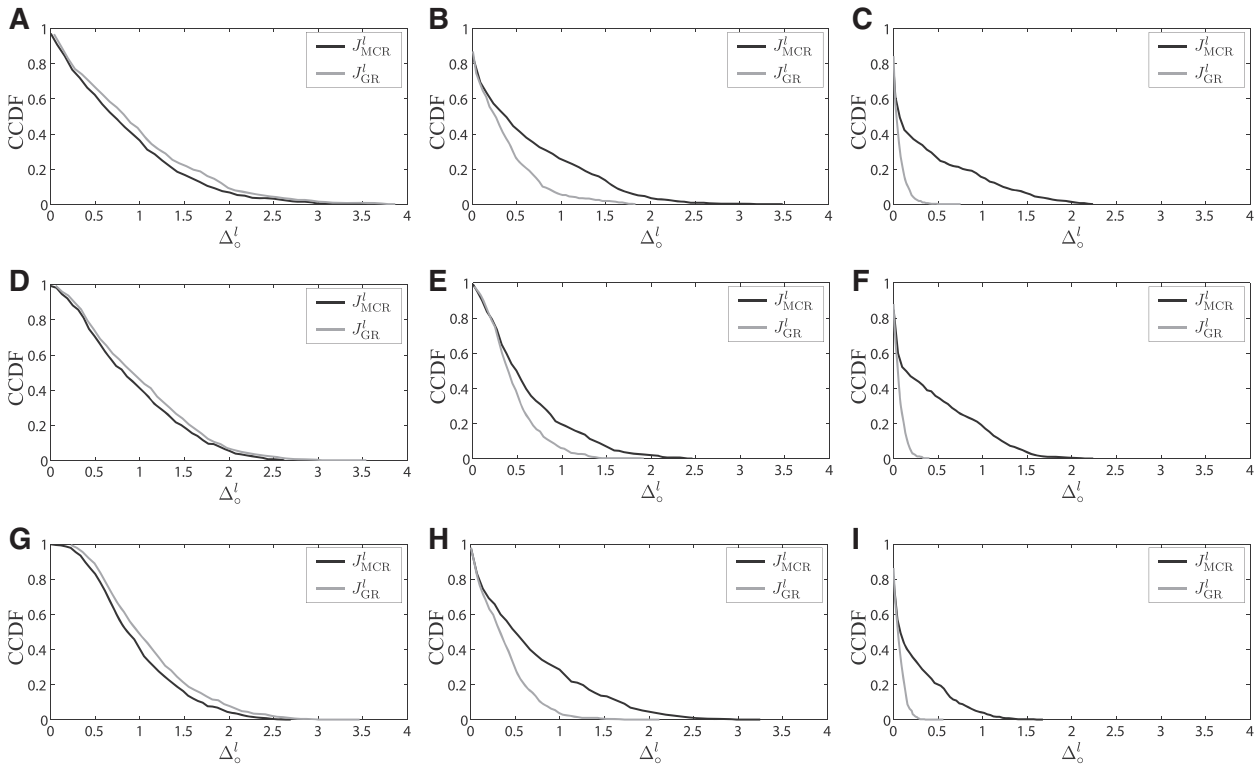

Figure 3 Empirical CCDF of $\Delta_\circ^l$ for different suboptimal intervention policies across randomly generated PBNs with four genes. (A) $(\kappa, \varepsilon) = (0.1, 0.0)$. **(B)** $(\kappa, \varepsilon) = (1.0, 0.0)$. **(C)** $(\kappa, \varepsilon) = (5.0, 0.0)$. **(D)** $(\kappa, \varepsilon) = (0.1, 0.1)$. **(E)** $(\kappa, \varepsilon) = (1.0, 0.1)$. **(F)** $(\kappa, \varepsilon) = (5.0, 0.1)$. **(G)** $(\kappa, \varepsilon) = (0.1, 0.25)$. **(H)** $(\kappa, \varepsilon) = (1.0, 0.25)$. **(I)**$(\kappa, \varepsilon) = (5.0, 0.25)$.

difference between these suboptimal policies gets smaller and smaller. However, it can be seen that the GR policy outperforms MCR for larger κ, which could be due to the fact that GR and AGR policies differ very little when the effect of observations on the posterior distribution is dominated by the prior hyperparameters.

Real network

We construct a PBN corresponding to a reduced network from a mutated mammalian cell cycle network proposed in [25]. The original GRN is a BN with ten genes. Three key genes in the model are Cyclin D (CycD), retinoblastoma (Rb), and p27, where cell division is coordinated with the overall growth of the organism through extracellular signals controlling the activation of CycD in the cell. A proposed mutation for this network is that p27 can never be activated (always OFF), creating a situation where both CycD and Rb might be inactive [25]. Under these conditions, the cell can cycle in the absence of any growth factor, thereby causing undesirable proliferation. Table 3 lists the Boolean functions for this real network.

Since the size of the network is too large for the Bayesian treatment, we need to first reduce the number of genes to a more manageable size while preserving important

Table 3 Boolean regulatory functions of a mutated mammalian cell cycle

Gene	Node	Predictor functions
CycD	v_1	Extracellular signal
Rb	v_2	$(\overline{v_1} \wedge \overline{v_4} \wedge \overline{v_5} \wedge \overline{v_9})$
E2F	v_3	$(\overline{v_2} \wedge \overline{v_5} \wedge \overline{v_9})$
CycE	v_4	$(v_3 \wedge \overline{v_2})$
CycA	v_5	$(v_3 \wedge \overline{v_2} \wedge \overline{v_6} \wedge (\overline{v_7 \wedge v_8})) \vee (v_5 \wedge \overline{v_2} \wedge \overline{v_6} \wedge (\overline{v_7 \wedge v_8}))$
Cdc20	v_6	v_9
Cdh1	v_7	$(\overline{v_5} \wedge \overline{v_9}) \vee v_6$
UbcH10	v_8	$\overline{v_7} \vee (v_7 \wedge v_8 \wedge (v_6 \vee v_5 \vee v_9))$
CycB	v_9	$(\overline{v_6} \wedge \overline{v_7})$

Table 4 Boolean regulatory functions of a reduced mutated mammalian cell cycle

Gene	Node	Predictor functions
CycD	v_1	Extracellular signal
Rb	v_2	$(\overline{v_1} \wedge v_2 \wedge \overline{v_3} \wedge \overline{v_5})$
CycA	v_3	$(\overline{v_2} \wedge v_3 \wedge \overline{v_5}) \vee (\overline{v_2} \wedge \overline{v_4} \vee \overline{v_5})$
UbcH10	v_4	$(v_4 \wedge v_5) \vee (v_3 \wedge \overline{v_5})$
CycB	v_5	$v_3 \wedge \overline{v_5}$

Table 5 Total discounted cost of different suboptimal policies for the reduced cell cycle network

	J_{LB}	J_{T}	J_{ET}	J_{MCR}	J_{GR}	J_{AGR}
$(\kappa,\varepsilon)=(0.1,0.0)$	0.7507	0.9685	0.9326	0.9465	0.9326	0.9316
$(\kappa,\varepsilon)=(1.0,0.0)$	0.4990	0.9685	0.9675	0.9614	0.9675	0.9571
$(\kappa,\varepsilon)=(5.0,0.0)$	0.6136	0.9685	0.9658	0.9774	0.9658	0.9605
$(\kappa,\varepsilon)=(0.1,0.1)$	0.4501	0.9685	0.9239	0.9268	0.9239	0.9144
$(\kappa,\varepsilon)=(1.0,0.1)$	0.5752	0.9685	0.9340	0.9526	0.9340	0.9294
$(\kappa,\varepsilon)=(5.0,0.1)$	0.7507	0.9685	0.9326	0.9465	0.9326	0.9316
$(\kappa,\varepsilon)=(0.1,0.25)$	0.3885	0.9685	0.8643	0.8674	0.8623	0.8550
$(\kappa,\varepsilon)=(1.0,0.25)$	0.5140	0.9685	0.8728	0.8860	0.8730	0.8694
$(\kappa,\varepsilon)=(5.0,0.25)$	0.7014	0.9685	0.8864	0.9002	0.8864	0.8861

dynamical properties of the network. We have implemented the methodology proposed in [26] and reduced the size of the network to the five genes shown in Table 4. Even for a network of this size, finding the OBR policy is computationally too expensive. Therefore, we only report results for suboptimal policies.

We first construct an instantaneously random PBN for the reduced network. The PBN consists of five genes, CycD, Rb, CycA, UbcH10 and CycB, ordered from the most significant bit to the least significant bit in the binary representation. In the mutated network, depending on the state of the extracellular signal determining the state of CycD as being ON or OFF, we obtain two BNs. These two will serve as two equally likely constituent BNs. It is also assumed that the gene perturbation probability is 0.01. Since cell growth in the absence of growth factors is undesirable, we define undesirable states of the state space to be those for which CycD and Rb are both down-regulated. We also choose CycA as the control gene. The immediate cost function is defined similarly to that of the synthetic network simulations (Equation 14). The discounting factor is $\lambda = 0.2$. We calculate the TPM of this network and construct prior hyperparameter matrices α using (15) for various pairs of κ and ε. We generate $10,000$ random TPMs from the prior distribution to represent the uncertainty class Ω. We also generate $10,000$ different sampling paths of length 10 for each random TPM. The total costs are reported in Table 5, where we can see that the results are consistent with those obtained from synthetic networks.

Conclusions

Due to the complex nature of Markovian genetic regulatory networks, it is commonplace not to possess accurate knowledge of their parameters. Under the latter assumption, we have treated the system of interest as an uncertainty class of TPMs governed by a prior distribution. The goal is to find a robust intervention policy minimizing the expected infinite-horizon discounted cost relative to the prior distribution. We have taken a Bayesian approach and formulated the intervention policy optimizing this cost, thereby resulting in an intrinsically robust policy. Owing to extreme computational complexity, the resulting OBR policy is, from a practical sense, infeasible. Using only a few genes, we have compared it to several suboptimal polices on synthetically generated PBNs. In this case, although there are PBNs where the OBR policy significantly outperforms the suboptimal AGR policy, on average there is very little difference. Hence, one can feel somewhat comfortable using the AGR policy while losing only negligible performance. Unfortunately, even the AGR policy is computationally burdensome. Hence, when applying it to the mammalian cell cycle network, we are restricted to five genes.

The twin issues of uncertainty and computational complexity are inherent to translational genomics. Here we have examined the problem in the context of therapy, where the uncertainty is relative to network structure. It occurs to also in the other major area of translational genomics, gene-based classification. Whereas here the prior distribution is over an uncertainty class of networks, in classification it is over an uncertainty class of feature-label distributions and one looks for a classifier that is optimal, on average, across that prior distribution [27,28]. There is no doubt, however, that the complexity issue is much graver in the case of dynamical intervention. Hence, much greater effort should be placed on gaining knowledge regarding biochemical pathways and thereby reducing the uncertainty when designing intervention strategies [29]. This means more attention should be paid to classical biological regulatory experiments and less reliance on blind data mining [30].

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MRY contributed to the main idea, designed and implemented the algorithms, designed and carried out the simulation, analyzed the results, and drafted the manuscript. ERD conceived the study, contributed in the design of the simulation, and revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgements
The authors thank the High-Performance Biocomputing Center of TGen for providing the clustered computing resources used in this study; this includes the Saguaro-2 cluster supercomputer, partially funded by NIH grant 1S10RR025056-01.

Author details
[1]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA. [2]Center for Bioinformatics and Genomic Systems Engineering, Department of Electrical and Computer Engineering, Texas A & M University, College Station, TX 77843, USA.

Received: 31 December 2013 Accepted: 18 March 2014
Published: 3 April 2014

References

1. ER Dougherty, R Pal, X Qian, ML Bittner, A Datta, Stationary and structural control in gene regulatory networks: basic concepts. Int. J. Syst. Sci. **41**(1), 5–16 (2010)
2. I Shmulevich, ER Dougherty, *Genomic Signal Processing* (Princeton University, Princeton, 2007)
3. I Shmulevich, ER Dougherty, S Kim, W Zhang, Probabilistic, Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics. **18**(2), 261–274 (2002)
4. R Pal, A Datta, ER Dougherty, Optimal infinite-horizon control for probabilistic Boolean networks. IEEE Trans. Signal Process. **54**(6), 2375–2387 (2006)
5. B Faryabi, G Vahedi, J-F Chamberland, A Datta, ER Dougherty, Optimal constrained stationary intervention in gene regulatory networks. EURASIP J. Bioinform. Syst. Biol. **2008**, 620767 (2008)
6. B Faryabi, J-F Chamberland, G Vahedi, A Datta, ER Dougherty, Optimal intervention in asynchronous genetic regulatory networks. IEEE J. Sel. Top. Signal. Process. **2**(3), 412–423 (2008)
7. MR Yousefi, A Datta, ER Dougherty, Optimal intervention in Markovian gene regulatory networks with random-length therapeutic response to antitumor drug. IEEE Trans. Biomed. Eng. **60**(12), 3542–3552 (2013)
8. MR Yousefi, ER Dougherty, Intervention in gene regulatory networks with maximal phenotype alteration. Bioinformatics. **29**(14), 1758–1767 (2013)
9. R Pal, A Datta, ER Dougherty, Robust intervention in probabilistic Boolean networks. IEEE Trans. Signal Process. **56**(3), 1280–1294 (2008)
10. R Pal, A Datta, ER Dougherty, Bayesian robustness in the control of gene regulatory networks. IEEE Trans. Signal Process. **57**(9), 3667–3678 (2009)
11. AM Grigoryan, ER Dougherty, Bayesian robust optimal linear filters. Signal Process. **81**(12), 2503–2521 (2001)
12. PR Kumar, A survey of some results in stochastic adaptive control. SIAM J. Contr. Optim. **23**(3), 329–380 (1985)
13. R Bellman, A problem in the sequential design of experiments. Sankhya: Indian J. Stat. **16**(3/4), 221–229 (1956)
14. R Bellman, R Kalaba, Dynamic programming and adaptive processes: mathematical foundation. IRE Trans. Automatic Control. **AC-5**(1), 5–10 (1960)
15. EA Silver, Markovian decision processes with uncertain transition probabilities or rewards. Technical report, DTIC document, (1963)
16. JM Gozzolino, R Gonzalez-Zubieta, RL Miller, Markovian decision processes with uncertain transition probabilities. Technical report, DTIC document (1965)
17. JJ Martin, *Bayesian Decision Problems and Markov Chains* (Wiley, New York, 1967)
18. SA SA Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. **22**(3), 437–467 (1969)
19. B Faryabi, G Vahedi, J-F Chamberland, A Datta, ER Dougherty, Intervention in context-sensitive probabilistic Boolean networks revisited. EURASIP J. Bioinform. Syst. Biol. **2009**(5) (2009)
20. X Qian, ER Dougherty, Effect of function perturbation on the steady-state distribution of genetic regulatory networks: optimal structural intervention. IEEE Trans. Signal Process. **56**(10), 4966–4976 (2008)
21. C Derman, *Finite State Markovian Decision Processes* (Academic, Orlando, 1970)
22. JK Satia, RE Lave, Markovian decision processes with uncertain transition probabilities. Oper. Res. **21**(3), 728–740 (1973)
23. MO Duff, Optimal learning: computational procedures for Bayes-adaptive Markov decision processes. PhD thesis, University of Massachusetts, Amherst (2002)
24. MR Yousefi, A Datta, ER Dougherty, Optimal intervention strategies for therapeutic methods with fixed-length duration of drug effectiveness. IEEE Trans. Signal Process. **60**(9), 4930–4944 (2012)
25. A Faure, A Naldi, C Chaouiya, D Thieffry, Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics. **22**(14), 124–131 (2006)
26. A Veliz-Cuba, Reduction of Boolean network models. J. Theor. Biol. **289**, 167–172 (2011)
27. LA Dalton, ER Dougherty, Optimal classifiers with minimum expected error within a Bayesian framework - Part I: discrete and gaussian models. Pattern Recogn. **46**(5), 1301–1314 (2013)
28. LA Dalton, ER Dougherty, Optimal classifiers with minimum expected error within a Bayesian framework - Part II: properties and performance analysis. Pattern Recogn. **46**(5), 1288–1300 (2013)
29. B-J Yoon, X Qian, ER Dougherty, Quantifying the objective cost of uncertainty in complex dynamical systems. IEEE Trans. Signal Process. **61**(9), 2256–2266 (2013)
30. ER Dougherty, ML Bittner, *Epistemology of the Cell: A Systems Perspective on Biological Knowledge* (Wiley, Hoboken, 2011)

doi:10.1186/1687-4153-2014-6
Cite this article as: Yousefi and Dougherty: **A comparison study of optimal and suboptimal intervention policies for gene regulatory networks in the presence of uncertainty.** *EURASIP Journal on Bioinformatics and Systems Biology* 2014 **2014**:6.