EURASIP Journal on Applied Signal Processing 2004:1, 125–131

Reduction Mappings between Probabilistic Boolean Networks

Ivan Ivanov
Department of Electrical Engineering, Texas A&M University, College Station, TX 77843, USA
Email: ivanov@ee.tamu.edu

Edward R. Dougherty
Department of Electrical Engineering, Texas A&M University, 3128 TAMU College Station, TX 77843-3128, USA
Email: e-dougherty@tamu.edu

Received 11 April 2003; Revised 28 August 2003

Probabilistic Boolean networks (PBNs) comprise a model describing a directed graph with rule-based dependences between its nodes. The rules are selected, based on a given probability distribution which provides a flexibility when dealing with the uncertainty which is typical for genetic regulatory networks. Given the computational complexity of the model, the characterization of mappings reducing the size of a given PBN becomes a critical issue. Mappings between PBNs are important also from a theoretical point of view. They provide means for developing a better understanding about the dynamics of PBNs. This paper considers two kinds of mappings reduction and projection and their effect on the original probability structure of a given PBN.

Keywords and phrases: Boolean network, genetic network, graphical models, projection, reduction.

1. INTRODUCTION

Given a set of genes, the evolution of their expressions constitutes a dynamical system over time. Owing to the complexity of gene interaction and the paucity of data, homogeneous transitions are customarily assumed. Many different gene-regulatory-network models have been proposed. Among deterministic dynamical systems, perhaps, the most attention has been given to the Boolean network model [1, 2, 3]. In this model, gene expression is quantized to only two levels: ON and OFF. The expression level (state) of a gene is functionally related, via a logical rule, to the expression states of some other genes. The Boolean network model has yielded insights into the overall behavior of large genetic networks [4, 5, 6, 7], thereby facilitating the study of large data sets in a global fashion. Here, we are concerned with a stochastic extension of the Boolean model that results in probabilistic Boolean networks [8, 9]. For these, similarities exist with Bayesian networks [10, 11, 12, 13] and, more generally, with models including stochastic components on the molecular level [14, 15, 16].

The dynamical behavior of such networks can be used to model many biologically meaningful phenomena—for instance, cellular state dynamics, possessing switch-like behavior, stability, and hysteresis [17]. Besides the conceptual framework offered by such models, there are practical uses, such as the identification of suitable drug targets in cancer therapy or inferring the structure of the genetic models from experimental data, for example, from the gene expression profiles [17]. To that end, a significant effort has gone into identifying the structure of gene regulatory networks from expression data [8, 18, 19, 20, 21, 22, 23].

Probabilistic Boolean networks (PBNs) [8, 9] constitute a probabilistic generalization of Boolean networks and offer a more powerful and flexible modeling framework. They share the appealing rule-based properties of the Boolean networks, are robust to uncertainty both in the data and model selection, and can be studied in the probabilistic context of Markov chains (see also [23]). PBNs enable the systematic study of global network dynamics and permit quantification of the relative influence and sensitivity of genes in their interactions with other genes. While the Boolean assumption is useful for a simple up- or down-regulated model and also useful for reducing the complexity of the network, the basic model extends directly to a finite-state-space model, and inference has been studied in that context in [22].

A principle reason for studying regulatory models is to develop intervention strategies to help in guiding the time evolution of the network towards more desirable states. Three distinct approaches to the intervention problem have been considered in the context of PBNs by exploiting their Markovian nature. First, one can toggle the expression status

of a particular gene from ON to OFF or vice versa to facilitate the transition to some other desirable state or set of states. Specifically, using the concept of the mean first passage time, it has been demonstrated how the particular gene, whose transcription status is to be momentarily altered to initiate the state transition, can be chosen to "minimize" (in a probabilistic sense) the time required to achieve the desired state transitions [24]. A second approach has aimed at changing the steady-state (long-run) behavior of the network by minimally altering its rule-based structure [25]. A third approach has focused on applying ideas from control theory to develop an intervention strategy in the general context of Markovian genetic regulatory networks whose state transition probabilities depend on an external (control) variable [26].

An obstacle in applying PBNs is the computational complexity of the model. Owing to the large number of states often present in full networks, it is sometimes necessary to construct computationally tractable subnetworks while still carrying sufficient structure for the application at hand—hence, the need for size reducing mappings between PBNs. Construction of mappings to alter PBN structure while at the same time maintaining consistency with the original probability structure have previously been studied [27]. These include projections onto subnetworks. Unfortunately, while projections maintain the probabilistic structure by reducing the number of genes, they also increased the complexity of the Boolean function structure. This paper considers reduction mappings of a PBN that alter the structure of the network while maintaining maximum consistency with the original probability structure. Once this notion of maximum consistency has been defined, the problem reduces to one of optimization. Thus, a key issue to be addressed in this paper is the positing of consistency conditions.

2. DEFINITIONS AND BASIC PROPERTIES

This section provides the definitions and the basic properties of probabilistic Boolean networks as given in [8]. While there have been some generalization of the model [9, 24], we stay with the original definition, as has the original analysis of projection mappings between PBNs [27]—which plays a key role in the present paper. A PBN (V, F, C) is defined by a set of nodes (genes)

$$\begin{aligned} V &= \{x_1, \ldots, x_n\}, \\ x_i &\in \{0, 1\}, \\ i &= 1, \ldots, n, \end{aligned} \tag{1}$$

a list of predictors

$$\begin{aligned} F &= (F_1, \ldots, F_n), \\ F_i &= \left\{f_1^{(i)}, \ldots, f_{l(i)}^{(i)}\right\}, \\ f_j^{(i)} &: \{0, 1\}^n \longrightarrow \{0, 1\}, \end{aligned} \tag{2}$$

and a list

$$\begin{aligned} C &= (C_1, \ldots, C_n), \\ C_i &= \left\{c_1^{(i)}, \ldots, c_{l(i)}^{(i)}\right\} \end{aligned} \tag{3}$$

of selection probabilities $c_j^{(i)} = \Pr\{f^{(i)} = f_j^{(i)}\}$ with respect to a list (vector) of probability distributions $(\nu^{(1)}, \ldots, \nu^{(n)})$, where $\mathbf{f} = (f^{(1)}, \ldots, f^{(n)})$ is a random vector taking values in F. Each node x_i represents the state (expression) of the gene i, where $x_i = 0$ means that the gene i is not expressed and $x_i = 1$ means that it is expressed. Every set F_i contains the possible rules $f_j^{(i)}$ of regulatory interactions for the gene i. These functions are also called *predictors* for the corresponding gene. Updating of the states of all genes in the network is done synchronously according to the functions assigned to the genes, and then the process is repeated. The predictors for every gene x_i are selected simultaneously and randomly (according to the list C) from the sets F_i at every time step.

A *realization* of a PBN is determined at every time step by the vector $\mathbf{f}$. If the predictor for each gene is chosen independently of the other predictors, then the number of all possible realizations $\mathbf{f}_k = (f_{k_1}^{(1)}, \ldots, f_{k_n}^{(n)})$, $k = 1, \ldots, N$, of the PBN is $N = \prod_{j=1}^{n} l(j)$. Even though the domain of every predictor $f_j^{(i)}$ is assumed to be $\{0, 1\}^n$, there are only a few input genes that actually regulate x_i at any given time step. This simplification can be justified by some biological and practical considerations [8]. In general, there is no need of the assumption that $f^{(1)}, f^{(2)}, \ldots, f^{(n)}$ are selected independently; however we make this assumption. A PBN that satisfies this assumption is called *independent*. For an independent PBN, we have

$$P_k = \Pr\{\mathbf{f} = \mathbf{f}_k\} = \prod_{j=1}^{n} \Pr\left\{f^{(i)} = f_{k_j}^{(i)}\right\} = \prod_{j=1}^{n} c_{k_j}^{(i)}. \tag{4}$$

In [8], the list C of selection probabilities is created using the *coefficient of determination* [28, 29].

A PBN can be interpreted as a homogeneous Markov chain relative to the states $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of the network with transition probabilities given by

$$\Pr\{\mathbf{x} \longrightarrow \mathbf{x}'\} = \sum_i P_i, \tag{5}$$

where the summation is over the indices i such that $i : f_{K_{i1}}^{(i)}(x_1, \ldots, x_n) = x_1', \ldots, f_{K_{in}}^{(n)}(x_1, \ldots, x_n) = x_n'$ and K is the matrix with rows given by the possible realizations of the PBN [8].

3. PBN PROJECTION MAPPING

Projection mappings of a PBN A are defined in [27]. They are introduced as an attempt to reduce the complexity of A while maintaining consistency with the original probability structure of the PBN. The basic projection Π_i is a mapping that

transforms the given PBN into a new one, where the number of the genes is reduced by one, that is, the gene x_i in the original network is "deleted." Without loss of generality, we may assume that the deleted gene is the last one, x_n. Thus, $\Pi_n : A \to \hat{A}, \hat{A}(\hat{V}, \hat{F}, \hat{C})$, where

$$\hat{V} = \{x_1, \ldots, x_{n-1}\}, \qquad \hat{F} = (\hat{F}_1, \ldots, \hat{F}_{n-1}), \\ \hat{C} = (\hat{C}_1, \ldots, \hat{C}_{n-1}). \tag{6}$$

Every $\hat{F}_i$ and every $\hat{C}_i$ have twice as many elements as the corresponding sets F_i and C_i in A. Every predictor $f_j^{(i)} \in F_i$ generates two predictors $\hat{f}_{0j}^{(i)}$ and $\hat{f}_{1j}^{(i)}$ according to the rule

$$\hat{f}_{kj}^{(i)}(x_1, \ldots, x_{n-1}) = f_j^{(i)}(x_1, \ldots, x_{n-1}, k), \tag{7}$$

where $k \in \{0, 1\}$ and $(x_1, \ldots, x_{n-1})$ is in $\hat{A}$. The new Boolean functions $\hat{f}_{kj}^{(i)}$, $k = 0, 1$, have transition probabilities

$$\hat{c}_{kj}^{(i)} = c_j^{(i)} \Pr\{x_n = k\}, \quad k \in \{0, 1\}. \tag{8}$$

It is noticed in [27] that there is a difficulty in defining the new selection probabilities $\hat{c}_{kj}^{(i)}$ because the probabilities for the gene x_n depend on the current state probability distribution of the underlying Markov chain. One way to go around the problem is to use the steady state distribution for A or, the stationary distribution for A if there is no steady state distribution. Another way is to estimate $\Pr\{x_n = k\}$, $k = 0, 1$, by running A for some time. In doing so, one has to be aware of the possible transient behavior of those probabilities. Yet, another way to find the values of $\Pr\{x_n = k\}$ is to use the data set from which the original PBN A was created.

4. PBN REDUCTION MAPPINGS

In this paper, we propose a new kind of mapping that also reduces the size of a given PBN. In contrast to the projection mapping discussed in the previous section, this new mapping does not increase the number of the predictors for the genes that remain in the new network. One has to keep in mind that any such mapping might not preserve the probability structure of the original PBN. For example, this will be the case if the deleted gene is *essential* for one of the predictors of the remaining genes [8].

Therefore, the problem is to find a reduction mapping that renders a PBN close to the original one. To be more specific, consider an independent PBN $A(V, F, C)$ and a mapping $\pi_n : A \to \tilde{A}, \tilde{A}(\tilde{V}, \tilde{F}, \tilde{C})$, where

$$\tilde{V} = \{x_1, \ldots, x_{n-1}\}, \qquad \tilde{F} = (\tilde{F}_1, \ldots, \tilde{F}_{n-1}), \\ \hat{C} = (\hat{C}_1, \ldots, \hat{C}_{n-1}), \tag{9}$$

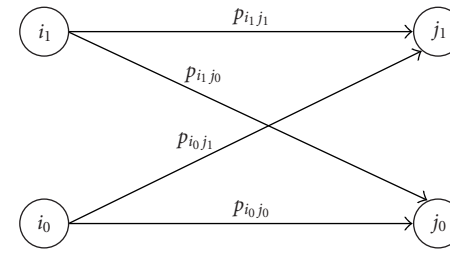
where $\tilde{A}(\tilde{V}, \tilde{F}, \tilde{C})$ is an independent PBN with $\tilde{F}_i = \{\tilde{f}_1^{(i)}, \ldots, \tilde{f}_{l(i)}^{(i)}\}$, $\tilde{f}_j^{(i)} : \{0, 1\}^{n-1} \to \{0, 1\}$, and $\tilde{c}_j^{(i)} = \tilde{\Pr}\{\tilde{f}^{(i)} = \tilde{f}_j^{(i)}\}$ with respect to some probability distribution vector $(\tilde{\nu}^{(1)}, \ldots, \tilde{\nu}^{(n-1)})$. Note that the cardinality of $\tilde{F}_i$ is the same as the cardinality of F_i. The new PBN $\tilde{A}$ is called a *reduced* PBN obtained from the original PBN A by deleting one of the genes in A. As in Section 3, we have assumed without loss of generality that the deleted gene is x_n.

The reduction π_n should yield a PBN that is "close" to the original, and there are various natural ways to interpret this closeness:

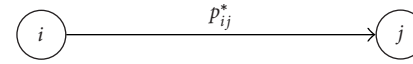
(A) for every $\tilde{c}_j^{(i)}$, $|\tilde{c}_j^{(i)} - c_j^{(i)}| \le \epsilon$ for some given $\epsilon \ge 0$;
(B) the transition probabilities for the state diagrams of A and $\tilde{A}$ are close;
(C) the stationary/steady-state distributions D of A and $\tilde{D}$ of $\tilde{A}$ are close;
(D) every new predictor function $\tilde{f}_j^{(i)}$ is selected as close as possible to both functions $\hat{f}_{kj}^{(i)}$, $k = 0, 1$, given by the projected PBN $\hat{A}$.

Some comments about the preceding conditions are in order.

(A′) In the context of gene regulatory networks, one can expect the number ϵ to be reasonably small, and perhaps even equal to zero, that is, the predictors for the genes in the reduced PBN $\tilde{A}$ have the same selection probabilities as their corresponding predictors from the original PBN A.
(B′) Consider the portion of the state diagram of A containing the states $i_1 = (x_1, \ldots, x_{n-1}, 1)$, $i_0 = (x_1, \ldots, x_{n-1}, 0)$, $j_1 = (x'_1, \ldots, x'_{n-1}, 1)$, and $j_0 = (x'_1, \ldots, x'_{n-1}, 0)$:



where $p_{i_1 j_1}$, $p_{i_0 j_0}$, $p_{i_1 j_0}$, and $p_{i_0 j_1}$ are the corresponding transition probabilities. If one "deletes" the node x_n, this diagram collapses to the following one:



where $i = (x_1, \ldots, x_{n-1})$ and $j = (x'_1, \ldots, x'_{n-1})$ are the corresponding states in $\tilde{A}$, and

$$p_{ij}^* = \Pr\{x_n = 1\}(p_{i_1 j_1} + p_{i_1 j_0}) \\ + \Pr\{x_n = 0\}(p_{i_0 j_1} + p_{i_0 j_0}). \tag{10}$$

The transition probabilities for the reduced PBN $\tilde{A}$ are given by (see [8])

$$\tilde{p}_{ij} = \sum_i \tilde{P}_i, \tag{11}$$

where the summation is over the indices i such that $i : \tilde{f}_{K_{i1}}^{(i)}(x_1, \ldots, x_{n-1}) = x'_1, \ldots, \tilde{f}_{K_{in-1}}^{(i)}(x_1, \ldots, x_{n-1}) = x'_{n-1}$.

Since the transition probability matrices for A and $\tilde{A}$ have different dimensions, one cannot compare them directly. This is why we compare the $\tilde{p}_{ij}$'s to the p^*_{ij}'s, and the term "close" in part (B) refers to the quantity $\max_{i,j}|\tilde{p}_{ij} - p^*_{ij}|$ being small.

(C′) Collapsing the state transition diagram, as described in part (B′), induces a probability distribution D^* on the state space of $\tilde{A}$ in the following way:

$$\begin{aligned}\Pr^*\{\text{state in } \tilde{A} = (x_1,\ldots,x_{n-1})\} &= \Pr\{\text{state in } A = (x_1,\ldots,x_{n-1},0)\} \\ &\quad + \Pr\{\text{state in } A = (x_1,\ldots,x_{n-1},1)\}.\end{aligned} \tag{12}$$

Notice that one cannot compare the distribution D to the distribution $\tilde{D}$ directly because they are defined over different state spaces. This is why the term "close" in (C) refers to the closeness of the distribution D^* and the stationary state distribution $\tilde{D}$ of $\tilde{A}$ in the l_1 sense, that is, to the quantity

$$\|D^* - \tilde{D}\|_{l_1} = \sum_{\mathbf{x}\in\tilde{S}} |\Pr^*\{\mathbf{x}\} - \tilde{\Pr}\{\mathbf{x}\}| \tag{13}$$

being small. Here, $\tilde{S} = \{0,1\}^{n-1}$ is the set of all states in $\tilde{A}$, and $\tilde{\Pr}$ is associated with $\tilde{D}$.

(D′) Using the notation from (D), we have the following proposition.

Proposition 1. *Given a PBN A with a stationary state distribution D, consider the projected PBN $\hat{A}$. Then*

$$E_D\left(\frac{\partial f_j^{(i)}}{\partial x_n}\right) = \left\|\hat{f}_{0j}^{(i)} - \hat{f}_{1j}^{(i)}\right\|_{l_1^{2^{n-1}}} = E_{D^*}\left(\left|\hat{f}_{0j}^{(i)} - \hat{f}_{1j}^{(i)}\right|\right). \tag{14}$$

Here the space $l_1^{2^{n-1}}$ is endowed with the probability measure $\tilde{\Pr}$ defined by the distribution D^*, and E_D means the expectation of the corresponding random variable with respect to the distribution D.

Proof. The claim in this proposition becomes obvious if one notices that for every state $(x_1,\ldots,x_{n-1},x_n) \in \{0,1\}^n$, where $\partial f_j^{(i)}/\partial x_n = 1$, there are two terms in the sum that compute $E_D(\partial f_j^{(i)}/\partial x_n)$, namely, $\Pr\{(x_1,\ldots,x_{n-1},0)\}$ and $\Pr\{(x_1,\ldots,x_{n-1},1)\}$. □

The proposition plays an important role in selecting the new predictor function $\tilde{f}_j^{(i)}$. Notice that the expectation $E_D(\partial f_j^{(i)}/\partial x_n)$ represents the influence $I_n(f_j^{(i)})$ of the gene x_n on the predictor $f_j^{(i)}$ (cf. [8]). In the special case when x_n is not essential for the function $f_j^{(i)}$, the new predictor can be selected to be identically equal to either of the two possible predictors $\hat{f}_{0j}^{(i)}$ and $\hat{f}_{1j}^{(i)}$ in the projected PBN $\hat{A}$. Generally speaking, the selection of the new predictor $\tilde{f}_j^{(i)}$ should minimize both $E_{D^*}(|\hat{f}_{0j}^{(i)} - \tilde{f}_j^{(i)}|)$ and $E_{D^*}(|\tilde{f}_j^{(i)} - \hat{f}_{1j}^{(i)}|)$. The inequality

$$\begin{aligned}E_{D^*}\left(\left|\hat{f}_{0j}^{(i)} - \hat{f}_{1j}^{(i)}\right|\right) &\le E_{D^*}\left(\left|\hat{f}_{0j}^{(i)} - \tilde{f}_j^{(i)}\right|\right) \\ &\quad + E_{D^*}\left(\left|\tilde{f}_j^{(i)} - \hat{f}_{1j}^{(i)}\right|\right)\end{aligned} \tag{15}$$

provides a measurement of how well the reduction mapping preserves the predictors from the original PBN. "Deleting" a gene x_k with bigger influence $I_k(f_j^{(i)})$ on the predictor $f_j^{(i)}$ produces a new predictor $\tilde{f}_j^{(i)}$ which cannot be closer to $f_j^{(i)}$ when compared to the new predictor resulting from the "deletion" of a gene x_l with smaller influence $I_l(f_j^{(i)})$ on $f_j^{(i)}$. In other words, "deleting" essential genes from the original PBN comes with a "price"—the predictor functions for the reduced PBN cannot be too close to the original predictors.

The selection of every function $\tilde{f}_j^{(i)} \in \tilde{F}_i$ has to be performed pointwise, that is, for each state in $\tilde{S}$, define

$$U = \left\{\mathbf{x} = (x_1,\ldots,x_{n-1}) \in \tilde{S} : \hat{f}_{0j}^{(i)}(\mathbf{x}) = \hat{f}_{1j}^{(i)}(\mathbf{x})\right\} \tag{16}$$

and $W = \tilde{S} \setminus U$. Clearly, $\tilde{f}_j^{(i)} \equiv \hat{f}_{0j}^{(i)} \equiv \hat{f}_{1j}^{(i)}$ on the set U. For the states in the remaining set W, one has to decide to what degree one favors certain states in $S = \{0,1\}^n$ which in its turn defines $\tilde{f}_j^{(i)}$ as either equal to $\hat{f}_{0j}^{(i)}$ or to $\hat{f}_{1j}^{(i)}$. Motivated by the preceding remarks about the conditions (A), (B), (C), and (D), we now design a selection procedure for the functions $\tilde{f}_j^{(i)}$.

Selection procedure

(a) For all i, j, select numbers $-1 \le \omega_j^{(i)} \le 1$.

(b) For every state $\mathbf{x} = (x_1,\ldots,x_{n-1}) \in W$, define

$$\tilde{f}_j^{(i)}(\mathbf{x}) = \begin{cases} \hat{f}_{0j}^{(i)}(\mathbf{x}) & \text{if } \Pr\{(x_1,\ldots,x_{n-1},0)\} \\ & \quad > \omega_j^{(i)} + \Pr\{(x_1,\ldots,x_{n-1},1)\}; \\ \hat{f}_{1j}^{(i)}(\mathbf{x}) & \text{otherwise.} \end{cases} \tag{17}$$

(c) For every state $\mathbf{x} = (x_1,\ldots,x_{n-1}) \in U$, set $\tilde{f}_j^{(i)}(\mathbf{x}) = \hat{f}_{1j}^{(i)}(\mathbf{x})$.

Notice that the condition on the numbers $\omega_j^{(i)}$ is natural since we are dealing with probabilities.

Our selection procedure leads to the following optimization problem.

Problem 1. Find $\tilde{F}$ that achieves $\min_\Omega \max_{i,j}|\tilde{p}_{ij} - p^*_{ij}|$ subject to

(i) $\tilde{c}_j^{(i)} = c_j^{(i)}$, $1 \le i \le n-1$, $1 \le j \le l(n)$,

(ii) $\Omega = \{\omega_j^{(i)} : -1 \le \omega_j^{(i)} \le 1, 1 \le i \le n-1, 1 \le j \le l(n)\}$.

Remark 1. The above problem has a solution: it is enough to notice that Ω is a compact set.

Remark 2. From a computational point of view, the only values for $\omega_j^{(i)}$ one should consider are the differences

$\Pr\{(x_1,\ldots,x_{n-1},0)\} - \Pr\{(x_1,\ldots,x_{n-1},1)\}$. This essentially reduces Ω to a finite set. Notice that if all $\omega_j^{(i)} \equiv -1$, then $\tilde{f}_j^{(i)}(\mathbf{x}) \equiv \hat{f}_{0j}^{(i)}(\mathbf{x})$. The other extreme choice is when all $\omega_j^{(i)} \equiv 1$ which produces $\tilde{f}_j^{(i)}(\mathbf{x}) \equiv \hat{f}_{1j}^{(i)}(\mathbf{x})$. One can see that different choices for the $\omega_j^{(i)}$'s could be based on how much one favors certain states in the original PBN. In the following simulations, we always set $\omega_j^{(i)} = 0$ which means that we do not assume any additional information, and the selection of the new predictor functions is based only on the probability distribution of the states of A.

5. COMPARISON BETWEEN THE PROJECTION AND THE REDUCTION MAPS

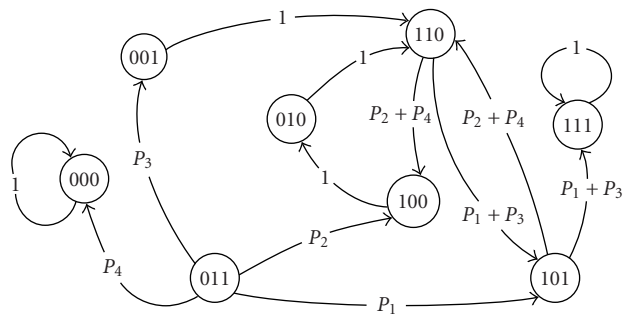
One should immediately notice the difference in defining the reduction and the projection mappings. While the projection is based on the probability distribution of a single gene, the reduction mapping is defined using the probability distribution of the entire collection of states of the given PBN. To illustrate this difference, we consider one particular example of a PBN (cf. [8]).

Example 1. Let $A(V,F,C)$ be a PBN consisting of three genes $V = \{x_1,x_2,x_3\}$ and function sets $F = (F_1,F_2,F_3)$, where $F_1 = \{f_1^{(1)}, f_2^{(1)}\}$, $F_2 = \{f_1^{(2)}\}$, and $F_3 = \{f_1^{(3)}, f_2^{(3)}\}$, and the predictor functions are given by the truth table (Table 1).

Table 1

$x_1x_2x_3$	$f_1^{(1)}$	$f_2^{(1)}$	$f_1^{(2)}$	$f_1^{(3)}$	$f_2^{(3)}$
000	0	0	0	0	0
001	1	1	1	0	0
010	1	1	1	0	0
011	1	0	0	1	0
100	0	0	1	0	0
101	1	1	1	1	0
110	1	1	0	1	0
111	1	1	1	1	1
$c_j^{(i)}$	0.6	0.4	1	0.5	0.5

After computing the transition probabilities, (cf. [8]), we arrive at the following directed graph/state transition diagram:



Here, $P_1 = 0.3$, $P_2 = 0.3$, $P_3 = 0.2$, and $P_4 = 0.2$. Next, we start with a uniform state probability distribution $D_{in} = \{1/8, 1/8, \ldots, 1/8\}$ for the states in the state space S of A, and then run the corresponding Markov chain for some large number of iterations. Notice that even if the given network does not possess a steady state distribution, the result after running the Markov chain sufficiently long time is approximately the stationary state distribution D that corresponds to D_{in}. The simulation gives $D = \{0.15, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.85\}$. Using this distribution, one can compute the projected $\hat{A}$ and the reduced $\tilde{A}$ networks, as well as their probability transition matrices. After running the Markov processes associated with these two transition probability matrixes, we obtain the stationary state distributions $\hat{D}$ for $\hat{A}$, and $\tilde{D}$ for $\tilde{A}$. For the case when the deleted gene is x_3, we get

$$\begin{aligned} \hat{D} &= \{0.006734, 0.022778, 0.138384, 0.831647\}, \\ \tilde{D} &= \{0.35, 0.0, 0.0, 0.65\}. \end{aligned} \tag{18}$$

The stationary state distribution for the transition probability matrix $(p_{i,j}^*)_{i=1,\,j=1}^{4,4}$, produced after the collapsing procedure described in part (D) (Section 4), is $D_1 = \{0.008145, 0.020362, 0.135747, 0.835747\}$. One can notice the similarity between D_1 and $\hat{D}$ and their apparent difference from $\tilde{D}$. At the same time, the distribution $D^* = \{0.15, 0.0, 0.0, 0.85\}$ described in part (D), Section 4, is similar to $\tilde{D}$. This should not be surprising—both the projection and the "collapsing" mappings are based on the probability distribution of a single gene, x_3 in our example, while the reduction mapping is based on the probability distribution of the entire collection of states in the original PBN. Thus the optimization criterion described in Problem 1 becomes a natural compromise between these two possible approaches of reducing the original PBN size.

Inequality (15) can be used in deciding which gene, after being eliminated from the network, will have a minimal impact on the stationary distribution of the original PBN. Since the left-hand side of (15) represents the influence $I_n(f_j^{(i)})$ of a gene x_n on the predictor $f_j^{(i)}$, one can say that, in general, deleting genes with smaller influences on the remaining predictors will result in a better chance of preserving the stationary state distribution of the original PBN. Here, we provide the values for the influences of x_3 on the remaining predictors and then two more simulations for the same example, where the other two possible genes x_2 and x_1 are deleted from the original PBN. The influences of x_3 on the remaining predictors are $I_3(f_1^{(1)}) = 0.15$, $I_3(f_2^{(1)}) = 0.15$, and $I_3(f_1^{(2)}) = 1$. After deleting x_2 from A, the corresponding stationary state distribution is $\tilde{D} = \{0.25, 0.0, 0.0, 0.75\}$, and the influences of x_2 on the remaining predictors are $I_2(f_1^{(1)}) = 0.15$, $I_2(f_2^{(1)}) = 0.15$, $I_2(f_3^{(1)}) = 0$, and $I_2(f_3^{(2)}) = 0.85$. After deleting x_1 from A, the corresponding stationary state distribution is $\tilde{D} = \{0.5, 0.0, 0.0, 0.5\}$, and the influences of x_1 on the remaining predictors are $I_1(f_2^{(1)}) = 1$, $I_1(f_3^{(1)}) = 0$, and $I_1(f_3^{(2)}) = 0.85$.

It appears that the gene x_1 with the biggest total influence distorts the stationary state distribution the most but one should be careful when generalizing this observation. Gene influences can be computed based on different probability distributions (cf. [8]). In addition, deleting different genes from the original PBN results in reduced PBNs with different state spaces. Finally, the left-hand side of (15) is just a lower bound that governs the selection procedure in constructing $\tilde{A}$, and that the lower bound might not be achieved during the selection procedure.

6. SIMULATION RESULTS

The reduction mapping has been tested using coefficient of determination (COD) microarray data for a network A consisting of 10 genes [23]. The genes of interest in the network are *PIRIN, WNT5A, S100P, RET-1, MMP-3, PHO-C, STC2, MART-1, HADHB*, and *SYNUCLEIN*. The network is reduced down to 7 genes by subsequently deleting the last three genes, starting with *SYNUCLEIN*. Table 2 presents lists of some of the states in the stationary/steady distributions for the full network A and the reduced networks $\tilde{A}_{10}$, $\tilde{A}_{10,9}$, and $\tilde{A}_{10,9,8}$, where the indices indicate which genes in A are deleted. For example, $\tilde{A}_{10,9,8}$ is the reduced network after deleting the genes *SYNUCLEIN*, *HADHB*, and *MART-1*. The states are presented by binary strings of ten digits, where 0 indicates that the corresponding gene is "OFF" and 1 indicates that the corresponding gene is "ON." The leftmost digit represents *PIRIN* and then the remaining digits represent the following genes in the network with the rightmost digit representing *SYNUCLEIN*. Next to every given state, its corresponding weight in the stationary state distribution of the network is given. Only states with weight bigger than 0.0001 are shown.

One can notice the presence of a very "heavy" state, 1010000111, in the stationary/steady state distribution of the full network. That is in agreement with the COD data set, where the same state is present in 8 out of 31 samples (see [23] for a related discussion). The reduction mapping maintains the structure of the stationary state distribution of the full network, specifically, the states 101000011, 10100001, and 1010000 carry most of the weight in the stationary/steady state distributions of their corresponding reduced networks.

7. CONCLUSION

The new mapping introduced in this paper offers a way of reducing the size of a given PBN by using the stationary probability distribution on the state space of the PBN. At the same time, it minimizes the distance between the reduced network and the projected PBN introduced in [27]. The distance is given in terms of the distance between their corresponding probability transition matrices. One should notice that the construction of the projected PBN is based on the probability distribution of a single gene, and that the same single gene probability distribution could happen under many different stationary distributions on the state space of the original PBN.

Table 2

For the full network A:
(0000111000, 0.003773); (0000111001, 0.003117);
(0001111000, 0.001905); (0010000111, 0.001715);
(0100011000, 0.001030); (0100101000, 0.010710);
(0100101001, 0.012694); (0100111000, 0.023957);
(0100111001, 0.026482); (0101101000, 0.004985);
(0101101001, 0.001685); (0101111000, 0.011730);
(0101111001, 0.003710); (0110111001, 0.001352);
(0111111000, 0.001416); (1010000101, 0.002299);
(1010000111, 0.832929)
For $\tilde{A}_{10}$:
(000011100, 0.010795); (000111100, 0.002513);
(001011100, 0.001944); (010010100, 0.140539);
(010011100, 0.083694); (010110100, 0.020368);
(010111000, 0.001241); (010111100, 0.014547);
(011011000, 0.001116); (011011100, 0.005920);
(011111000, 0.001743); (011111100, 0.003310);
(101000010, 0.001003); (101000011, 0.689413)
For $\tilde{A}_{10,9}$:
(00000110, 0.001528); (00001110, 0.017951);
(00011110, 0.004141); (00101110, 0.003209);
(01000110, 0.001423); (01001010, 0.230668);
(01001100, 0.001481); (01001110, 0.137728);
(01011010, 0.033485); (01011100, 0.002186);
(01011110, 0.024005); (01101100, 0.001911);
(01101110, 0.009774); (01111100, 0.003293);
(01111110, 0.005520); (10100001, 0.499967)
For $\tilde{A}_{10,9,8}$:
(0000011, 0.001523); (0000111, 0.020527);
(0001110, 0.001065); (0001111, 0.004695);
(0010111, 0.003700); (0100011, 0.001814);
(0100101, 0.269151); (0100110, 0.001628);
(0100111, 0.160629); (0101101, 0.039116);
(0101110, 0.002575); (0101111, 0.027858);
(0110110, 0.002169); (0110111, 0.011428);
(0111110, 0.004034); (0111111, 0.006365);
(1010000, 0.429211)

ACKNOWLEDGMENTS

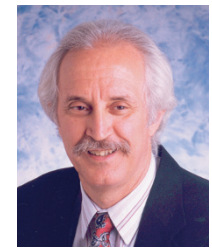
This research was supported by the National Cancer Institute (CA90301), the National Human Genome Research Institute, and the Translational Genomics Research Institute.

REFERENCES

[1] L. Glass and S. A. Kauffman, "The logical analysis of continuous, nonlinear biochemical control networks," *Journal of Theoretical Biology*, vol. 39, no. 1, pp. 103–129, 1973.
[2] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, no. 3, pp. 437–467, 1969.
[3] S. A. Kauffman, *Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, NY, USA, 1993.
[4] R. Somogyi and C. Sniegoski, "Modeling the complexity of

genetic networks: understanding multigenic and pleiotropic regulation," *Complexity*, vol. 1, no. 6, pp. 45–63, 1996.
[5] Z. Szallasi and S. Liang, "Modeling the normal and neoplastic cell cycle with "realistic Boolean genetic networks": their application for understanding carcinogenesis and assessing therapeutic strategies," in *Proc. Pacific Symposium on Biocomputing (PSB '98)*, vol. 3, pp. 66–76, Maui, Hawaii, USA, January 1998.
[6] R. Thomas, D. Thieffry, and M. Kaufman, "Dynamical behavior of biological regulatory networks. I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state," *Bulletin of Mathematical Biology*, vol. 57, no. 2, pp. 247–276, 1995.
[7] A. Wuensche, "Genomic regulation modeled as a network with basins of attraction," in *Proc. Pacific Symposium on Biocomputing (PSB '98)*, vol. 3, pp. 89–102, Maui, Hawaii, USA, January 1998.
[8] I. Shmulevich, E. R. 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.
[9] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," *Proceedings of the IEEE*, vol. 90, no. 11, pp. 1778–1792, 2002.
[10] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data," *Journal of Computational Biology*, vol. 7, no. 3/4, pp. 601–620, 2000.
[11] A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, and R. A. Young, "Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks," in *Proc. Pacific Symposium on Biocomputing (PSB '01)*, pp. 422–433, Honolulu, Hawaii, USA, January 2001.
[12] E. J. Moler, D. C. Radisky, and I. S. Mian, "Integrating naive Bayes models and external knowledge to examine copper and iron homeostasis in S. cerevisiae," *Physiological Genomics*, vol. 4, no. 2, pp. 127–135, 2000.
[13] K. Murphy and I. S. Mian, "Modeling gene expression data using dynamic Bayesian networks," Tech. Rep., University of California, Berkeley, Calif, USA, 1999.
[14] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells," *Genetics*, vol. 149, no. 4, pp. 1633–1648, 1998.
[15] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins, "Computational studies of gene regulatory networks: in numero molecular biology," *Nature Reviews Genetics*, vol. 2, no. 4, pp. 268–279, 2001.
[16] P. Smolen, D. Baxter, and J. Byrne, "Mathematical modeling of gene networks," *Neuron*, vol. 26, no. 3, pp. 567–580, 2000.
[17] S. Huang, "Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery," *Journal of Molecular Medicine*, vol. 77, no. 6, pp. 469–480, 1999.
[18] T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano, "Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions," in *Proc. 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '98)*, pp. 695–702, San Francisco, Calif, USA, January 1998.
[19] T. Akutsu, S Miyano, and S. Kuhara, "Inferring qualitative relations in genetic networks and metabolic pathways," *Bioinformatics*, vol. 16, no. 8, pp. 727–734, 2000.
[20] P. D'Haeseleer, S. Liang, and R. Somogyi, "Genetic network inference: from co-expression clustering to reverse engineering," *Bioinformatics*, vol. 16, no. 8, pp. 707–726, 2000.
[21] S. Liang, S. Fuhrman, and R. Somogyi, "REVEAL, a general reverse engineering algorithm for inference of genetic network architectures," in *Proc. Pacific Symposium on Biocomputing (PSB '98)*, vol. 3, pp. 18–29, Maui, Hawaii, USA, January 1998.
[22] X. Zhou, X. Wang, and E. R. Dougherty, "Construction of genomic networks using mutual-information clustering and reversible-jump Markov-Chain-Monte-Carlo predictor design," *Signal Processing*, vol. 83, no. 4, pp. 745–761, 2003.
[23] S. Kim, H. Li, E. R. Dougherty, et al., "Can Markov chain models mimic biological regulation?," *Journal of Biological Systems*, vol. 10, no. 4, pp. 337–357, 2002.
[24] I. Shmulevich, E. R. Dougherty, and W. Zhang, "Gene perturbation and intervention in probabilistic Boolean networks," *Bioinformatics*, vol. 18, no. 10, pp. 1319–1331, 2002.
[25] 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.
[26] A. Datta, A. Choudhary, M. L. Bittner, and E. R. Dougherty, "External control in Markovian genetic regulatory networks," *Machine Learning*, vol. 52, no. 1/2, pp. 169–191, 2003.
[27] E. R. Dougherty and I. Shmulevich, "Mappings between probabilistic Boolean networks," *Signal Processing*, vol. 83, no. 4, pp. 799–809, 2003.
[28] E. R. Dougherty, S. Kim, and Y. Chen, "Coefficient of determination in nonlinear signal processing," *Signal Processing*, vol. 80, no. 10, pp. 2219–2235, 2000.
[29] S. Kim, E. R. Dougherty, M. L. Bittner, et al., "General nonlinear framework for the analysis of gene interaction via multivariate expression arrays," *Journal of Biomedical Optics*, vol. 5, no. 4, pp. 411–424, 2000.

Ivan Ivanov is a Visiting Research Scholar in the Department of Electrical Engineering at Texas A&M University in College Station and a trainee in the training program in bioinformatics at the Department of Statistics at Texas A&M University. He received his M.S. degree in mathematics modeling from Sofia University St. Kilmet Ohridski in 1987 and his Ph.D. degree in mathematics from the University of South Florida in 1999. His current research focuses on genomic signal processing and, in particular, on modeling the genomic regulatory mechanisms. He is working in the Genomic Signal Processing Laboratory at Texas A&M University.



Edward R. Dougherty is a Professor in the Department of Electrical Engineering at Texas A&M University in College Station. He holds an M.S. degree in computer science from Stevens Institute of Technology in 1986 and a Ph.D. degree in mathematics from Rutgers University in 1974. He is the author of eleven books and the editor of other four books. He has published more than one hundred journal papers, is an SPIE Fellow, and has served as an Editor of the Journal of Electronic Imaging for six years. He is currently Chair of the SIAM Activity Group on Imaging Science. Prof. Dougherty has contributed extensively to the statistical design of nonlinear operators for image processing and the consequent application of pattern recognition theory to nonlinear image processing. His current research focuses on genomic signal processing, with the central goal being to model genomic regulatory mechanisms. He is Head of the Genomic Signal Processing Laboratory at Texas A&M University.