East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.030511.060911a

Vol. **2**, No. 1, pp. 1-18
February 2012

On Construction of Sparse Probabilistic Boolean Networks

Xi Chen*, Hao Jiang and Wai-Ki Ching

Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.

Received 3 May 2011; Accepted (in revised version) 6 September 2011

Available online 10 February 2012

Abstract. In this paper we envisage building Probabilistic Boolean Networks (PBNs) from a prescribed stationary distribution. This is an inverse problem of huge size that can be subdivided into two parts — viz. (i) construction of a transition probability matrix from a given stationary distribution (Problem ST), and (ii) construction of a PBN from a given transition probability matrix (Problem TP). A generalized entropy approach has been proposed for Problem ST and a maximum entropy rate approach for Problem TP respectively. Here we propose to improve both methods, by considering a new objective function based on the entropy rate with an additional term of L_α-norm that can help in getting a sparse solution. A sparse solution is useful in identifying the major component Boolean networks (BNs) from the constructed PBN. These major BNs can simplify the identification of the network structure and the design of control policy, and neglecting non-major BNs does not change the dynamics of the constructed PBN to a large extent. Numerical experiments indicate that our new objective function is effective in finding a better sparse solution.

AMS subject classifications: 65C20, 92B05

Key words: Probabilistic Boolean Networks, entropy, stationary distribution, sparsity, transition probability matrix.

1. Introduction

Coordinated interactions and regulations among genes and gene products form so-called gene regulatory networks, an important research topic in genomic research [3, 16] where inference from gene expression data plays an important role. In recent years, many formalisms have been proposed for modeling gene regulatory networks — including Bayesian networks [20], Boolean Networks (BNs) [18], multivariate Markov chain [7] and regression [31] models, and Probabilistic Boolean Networks (PBNs) [23, 24]. The various mathematical models are reviewed in Refs. [15, 25].

*Corresponding author. *Email addresses:* dlkcissy@hku.hk (Xi Chen), haohao@hkusuc.hku.hk (Hao Jiang), wching@hku.hk (Wai-Ki Ching)

The Boolean Network (BN) model and its Probabilistic Boolean Network (PBN) extension have received considerable attention, as they capture some fundamental characteristics of the gene regulations that occur in gene regulatory networks [28]. Consequently, one can understand a particular gene regulatory network and study the influence of different genes. In a BN model, first introduced by Kauffman [18, 19], each gene is represented as a node and each node can take two possible values (1 and 0). The value of a target node is determined by several input nodes (regulators) via a Boolean function. A BN model is deterministic, and randomness only arises from its initial state. Given this inherent deterministic directionality and also the finite number of possible states, the state transitions allow a BN network to enter a set of states and then cycle among them in a fixed order forever, so the set of states is an attractor. If the attractor contains only one state, it is called a singleton attractor; and if it contains more than one state, it is called an attractor cycle [1, 18, 19]. Since attractors represent stable states in a dynamic system, they can reflect the long term behavior of a BN. In particular, it has been demonstrated that attractors are associated with cellular phenotypes [28].

A BN is not only inherently deterministic but also a closed system and therefore has modeling limitations, but a PBN extension provides a stochastic aspect. A PBN consists of a cluster of BNs with selection probabilities assigned, and each BN can be considered a "context". At any given time instant, gene regulations are governed by one of the component BNs. At the next time instant, the system may switch to another BN with a certain switching probability, when the genes can interact under a different context. Thus a PBN model is more flexible than BN model, and it can be described via a Markov chain [8, 23, 24]. Since a PBN also has a finite number of states, its long term behavior can be characterized by the stationary distribution, providing a possible way to infer the PBN from gene expression data.

Time-independent gene expression data can be obtained from micro-array studies, usually by sampling steady states of the network. Using this data, one can estimate a stationary distribution of the network and hence consider building a PBN. This construction problem involves identifying all the component BNs and their corresponding selection probabilities, such that the long term behavior of the constituting PBN is consistent with the prescribed stationary distribution. There has been some preliminary work based on entropy theory [11, 12, 32], using the entropy rate as the objective function. We recall from information theory that the entropy can measure the amount of information missing before reception. Indeed, one can minimize the amount of missing information during the construction of PBNs from gene expression data, using entropy as the objective function. Motivated by the results in [12, 32], we tackle the inverse problem by splitting it into two different inverse problems — viz. (i) construction of a transition probability matrix from a given stationary distribution (Problem ST), and (ii) construction of a PBN from a given transition probability matrix (Problem TP). For the Problem ST, we propose to construct a transition probability matrix from the prescribed stationary distribution. The state transitions in a PBN can be regarded as a Markov chain, and our aim is to find a transition probability matrix that has the prescribed stationary distribution. For Problem TP the main aim is to construct a PBN from a given transition probability matrix by identifying all the

component BNs and their corresponding selection probabilities. The key issue is the large number of feasible solutions, from which we must select a good PBN candidate and the corresponding transition probability matrix, consistent with the given stationary distribution. A reasonable criterion is needed to evaluate all the feasible solutions in order to find an optimal one. We consider the sparsity of the transition probability matrix and the selection probability of component BNs, since a "sparser" solution allows us to identify several major BNs from a huge set of component BNs. Thus in the control policy design for the PBN, we can focus on these major BNs and neglect all others with selection probabilities very close to zero — i.e. we can neglect those that do not change the dynamics of the PBN to a large extent. A sparse solution can help us to better identify the network structure and simplify the design of the control policy. In both problems, we consider adding a term of L_α-norm to the objective function, to more likely get a sparse solution [2,29].

In Section 2, a brief review on BNs and PBNs is given. Section 3 gives a mathematical formulation of the inverse problems, where we present Newton's method in conjunction with the CG method for solving the inverse problems. Section 4 gives some numerical examples to demonstrate our proposed methods. Finally, our concluding remarks are made in Section 5.

2. A Brief Review on Boolean Networks and Probabilistic Boolean Networks

A Boolean Network (BN) $G(V,F)$ consists of a set of nodes

$$V = \{v_1, v_2, \cdots, v_n\}$$

and a list of Boolean functions

$$F = \{f_1, f_2, \cdots, f_n\} \quad \text{where} \quad (f_i : \{0,1\}^n \to \{0,1\}).$$

Here $v_i(t)$ defines the state (0 or 1) of the node v_i at time t, and the Boolean functions represent the rules of regulatory interactions among the nodes — i.e.

$$v_i(t+1) = f_i(\mathbf{v}(t)), \quad i = 1,2,\cdots,n$$

where

$$\mathbf{v}(t) = (v_1(t), v_2(t), \cdots, v_n(t))^T$$

is the Gene Activity Profile (GAP). The GAP can take any possible form (state) from the set

$$S = \{(v_1, v_2, \cdots, v_n)^T : v_i \in \{0,1\}\}, \tag{2.1}$$

and in total there are 2^n possible states.

For example, a BN with three nodes and truth table given in Table 1 has eight states $\{(0,0,0),(0,0,1),(0,1,0),(1,0,0),(1,1,1),(1,1,0),(1,0,1),(0,1,1)\}$. Let us label them by $1,2,3,4,5,6,7$ and 8 respectively. We note that if the current state of the network is 1, the network will go to State 2 in the next step (with probability one); if the current state

Table 1: Truth Table.

State	$v_1(t)$	$v_2(t)$	$v_3(t)$	$f^{(1)}$	$f^{(2)}$	$f^{(3)}$
1	0	0	0	0	0	1
2	0	0	1	0	1	0
3	0	1	0	1	0	0
4	1	0	0	0	0	0
5	1	1	1	1	1	0
6	1	1	0	1	1	1
7	1	0	1	1	0	1
8	0	1	1	1	0	1

is 2, the network will go to State 3 in the next step (with probability one); if the current state is 3, the network will go to State 4 in the next step (with probability one); and if the current state is 4, the network will go to State 1 in next step (with probability one). Thus there is a cycle of length four: $1 \to 2 \to 3 \to 4 \to 1$. One can also check that there is a cycle of length two: $5 \to 6 \to 5$ and a cycle of length one: $7 \to 7$. The transition probability matrix of this 3-gene BN is then given by

$$B = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.2}$$

The truth table provides the one-step transition probability (0 or 1 in the case of the BN) between any two states. Let column vectors **a** and **b** be any two states in the set S. By letting **a** and **b** take all possible states in S, one can get the transition probability matrix of the 3-gene BN. Since the network is a deterministic one, each column in B (the Boolean network matrix) has only one non-zero element and the column sum is one. We remark that there is an one-to-one relation between a BN and its corresponding transition probability matrix.

Since a BN model is deterministic, a natural extension is to a stochastic PBN model. Then instead of only one Boolean function in the case of a BN, in a PBN there are several Boolean functions (predictor functions) $f_j^{(i)}(j = 1, 2, \cdots, l(i))$ to be chosen in determining the state of each node v_i, where $l(i) \leq 2^{2^n}$ is the total number of possible BNs of node v_i available. Since a PBN is a cluster of BNs, there are

$$N = \prod_{i=1}^{n} l(i) \tag{2.3}$$

different possible BNs in the PBN. If $\mathbf{f}_j$ denotes the set of Boolean functions for the jth component BN where

$$\mathbf{f}_j = (f_{j_1}^{(1)}, f_{j_2}^{(2)}, \cdots, f_{j_n}^{(n)}), \quad 1 \le j_i \le l(i), \quad i = 1, 2, \cdots, n\,,$$

then we can get N transition probability matrices $A_1, \cdots, A_N$ of component BNs from the N sets of Boolean functions $\mathbf{f}_1, \cdots, \mathbf{f}_N$.

In an independent PBN (i.e. where the selection of the Boolean function for each node is assumed to be independent), the selection probability for the jth component BN is

$$q_j = \prod_{i=1}^{n} c_{j_i}^{(i)}, \quad j = 1, 2, \cdots, N \tag{2.4}$$

where $c_{j_i}^{(i)}$ is the probability of choosing $f_{j_i}^{(i)}$ as the Boolean function for node v_i. The probability $c_{j_i}^{(i)}$ can be estimated from gene expression data by using a statistical method — viz. the coefficient of determination method [14]. The state transitions in a PBN follow a Markov chain process [8]. If $\mathbf{a}$ and $\mathbf{b}$ denote any two states in the set S, the transition probability is given by

$$\begin{aligned} &\text{Prob}\,\{\mathbf{v}(t+1) = \mathbf{a} \mid \mathbf{v}(t) = \mathbf{b}\} \\ &= \sum_{j=1}^{N} \text{Prob}\,\{\mathbf{v}(t+1) = \mathbf{a} \mid \mathbf{v}(t) = \mathbf{b}, \text{the } j\text{th BN is selected}\,\} \cdot q_j. \end{aligned} \tag{2.5}$$

In fact, it can be shown that the transition probability matrix A of PBN can be written as the sum of the transition probability matrices A_i of component BNs ([9]):

$$A = \sum_{i=1}^{N} q_i A_i. \tag{2.6}$$

3. The Inverse Problems

We now discuss the mathematical formulations of the two problems ST and TP. We first construct a sparse transition probability matrix from the prescribed stationary distribution (Problem ST), and then identify the major component BNs and corresponding selection probabilities (Problem TP).

3.1. New mathematical formulation for Problem ST

Given a stationary distribution π with 2^n states, we construct a corresponding $2^n \times 2^n$ transition probability matrix P such that

$$P\pi = \pi \quad \text{and} \quad (1, 1, \cdots, 1)P = (1, 1, \cdots, 1). \tag{3.1}$$

There can be infinitely many possible solutions for the captured problem, so some measure has be introduced to find an optimal solution. In [11], the generalized entropy rate was proposed as a measure — i.e.

$$\sum_{j=1}^{2^n} w_j \left(-\sum_{i=1}^{2^n} P_{ij} \log P_{ij} \right) \tag{3.2}$$

where

$$0 \leq w_j \leq 1 \quad \text{and} \quad \sum_{i=1}^{2^n} w_i = 1.$$

The parameter w_j represents the weighting (importance) of State j and

$$-\sum_{i=1}^{2^n} P_{ij} \log P_{ij}$$

is the entropy of the conditional probability distribution when the chain is in State j. Thus here the proposal is that the entropy measure the amount of missing information in the gene expression data.

To construct a sparse transition probability matrix from the stationary distribution $\boldsymbol{\pi}$, we consider modifying the objective function by adding a term of L_α-norm. In [2], there is an algorithm for reconstructing a sparse solution $\mathbf{x} = (x_1, \cdots, x_n)$ from a small number of constraints by solving a linear system. On adding the L_1-norm of $\mathbf{x}$ defined by

$$\sum_{i=1}^{n} |x_i|$$

to the objective function, a sparse solution is more likely. We would like to introduce the L_1-norm in (3.2), but in our problem there is the constraint

$$\sum_{i=1}^{2^n} P_{ij} = 1 \quad j = 1, \cdots, 2^n$$

such that the L_1-norm actually has no effect. In view of this, we modify the idea by considering the following term for some $\alpha \in (0,1)$:

$$\sum_{j=1}^{2^n} \left(\beta \sum_{i=1}^{2^n} P_{ij}^{\alpha} \right), \tag{3.3}$$

where β is a non-negative weighting to be assigned. The new optimization problem then becomes

$$\max_{P_{ij}} \left\{ \sum_{j=1}^{2^n} \boldsymbol{\pi}_j \left(-\sum_{i=1}^{2^n} P_{ij} \log P_{ij} \right) - \sum_{j=1}^{2^n} \left(\beta \sum_{i=1}^{2^n} P_{ij}^{\alpha} \right) \right\} \tag{3.4}$$

subject to

$$\begin{cases} \sum_{i=1}^{2^n} P_{ij} = 1, \quad j = 1, 2, \cdots, 2^n \\ P\boldsymbol{\pi} = \boldsymbol{\pi} \\ P_{ij} \geq 0, \qquad i, j = 1, 2, \cdots, 2^n. \end{cases} \tag{3.5}$$

The parameter π_j represents the weighting of state j, the first term in the objective function is the entropy rate of the Markov chain, and the second term

$$\beta \sum_{i=1}^{N} P_{ij}^{\alpha}$$

is employed to obtained a sparse solution. Here α and β are two parameters, and by varying them we can adjust the sparsity of the solution. In our experiments, we set the range of α to be $[0.01, 0.99]$ and the range of β as $[0.1, 2.0]$.

To evaluate the performance of the new method and to obtain the best α and β pair, we employ both the entropy rate

$$\sum_{j=1}^{2^n} \pi_j \left(-\sum_{i=1}^{2^n} P_{ij} \log P_{ij} \right)$$

and the weighted variance of P

$$\sum_{j=1}^{2^n} \pi_j Var\left((P_{1j}, \cdots, P_{2^n j}) \right)$$

as two measures of the solutions. Here $Var(\mathbf{p})$ is the variance of the probability distribution vector $\mathbf{p}$. Experiments indicate that the two measures give consistent results.

3.2. New mathematical formulation for Problem TP

In Problem TP , given a transition probability matrix P we seek to identify the major component BNs and the corresponding selection probabilities, and then construct a PBN using these BNs. The transition probability matrix of the constituting PBN can then be approximated by the given transition probability matrix P. Suppose matrix P has at most m non-zero entries in each column. Although there can be infinite many PBNs with the transition probability matrix P, they share the same set of possible component BNs. It is easy to find there are at most m^{2^n} possible component BNs to form a PBN with the given transition probability matrix P. For example, a transition probability matrix is ([4])

$$P_1 = \begin{pmatrix} 0.1 & 0.3 & 0.5 & 0.6 \\ 0.0 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 \\ 0.9 & 0.0 & 0.0 & 0.4\,, \end{pmatrix}. \tag{3.6}$$

where the maximum number of non-zero entries in each column is 2, Then there are at most 16 possible component BNs for a PBN with the above transition probability matrix — i.e.

$$A_1=\begin{pmatrix}1&1&1&1\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}A_2=\begin{pmatrix}1&1&1&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{pmatrix}A_3=\begin{pmatrix}1&1&0&1\\0&0&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}A_4=\begin{pmatrix}1&1&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

$$A_5=\begin{pmatrix}1&0&1&1\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}A_6=\begin{pmatrix}1&0&1&0\\0&1&0&0\\0&0&0&0\\0&0&0&1\end{pmatrix}A_7=\begin{pmatrix}1&0&0&1\\0&1&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}A_8=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

$$A_9=\begin{pmatrix}0&1&1&1\\0&0&0&0\\0&0&0&0\\1&0&0&0\end{pmatrix}A_{10}=\begin{pmatrix}0&1&1&0\\0&0&0&0\\0&0&0&0\\1&0&0&1\end{pmatrix}A_{11}=\begin{pmatrix}0&1&0&1\\0&0&0&0\\0&0&1&0\\1&0&0&0\end{pmatrix}A_{12}=\begin{pmatrix}0&1&0&0\\0&0&0&0\\0&0&1&0\\1&0&0&1\end{pmatrix}$$

$$A_{13}=\begin{pmatrix}0&0&1&1\\0&1&0&0\\0&0&0&0\\1&0&0&0\end{pmatrix}A_{14}=\begin{pmatrix}0&0&1&0\\0&1&0&0\\0&0&0&0\\1&0&0&1\end{pmatrix}A_{15}=\begin{pmatrix}0&0&0&1\\0&1&0&0\\0&0&1&0\\1&0&0&0\end{pmatrix}A_{16}=\begin{pmatrix}0&0&0&0\\0&1&0&0\\0&0&1&0\\1&0&0&1\end{pmatrix}.$$

To construct a PBN from P_1 in (3.6), we may consider equation (2.6) such that

$$P_1=\sum_{i=1}^{16}q_iA_i.$$

If we can find a feasible solution of $\mathbf{q}=(q_1,\cdots,q_{16})$, then we can construct a PBN with the transition probability matrix P.

Given any transition probability matrix P, from the above example it is evidently possible to find a set of possible component BNs with the transition probability matrices $A_1,\cdots,A_{m^{2^n}}$, so Problem TP is then reduced. Thus given the transition probability matrix P and transition probability matrices $A_1,\cdots,A_{m^{2^n}}$ of the component BNs, we seek the selection probabilities $(q_1,\cdots,q_{16})$ satisfying

$$P=\sum_{i=1}^{m^{2^n}}q_iA_i \tag{3.7}$$

and

$$0\le q_i\le 1\quad\text{and}\quad\sum_{i=1}^{m^{2^n}}q_i=1.$$

Similar to Problem ST, there can be infinitely many solutions, so we need some measure to select an optimal solution. In [32], the entropy rate is employed as a measure of solution, similar to Problem ST. Here we take the sparsity of $\mathbf{q}$ into consideration, to identify several

major component BNs and neglect many unimportant BNs for a sparse solution. These major BNs can simplify the identification of the network structure and design of control policy, and neglecting the non-major BNs will not change the dynamics of the PBN to a large extent. To find a sparse solution, we consider adding the following L_α norm to the objective function:

$$\beta \sum_{i=1}^{m^{2^n}} q_i^\alpha. \tag{3.8}$$

Then the objective function becomes

$$\max_{\mathbf{q}} \left\{ -\sum_{i=1}^{m^{2^n}} q_i \log q_i - \beta \sum_{i=1}^{m^{2^n}} q_i^\alpha \right\}, \tag{3.9}$$

which is subject to

$$P = \sum_{i=1}^{m^{2^n}} q_i A_i, \tag{3.10}$$

$$0 \le q_i \le 1 \quad \text{and} \quad \sum_{i=1}^{m^{2^n}} q_i = 1. \tag{3.11}$$

Similar to Problem ST, the first term in the objective function is the entropy rate of $\mathbf{q}$, and the second term

$$\beta \sum_{i=1}^{m^{2^n}} q_i^\alpha$$

is employed to obtained a sparse solution. Here α and β are two parameters, and by varying them we can adjust the sparsity of the solution. In our experiments, we set the range of α to be $[0.01, 0.99]$ and the range of β as $[0.1, 2.0]$.

To evaluate the performance of the new method and obtain the best pair of α and β, we employ both the entropy rate

$$-\sum_{i=1}^{2^n} q_i \log q_i$$

and the weighted variance of $\mathbf{q}$, as two measures of the solutions. Experiments indicate that the two measures give consistent results.

3.3. The modified Newton's method

We now discuss how to solve the optimization problems (3.4) - (3.5) and (3.9) - (3.11). The solution of these two optimization problems is equivalent to solving the following optimization problem:

$$\max_{\mathbf{x}} \left\{ -x_i \sum_{i=1}^{N} \log x_i - \beta \sum_{i=1}^{N} x_i^\alpha \right\} \tag{3.12}$$

subject to

$$\mathbf{Cx} = \mathbf{d}, \tag{3.13}$$

$$x_i \geq 0, \quad i = 1, 2, \cdots, N. \tag{3.14}$$

where $\mathbf{x} = (x_1, \cdots, x_N)^T$, $\mathbf{C}$ is a l-by-N matrix and $\mathbf{d}$ is an N-by-1 vector. We follow analysis similar to that in Ref. [4], and apply the Lagrange multiplier method to this equivalent optimization problem. The Lagrange function only involves the constraint $\mathbf{Cz} = \mathbf{d}$, but the constraint $x_i \geq 0$ is checked in the whole process. Thus if $\mathbf{y}$ denotes the multiplier and $L(.,.)$ the Lagrangian function, we have

$$L(\mathbf{x}, y) = \max_{\mathbf{x}} \left\{ -\sum_{i=1}^{N} x_i \log x_i - \beta \sum_{i=1}^{N} x_i^{\alpha} + \mathbf{y}^T(\mathbf{d} - \mathbf{Cx}) \right\}, \tag{3.15}$$

so the optimal solution to the equivalent problem follows by solving

$$\begin{aligned} \nabla_{q_i} L(\mathbf{x}, \mathbf{y}) =& -\log x_i - 1 - \alpha\beta x_i^{\alpha-1} - \mathbf{y}^T \mathbf{C}_{.i} \\ =& 0, \quad i = 1, \cdots, N \end{aligned} \tag{3.16}$$

and

$$\nabla_{y_j} L(\mathbf{x}, \mathbf{y}) = d_j - \mathbf{C}_{j.}\mathbf{x} = 0, \quad j = 1, \cdots, l, \tag{3.17}$$

where $\mathbf{C}_{.i}$ is the ith column of matrix $\mathbf{C}$ and $\mathbf{C}_{j.}$ is the jth row of matrix $\mathbf{C}$.

Many numerical methods are available to solve Eqs. (3.16) and (3.17), but here we choose to use our modified Newton's method in conjunction with the CG method [11]. Let $\mathbf{z} = (x_1, \cdots, x_N, y_1, \cdots, y_l)^T$, and consider

$$\mathbf{F}(\mathbf{z}) = (F_1(\mathbf{z}), F_2(\mathbf{z}), \cdots, F_N(\mathbf{z}), F_{N+1}(\mathbf{z}), \cdots, F_{N+l}(\mathbf{z}))^T$$

where

$$F_i(\mathbf{z}) = -\log x_i - 1 - \alpha\beta x_i^{\alpha-1} - \mathbf{y}^T \mathbf{C}_{.i}, \quad i = 1, 2, \cdots, N, \tag{3.18}$$

$$F_{N+j}(\mathbf{z}) = d_j - \mathbf{C}_{j.}\mathbf{x}, \quad j = 1, 2, \cdots, l. \tag{3.19}$$

Let $M(\mathbf{z})$ be the diagonal matrix with diagonal entries

$$-\frac{1}{x_i} - \alpha\beta(\alpha - 1)x_i^{\alpha-2}, \quad i = 1, \cdots, N.$$

Then we have the following modified Newton's method for solving Eqs. (3.16) and (3.17).
Modified Newton's Method [11]:
Choose starting point $\mathbf{z}_0 \in \mathbf{v}^{\perp}$
$k = 1$;
while $||F(\mathbf{z}_k) - F(\mathbf{z}_{k-1})||_2 > tolerance$

find $\bar{\mathbf{b}}_k \in \bar{\mathbf{v}}^{\perp}$ with

$$\begin{pmatrix} M(\mathbf{z}_{k-1}) & \mathbf{0} \\ \mathbf{0} & -\mathbf{C}M(\mathbf{z}_{k-1})^{-1}\mathbf{C}^T \end{pmatrix} \bar{\mathbf{b}}_k = -\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}M(\mathbf{z}_{k-1})^{-1} & \mathbf{I} \end{pmatrix} F(\mathbf{z}_{k-1}); \tag{3.20}$$

set

$$\mathbf{b}_k = \begin{pmatrix} \mathbf{I} & M(\mathbf{z}_{k-1})^{-1}\mathbf{C}^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \bar{\mathbf{b}}_k;$$

$\mathbf{z}_k = \mathbf{z}_{k-1} + \mathbf{b}_k$;
$k = k+1$;
end.

To ensure the first N entries in $\mathbf{z}_k$ (i.e. $x_1, x_2, \cdots, x_N$) are positive, we set a small positive lower bound $r > 0$ for all x_i, and this guarantees the transition probability matrix is aperiodic and irreducible. Henceforth, in the Newton's method, whenever $x_i < r$ ($i = 1, \cdots, N$) occurs in the iteration, we set $x_i = r$ in all further iterations. In this paper, we set $r = 10^{-5}$.

A possible starting point is the vector

$$\mathbf{z}_0 = \left(\underbrace{\frac{1}{N}, \frac{1}{N}, \cdots, \frac{1}{N}}_{N}, \underbrace{d_1, d_2, \cdots, d_l}_{l} \right)^T. \tag{3.21}$$

Furthermore, we can consider the problem of finding $\bar{\mathbf{b}}_k \in \bar{\mathbf{v}}^{\perp}$ satisfying (3.20) as two subproblems. We denote $\bar{\mathbf{b}} = (\mathbf{u}\ ,\ \mathbf{w})^T$, where $\mathbf{u}$ and $\mathbf{w}$ are $N \times 1$ and $l \times 1$ vectors, respectively. Correspondingly, we denote the right-hand-side of Equation (3.20) as $\left(\mathbf{e}_1\ ,\ \mathbf{e}_2\right)$. Thus Eq. (3.20) can be reconsidered as

$$M\mathbf{u} = \mathbf{e_1} \tag{3.22}$$

and

$$-\mathbf{C}M^{-1}\mathbf{C}^T\mathbf{w} = \mathbf{e_2}. \tag{3.23}$$

Here the positive definiteness of $-\mathbf{C}M^{-1}\mathbf{C}^T$ on $\mathbf{v}^{\perp}$ suggests that one possible choice for solving Eq. (3.23) is the CG method.

4. Numerical Experiments

In this section, we discuss some numerical examples to demonstrate our new approach for Problem TP and Problem ST, and compare the results with existing methods. At the same time, the process for choosing the parameters α and β is also considered.

4.1. Numerical examples for Problem TP

We first discuss the performance of our method in solving Problem TP (less complicated than for Problem ST), in comparison with results obtained from the method in Ref. [4].

Example 4.1. Consider the transition probability matrix [4] in Section 3.2 — i.e.

$$P_1 = \begin{pmatrix} 0.1 & 0.3 & 0.5 & 0.6 \\ 0.0 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 \\ 0.9 & 0.0 & 0.0 & 0.4 \end{pmatrix}.$$

As previously discussed, there are at most 16 possible BNs for a PBN with this transition probability matrix, and we now seek the following decomposition:

$$P_1 = \sum_{i=1}^{16} q_i A_i \quad \text{where} \quad \sum_{i=1}^{16} q_i = 1 \quad \text{and} \quad q_i \geq 0, \tag{4.1}$$

where $A_1, \cdots, A_{16}$ are given in Section 3.2.

Using our method, we obtain the solution shown in Fig. 1 (Right). Fig. 2 shows the distribution of entropies and variances for different pairs of

$$(\alpha, \beta) \in [0.01, 0.99] \times [0.10, 2.00].$$

Each of the two figures consists of 99×20 points. The higher (lower) the variances (entropies) are, the lighter the points are. Among these 99×20 points, the largest variance (entropy) is 9.9×10^{-3} (2.3231), while the smallest variance (entropy) is 2.4×10^{-3} (1.5112), so different pairs of (α, β) can bring significant changes to the variance (entropy) of $\mathbf{q}$. The "good" pairs of (α, β) lie in a parabolic shaped region; and α and β work in conjunction with each other to influence the variance (entropy) of $\mathbf{q}$. The optimal solution is reached when $\alpha = 0.63$ and $\beta = 1.40$, for both measurements. We note that the re-constructed PBN is dominated (over 99.7%) by the 6th, 8th, 10th, 12th, 13th and 15th BNs. From the dominated BNs, one can therefore construct the underlying regulatory rules — i.e. their truth tables. Here we see that our method can be used to identify the major components of the BNs constituting the PBN better than the method in [4] — cf. Fig. 1 (Left).

The stationary distribution of P_1 is $(0.4, 0.00, 0.00, 0.60)^T$. If we approximate P_1 by using the six major BNs A_6,A_8,A_{10},A_{12},A_{13} and A_{15} with a normalization — i.e. we approximate P_1 by

$$\tilde{P}_1 = \frac{1}{0.997}(0.049A_6 + 0.049A_8 + 0.1495A_{10} + 0.1495A_{12} + 0.30A_{13} + 0.30A_{15})$$

— then the stationary distribution of $\tilde{P}_1$ is $(0.4001, 0.0000, 0.0000, 0.5999)^T$ with initial probability distribution $(0.25, 0.25, 0.25, 0.25)^T$. This shows our method produces a good

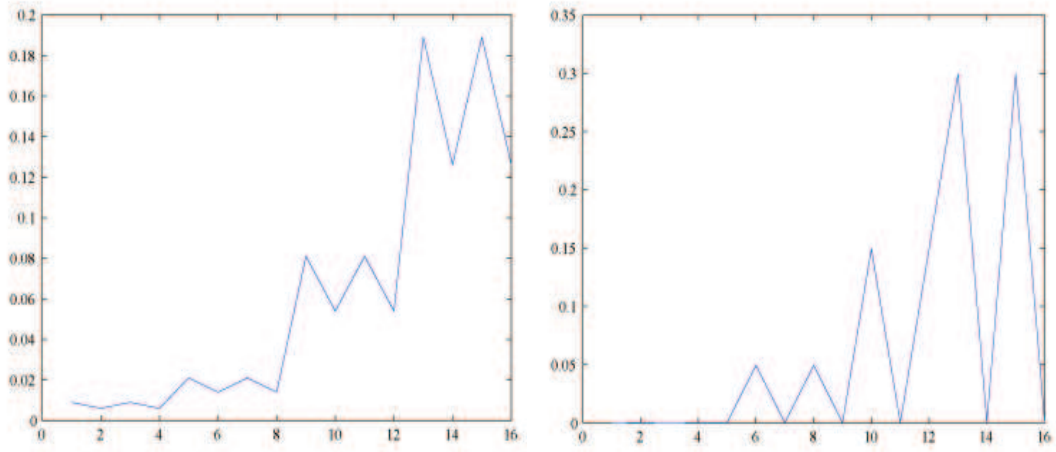

Figure 1: The Probability Distribution $\mathbf{q}$ for the Case of P_1. Method in [4] (Left) and Our method (Right)

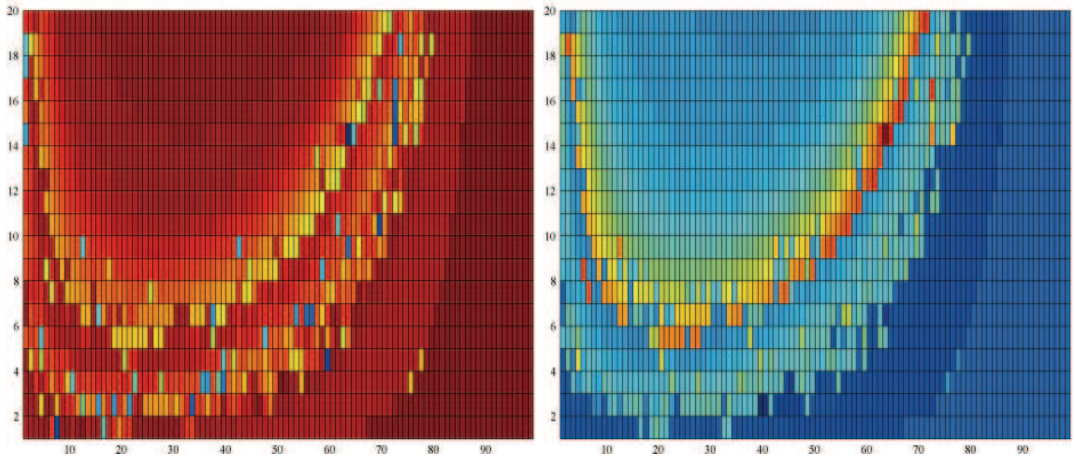

Figure 2: The Distribution of Entropies (Left) and the Distribution of Variances (Right)

Table 2: The stationary distributions with different approximations.

No. of BNs dropped	$\|\|\tilde{\mathbf{p}}-\mathbf{p}\|\|_2^2$	$\tilde{\mathbf{p}}$	$\|\|\hat{\mathbf{p}}-\mathbf{p}\|\|_2^2$	$\hat{\mathbf{p}}$
0	7.96×10^{-10}	$(0.4000, 0.0000, 0.0000, 0.6000)^T$	7.96×10^{-10}	$(0.4000, 0.0000, 0.0000, 0.6000)^T$
2	1.20×10^{-4}	$(0.3999, 0.0000, 0.0000, 0.6001)^T$	4.25×10^{-10}	$(0.4000, 0.0000, 0.0000, 0.6000)^T$
4	1.67×10^{-4}	$(0.3999, 0.0000, 0.0000, 0.6001)^T$	0.01	$(0.3927, 0.0000, 0.0000, 0.6073)^T$
6	3.57×10^{-4}	$(0.3997, 0.0000, 0.0000, 0.6003)^T$	0.01	$(0.3927, 0.0000, 0.0000, 0.6073)^T$
8	3.57×10^{-4}	$(0.3997, 0.0000, 0.0000, 0.6003)^T$	0.04	$(0.3750, 0.0000, 0.0000, 0.6250)^T$
10	1.87×10^{-4}	$(0.4001, 0.0000, 0.0000, 0.5999)^T$	0.01	$(0.4054, 0.0000, 0.0000, 0.5946)^T$
12	1.87×10^{-4}	$(0.4001, 0.0000, 0.0000, 0.5999)^T$	1.2329	$(0.0000, 1.0000, 0.0000, 0.0000)^T$

approximation of the PBN. In Table 2, we give the approximates $\tilde{\mathbf{p}}$ (our method) and $\hat{\mathbf{p}}$ (method in [4]) of the stationary distributions $\mathbf{p}$ obtained when we drop a number of BNs progressively, from the smallest selection probability to the highest selection probability.

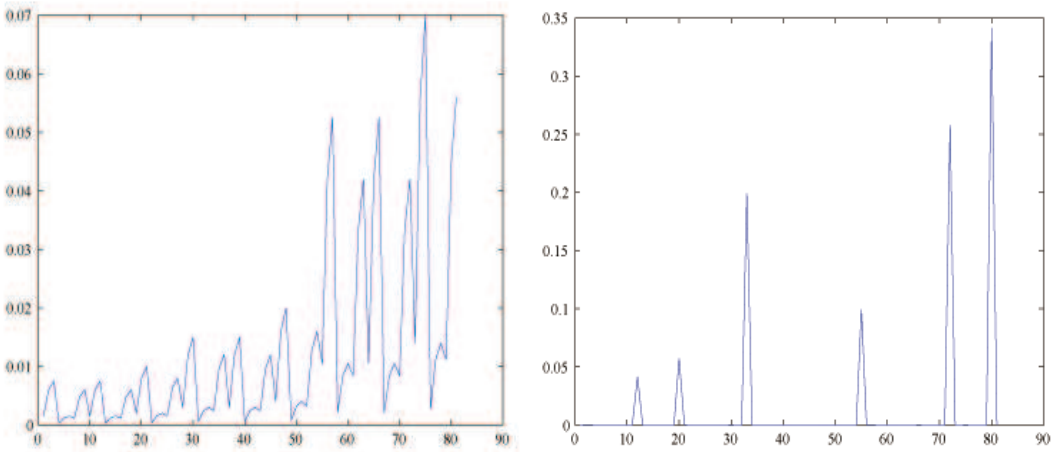

Figure 3: The Probability Distribution $\mathbf{q}$ for the Case of P_2. Method in [4] (Left) and Our method (Right)

We also compare the errors

$$||\tilde{\mathbf{p}} - \mathbf{p}||_2^2 \quad \text{and} \quad ||\hat{\mathbf{p}} - \mathbf{p}||_2^2$$

of the two methods. It is clear that our method is better, and that the solution (Fig. 1, right) obtained by our proposed method is "sparser" than the solution (Fig. 1, left) obtained by the method in Ref. [4].

Example 4.2. Here we consider two genes ($n = 2$) and a transition probability matrix with three non-zero entries ($m = 3$), and the observed transition probability matrix of the PBN

$$P_2 = \begin{pmatrix} 0.1 & 0.3 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.6 & 0.4 \\ 0.7 & 0.4 & 0.0 & 0.5 \end{pmatrix}. \tag{4.2}$$

Using our modified entropy approach, we obtain the solution as shown in Fig. 3 (Right). The optimal solution is reached when $\alpha = 0.61$ and $\beta = 0.6$. Compared with results in Ref. [4] (Fig. 3, left), our solution is much more sparse and we can more readily identify the major BNs in the PBN.

4.2. Numerical examples for Problem ST

We now demonstrate the performance of our method in solving Problem ST with some numerical examples, in comparison to results obtained from the method proposed in Ref [11].

Example 4.3. Given the stationary distribution [11]

$$\pi = (0.1, 0.2, 0.3, 0.4)$$

of a Markov chain with four states, we want to construct a sparse transition probability matrix corresponding to it. In Ref. [11], the optimal solution obtained is

$$P_3 = \begin{pmatrix} 0.1860 & 0.1344 & 0.0947 & 0.0653 \\ 0.2390 & 0.2220 & 0.2010 & 0.1784 \\ 0.2741 & 0.2918 & 0.3031 & 0.3083 \\ 0.3009 & 0.3518 & 0.4012 & 0.4480 \end{pmatrix}.$$

Using our method, we obtain the following "optimal" transition probability matrix:

$$P_4 = \begin{pmatrix} 0.0335 & 0.0358 & 0.0337 & 0.1984 \\ 0.0313 & 0.0312 & 0.0278 & 0.4557 \\ 0.9024 & 0.8990 & 0.0284 & 0.0536 \\ 0.0327 & 0.0341 & 0.9100 & 0.2923 \end{pmatrix}.$$

Clearly P_4 is more sparse than P_3.

The optimal solution is reached when $\alpha = 0.94$ and $\beta = 1.6$. In this numerical experiment, we employ a grid search method to adjust the values of α and β to reach the optimal solution. Here α ranges from 0.01 to 0.99 with grid size 0.01, and β ranges from 0.1 to 2.0 with grid size 0.1 — i.e. we tried 1980 pairs of values for α and β, and chose the pair that produces the maximum objective value in (3.4). We find there are fewer non-zero entries in our solution, so we obtain a sparser solution.

Example 4.4. Let us consider a 3-gene example, with the randomly generated stationary distribution

$$\pi = (0.1282, 0.2139, 0.0667, 0.1766, 0.1758, 0.0887, 0.1324, 0.0177).$$

Using our our method, we obtain the following transition probability matrix for our optimal solution:

$$P_5 = \begin{pmatrix} 0.0046 & 0.5758 & 0.0050 & 0.0085 & 0.0051 & 0.0049 & 0.0084 & 0.0094 \\ 0.0039 & 0.0044 & 0.0045 & 0.7506 & 0.0039 & 0.0043 & 0.5920 & 0.0089 \\ 0.0048 & 0.0068 & 0.0051 & 0.1948 & 0.0054 & 0.0050 & 0.2140 & 0.0095 \\ 0.0051 & 0.0081 & 0.9677 & 0.0120 & 0.0060 & 0.9685 & 0.1542 & 0.0097 \\ 0.0048 & 0.0067 & 0.0051 & 0.0093 & 0.9668 & 0.0050 & 0.0090 & 0.0095 \\ 0.0049 & 0.3884 & 0.0052 & 0.0101 & 0.0055 & 0.0051 & 0.0095 & 0.0095 \\ 0.9707 & 0.0089 & 0.0054 & 0.0136 & 0.0063 & 0.0054 & 0.0115 & 0.0098 \\ 0.0012 & 0.0009 & 0.0020 & 0.0011 & 0.0010 & 0.0017 & 0.0014 & 0.9338 \end{pmatrix}.$$

The result obtained by the method in Ref. [11] is

$$P_6 = \begin{pmatrix} 0.4777 & 0.0715 & 0.0878 & 0.0742 & 0.0743 & 0.0839 & 0.0783 & 0.1050 \\ 0.1193 & 0.5626 & 0.1317 & 0.1130 & 0.1131 & 0.1263 & 0.1186 & 0.1547 \\ 0.0456 & 0.0411 & 0.3825 & 0.0428 & 0.0428 & 0.0489 & 0.0454 & 0.0626 \\ 0.1022 & 0.0933 & 0.1133 & 0.5296 & 0.0967 & 0.1085 & 0.1017 & 0.1341 \\ 0.1018 & 0.0929 & 0.1129 & 0.0963 & 0.5288 & 0.1081 & 0.1013 & 0.1336 \\ 0.0580 & 0.0524 & 0.0651 & 0.0545 & 0.0545 & 0.4225 & 0.0577 & 0.0787 \\ 0.0808 & 0.0734 & 0.0901 & 0.0762 & 0.0763 & 0.0861 & 0.4828 & 0.1076 \\ 0.0145 & 0.0129 & 0.0165 & 0.0135 & 0.0135 & 0.0156 & 0.0144 & 0.2236 \end{pmatrix}.$$

Clearly, P_5 is much sparser than P_6.

5. Concluding Remarks

In this paper, we have presented two modified entropy methods for (1) constructing a sparse transition probability matrix from a given stationary distribution, and (2) for constructing a PBN from a given sparse transition probability matrix. Both are inverse problems of large size, which we solved by Newton's method in conjunction with the CG method. The entropy rate and weighted variance were both employed to measure the sparsity of the solution obtained, and give consistent results in our numerical experiments. We also found encouraging sparsity for some small size networks, compared with results from existing methods. In future, we intend to apply our approach to more practical examples.

Acknowledgments

A preliminary version of this paper was presented at the ISB2010 conference and published in the corresponding conference proceedings [5]. Our research was supported in part by HKRGC Grant 7017/07P, the HKU strategic theme grant on computational sciences, National Natural Science Foundation of China Grant No. 10971075 and the Guangdong Provincial Natural Science Grant No. 9151063101000021.

References

[1] T. Akutsu, M. Hayasida, W, Ching and M. Ng, *Control of Boolean Networks: Hardness Results and Algorithms for Tree Structured Networks*, Journal of Theoretical Biology, 2007, 244: 670–679.

[2] E. J. Candes and T. Tao, *Near-optimal Signal Recovery from Random Projections: Universal Encoding Strategies*, IEEE Transactions on Information Theory, 2006, 52: 5406–5425.

[3] J. E. Celis, M. Kruhøfferm, I. Gromova, C. Frederiksen, M. Østergaard and T. F. Ørntoft, *Gene Expression Profiling: Monitoring Transcription and Translation Products Using DNA Microarrays and Proteomics*, FEBS Letters, 2000, 480(1): 2–16.

[4] X. CHEN, W. CHING, X. CHEN, Y. CONG AND N. TSING, *Construction of Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix: A Maximum Entropy Rate Approach*, East Asian Journal of Applied Mathematics, 1 (2011) 132–154.

[5] X. CHEN, L. LI, W. CHING AND N. TSING, *A Modified Entropy Approach for Construction of Probabilistic Boolean Networks*, Lecture Notes in Operations Research 13, Series Editors: Ding-Zhu Du and Xiang-Sun Zhang The Fourth International Conference on Computational Systems Biology (ISB2010) Suzhou, China, September 9-11, 2010, 13: 243–250. Available at http://www.aporc.org/LNOR/13/ISB2010F32.pdf

[6] W. CHING, S. SCHOLTES AND S. ZHANG, *Numerical Algorithms for Estimating Traffic Between Zones in a Network*, Engineering Optimisation, 36 (2004), 379–400.

[7] W. CHING, E. FUNG, M. NG AND T. AKUTSU, *On Construction of Stochastic Genetic Networks Based on Gene Expression Sequences*, International Journal of Neural Systems, 15 (2005), 297–310.

[8] W. CHING AND M. NG, *Markov Chains: Models, Algorithms and Applications*, International Series on Operations Research and Management Science, Springer: New York, 2006.

[9] W. CHING, S. ZHANG, M. NG AND T. AKUTSU, *An Approximation Method for Solving the Steady-state Probability Distribution of Probabilistic Boolean Networks*, Bioinformatics, 2007, 23: 1511–1518.

[10] W. CHING, S. ZHANG, Y. JIAO, T. AKUTSU AND A. WONG, *Optimal Control Policy for Probabilistic Boolean Networks with Hard Constraints*, IET on Systems Biology, 2009, 3: 90–99.

[11] W. CHING AND Y. CONG, *A New Optimization Model for the Construction of Markov Chains*, CSO2009, Hainan, IEEE Computer Society Proceedings, 2009, 551–555.

[12] W. CHING, X. CHEN AND N. TSING, *Generating Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix*, IET on Systems Biology, 2009, 6: 453–464.

[13] A. CONN, B. GOULD AND P. TOINT, *Trust-region Methods*, SIAM Publications, Philadelphia, 2000.

[14] E. DOUGHERTY, S. KIM AND Y. CHEN, *Coefficient of Determination in Nonlinear Signal Processing*, Signal Processing, 2000, 80: 2219–2235.

[15] H. DE JONG, *Modeling and Simulation of Genetic Regulatory Systems: A Literature Review*, Journal of Computational Biology, 2002, 9: 69–103.

[16] S. HUANG, *Gene Expression Profiling, Genetic Networks, and Cellular States: An Integrating Concept for Tumorigenesis and Drug Discovery*, Journal of Molecular Medicine, 1999, 77: 469–480.

[17] S. HUANG AND D. E. INGBER, *Shape-dependent Control of Cell Growth, Differentiation, and Apoptosis: Switching Between Attractors in Cell Regulatory Networks*, Experimental Cell Research, 2000, 261: 91–103.

[18] S. KAUFFMAN, *Metabolic Stability and Epigenesis in Randomly Constructed Gene Nets*, Journal of Theoretical Biology, 1969, 22: 437–467.

[19] S. KAUFFMAN, *The Origins of Order: Self-organization and Selection in Evolution*, New York: Oxford Univ. Press, 1993.

[20] S. KIM, S. IMOTO AND S. MIYANO, *Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling of Gene Networks from time Series Gene Expression Data*, Proc. 1st Computational Methods in Systems Biology, Lecture Note in Computer Science, 2003, 2602: 104–113.

[21] R. PAL, I. IVANOV, A. DATTA, M. BITTNER AND E. DOUGHERTY, *Generating Boolean Networks with a Prescribed Attractor Structure*, Bioinformatics, 2005, 21: 4021–4025.

[22] C.E. SHANNON, *A Mathematical Theory of Communication*, Bell System Technical Journal, 1948, 27: 379–423.

[23] I. SHMULEVICH, E. DOUGHERTY, S. KIM AND W. ZHANG, *Probabilistic Boolean Networks: A Rule-based Uncertainty Model for Gene Regulatory Networks*, Bioinformatics, 2002, 18: 261–274.

[24] I. Shmulevich and E. Dougherty, *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*, SIAM Press, 2009.
[25] P. Smolen, D. Baxter and J. Byrne, *Mathematical Modeling of Gene Network*, Neuron, 2000, 26: 567–580.
[26] R. Somogyi and C. Sniegoski, *Modeling the Complexity of Gene Networks: Understanding Multigenic and Pleiotropic Regulation*, Complexity, 1996, 1: 45–63.
[27] A. Wilson, *Entropy in Urban and Regional Modelling*, Pion, London, 1970.
[28] Y. Xiao and E. R. Dougherty, *The Impact of Function Perturbations in Boolean Networks*, Bioinformatics, 2007, 23(10): 1265–1273.
[29] Z. Xu, H. Zhang, Y. Wang and X. Chang, $L_{1/2}$ *Regularization*, Science in China Series F-Information Sciences, (2010), 40: 1–11.
[30] S. Zhang, W. Ching, M. Ng and T. Akutsu, *Simulation Study in Probabilistic Boolean Network Models for Genetic Regulatory Networks*, Journal of Data Mining and Bioinformatics, 2007, 1: 217–240.
[31] S. Zhang, W. Ching, N. Tsing, H. Leung and D. Guo, *A New Multiple Regression Approach for the Construction of Genetic Regulatory Networks*, Journal of Artificial Intelligence in Medicine, 2009, 48: 153–160.
[32] S. Zhang, W. Ching, X. Chen and N. Tsing, *On Construction of PBNs from a Prescribed Stationary Distribution*, Information Sciences, 2010, 180: 2560–2570.