The Sixth International Symposium on Operations Research and Its Applications (ISORA'06)
Xinjiang, China, August 8–12, 2006
, pp. 368–379

An Optimization Model for Achieving Sparsity of Gene Regulatory Networks*

Jinshan Li[1,2,†] Xiang-Sun Zhang[2]

[1] School of Science, Beijing Forestry University, Beijing 100083, China
[2] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China

Abstract Dimension problem is the main difficulty in inferring gene regulatory networks and has not been solved in substance. Grounded on linear ordinary differential equation, we propose a simple optimization model for achieving sparseness for the derived gene regulatory networks. The model is applied to gene expression profile data related to breast cancer metastasis, and the computational outcome shows that this model has potential to find solution with biological plausibility. The advantage of the model lies also in its simplicity and time saving in computation.

Keywords optimization; gene regulatory networks; microarray technique

1 Introduction

Knowledge of mRNA levels under different conditions can help people understanding how the expression levels of each gene depend on an external stimuli and on the expression levels of other genes. With high throughput experimental methods, such as DNA microarrays, mRNA expression levels of a group of genes can be measured simultaneously [1]. While the amount of available gene expression data has been increasing rapidly, the required mathematical techniques to analyze such data are still in development. Modelling of genetic regulatory networks (GRN) is becoming increasingly appealing for gaining insight into the underlying processes of living systems, but deriving a gene regulatory network from gene expression data has been proved to be difficult.

Gene expression is a complex process regulated at several stages in the synthesis of proteins [7]. A simple GRN consists of one or more input signalling pathways, several target genes, and the RNA and protein produced from the target genes [6]. In order to understand the underlying structures of activities and interactions of intracellular processes, people have to understand the dependencies of gene products and their impact on the expression of other genes. Therefore, finding a GRN for a specific

*This work was supported by NSFC under contract No. 10471141 and the Important Research Direction Project of CAS.
†E-mail: lijinshan@gmail.com

process would explain this process from a logical point of view, thus explaining many other biological phenomena.

Therefore, the model construction of GRNs has become one of the important topics in bioinformatics. However, the large number of regulatory components requires a large experimental data to infer the networks. Recently, DNA microarrays have become one of the main tools in this research area. Microarray technology enables people to monitor the activities of thousands of genes in parallel and can be applied as a powerful tools to understand the regulatory mechanism of gene systems in a cell. Microarray experiments often result in time series of measured values indicating the levels of gene expression in a genome [2]. Using these data, a valid model is able to reflect the true regulatory networks, i.e. the dependencies of the biological components.

There are several approaches to address this problem. Most of them, based on the distance between the observed data and the simulated data from the mathematical model, can excavate some biological knowledge in a sense. But some plausible and known structure property of the GRNs are always neglected in the modelling process, e.g. the sparseness of biological regulation of gene networks, that is, one gene only depends on a small proportion of the components in the system.

In this paper we propose a simple optimization model to guide the reverse engineering of time series data, and apply it to a real time-course gene expression data related to breast cancer metastasis (see table 6 in Appendix A [8]). Our model builds on the general solution of the ordinary differential equations which have been widely used to analyze GRNs. In the process, the set of ordinary differential equations is transformed into a linear matrix equation. A special solution with least L_2 norm of the matrix equation is obtained by the singular valued decomposition (SVD) technique. The special solution is further removed in the solution space to gain a new regulatory network, which remains to fit the data as well as realizes the sparse connection of the system and is therefore biological plausible.

The rest of this paper is organized as follows. Section 2 presents an overview on related works and a list of associated publications. In Section 3, we detail the method proposed in this paper. The application of this model to a biological gene expression profile data will be shown in Section 4. Conclusion and an outlook on future research will be given in the last section.

2 Related Work

Understanding the mechanisms of gene regulatory system is very interesting and promotes researchers deriving the underlying networks. In this section, a brief description of related work is given.

The first computational models to infer gene network are boolean or random boolean networks (RBN) [9, 12, 11, 10]. Boolean networks have the advantage that they can be solved with light computation effort and allow large regulatory networks to be analyzed in an efficient way. In boolean network formalism a gene is considered to be either on or off, and intermediate expression levels are neglected, which causes

that certain behaviors may not be predicted by the simulation the designed network. There are situations in which these idealization are not appropriate, and more general methods are required.

In contrast to discrete models such as boolean networks, continuous models allow the expression of gene regulation to be numerical. An example for this kind of approach is the differential equation model given in [16, 17, 3], where several models were successfully inferred. Linear differential equations are attractive because of their lower number of parameters which implies that we are less likely to over fit the data, and they are sufficient for modelling complex interactions between the genes. Although gene regulations are often nonlinear, almost all of the existing approaches for GRN inference use linear or additive models due to unclear structures of biological systems and scarcity of data [18, 4, 3].

Another popular model for inferring gene networks is the Bayesian network or dynamic Bayesian network [13, 15, 14]. Friedman and his colleagues have proposed a heuristic algorithm for the induction of Bayesian network from expression data [13]. A Bayesian network approach towards modelling regulatory networks is attractive because of its solid basis in statistics, which enables it to deal with the stochastic aspects of gene expression and noisy measurements in a natural way. But they also confront some drawbacks as well, for example, they do not allow cyclic networks, which are known to exist in biological systems.

The computational biology literature abounds in various modelling approaches, all of which have particular goals along with their strengths and weaknesses [5].

3 Method

In this paper we consider a simple but not trivial dynamical model of a gene network with an optimization model to mimic the sparsity property of the GRN. Even though this oversimplified model may not be very realistic, it will be a fundamental tool for studying and gaining insight into the basic mechanism, thus providing a valuable numerical method for deriving biologically valid gene networks.

More precisely, we treat the problem of reconstruction of a gene network in a practical situation in which the number of available data is insufficient to uniquely determine the network, which is called the dimension problem. In order to try to find more reliable solution of the network, we adopt some additional biologically relevant *a priori* assumptions such as the sparseness of the GRNs.

Differential equations are used to model gene interactions under the assumption that the transcription rate over time of each gene expression level is a function of the expression levels of some (usually a few) other genes. Such modelling assumption is based on the reaction kinetics at the biochemical level.

A fully realistic model should consider a number of relevant biological issues such as the relationships between mRNA and protein concentrations, but only the mRNA is actually measured by microarrays. Clearly, a lot of work remains to be done in the field of gene network modelling. Nevertheless, a very simple linear

model has been proved to be useful in a number of cases [18] even if it is obvious that nonlinearity is an unavoidable issue since it reflects also the nature of biochemical interactions. From the viewpoint of dynamical systems, linear equations can at least capture the main features of the network or the function. Therefore, as in [19], we consider the linear system described by the following differential equations:

$$\dot{x_i}(t) = -\lambda_i x_i(t) + \sum_{j=1}^{N} W_{ij} x_j(t) + b_i(t) + \varepsilon_i(t) \tag{1}$$

for $i = 1, 2, \cdots, N$, where the state variables x_i's are the concentration of mRNA of gene i, λ_i's are the self-degradation rates, b_i's are the external stimuli, or environment conditions, which is set to zero when there is no external input, and ε_i's represent the noise. W_{ij} describes the type and strength of the effect of the j-th gene on the i-th gene, whose positive, zero or negative signs indicate activating, naught and repressing influence respectively. The differential equation (1) can be written in a compact form as follows:

$$\dot{\mathbf{x}} = A\mathbf{x} + B + \varepsilon \tag{2}$$

where A is an $N \times N$ matrix which incorporates both self-degradation rates (on its main diagonal entries) and the strength of the gene-to-gene interaction (on its off diagonal entries) and the columns of the $N \times m$ matrix B are the b_i's, where m is the number of time points. However, sometimes we do not have the information of the external stimuli, that is to say, B in (2) is nonexisting, so (2) is changed to

$$\dot{\mathbf{x}} = A\mathbf{x} + \varepsilon. \tag{3}$$

Microarray experiments often result in discrete time series of measured values. We assume that the number of measured time points to be m, $t_1, t_2, \cdots, t_m$, and (3) can be described in a discrete form as follows:

$$\begin{pmatrix} \Delta x_1(t_2) & \cdots & \Delta x_1(t_m) \\ \vdots & \cdots & \vdots \\ \Delta x_N(t_2) & \cdots & \Delta x_N(t_m) \end{pmatrix} = A_{N\times N} \begin{pmatrix} x_1(t_1) & \cdots & x_1(t_{m-1}) \\ \vdots & \ddots & \vdots \\ x_N(t_1) & \cdots & x_N(t_{m-1}) \end{pmatrix} + \varepsilon_{N\times(m-1)} \tag{4}$$

Where $\triangle x(t_i) = (x(t_i) - x(t_{i-1}))/(t_i - t_{i-1}), i = 2, 3, \cdots, m$. The error part $\varepsilon_{N\times(m-1)}$ is neglected in our model, and in a simple way (4) becomes

$$\Delta X = AX. \tag{5}$$

It is well known that the data sets created by microarray technology contain the number of genes far more than that of the time points. It leads in the solution of the above matrix equation (5) being not uniquely determined. We can get $X = VSU'$ by the SVD technique, where V and U are orthogonal matrix with order $N \times N$ and $(m-1) \times (m-1)$ respectively, and S is a diagonal matrix. Then the particular solution is $A_0 = \triangle XUS^{-1}V'$, which is a least-$L_2$-norm solution, usually the solution does not have the sparse property.

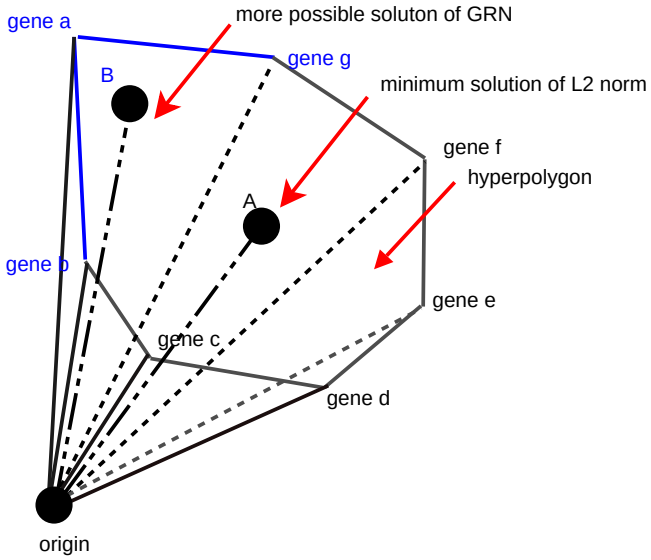

Figure 1: A more reliable solution of gene regulatory network maybe locate near to the margin of the hyper-polygon, which is the part of the hyperplane of (5) in the quadrant where the least-L_2-norm solution lies. Point A denotes the least-L_2-norm solution who comprise almost all variables, then is contradicting to the spareness assumption of biological networks. Point B may be a more reliable solution of GRN, because only gene a, b and c near to the point B have large regulating strength.

We think that a more reliable solution of gene regulatory networks should locate near to the margin of a hyper-polygon which is the part of the solution hyperplane in the quadrant where the least-L_2-norm solution lies, see figure 1. In this figure, point A denotes the least-L_2-norm solution, which possesses more variables than the point B. B comprises only three variables, e.g. a, b and c that maybe regulate the objective gene. We have known that $A_0^T + \alpha Y, \quad \forall Y \in nl, \forall \alpha \in \mathbb{R}$ are also solutions of (5), in which the real solution is contained. According to the above discussion, we establish the following optimization problem,

$$\begin{aligned} \max_{Z \in \mathbb{R}^{N-m+1}} \quad & \sum_{i=1}^{N} |a_{ij_0} + X^{\perp}(i)Z| \\ \text{s.t.} \quad & \text{sign}(a_{ij_0})(a_{ij_0} + X^{\perp}(i)Z) \geq 0, \quad i = 1,2,\cdots,N, \end{aligned} \tag{6}$$

where $A_0^T(j_0) = (a_{1j_0}, \cdots, a_{Nj_0})^T$ denotes the j_0th column of A_0^T, $X^{\perp}(i)$ is the ith $N-m+1$ dimensional row vector of the null space $X^{\perp}$, and $Z \in \mathbb{R}^{N-m+1}$, $j_0 = 1,2,\cdots,N$, where N is the number of genes. In order to avoid the objective function going to infinite, some constrained conditions are presented, which keep the sign of $a_{ij_0} + X^{\perp}(i)Z$ as same as that of a_{ij_0}, i.e., the reliable network will be searched in the hyper-

Table 1: The main computational steps of the model

Realization of the optimization model	
Step 1	Input a gene expression matrix with N genes and m time-points;
Step 2	A_0 is obtained by using the SVD technique;
Step 3	For $j_0 = 1, \cdots, N$, solve the optimization problem (6) to produce a revised solution A
Step 4	A threshold ε is determined and the elements in matrix A whose absolute value is less than the threshold are taken to zero, then output the matrix A.

polygon mentioned above. For the purpose of lucidity, the computational steps of the optimization model are listed in table 1.

4 Application

In this section, we apply our model to a real gene expression data related to breast cancer metastasis which has 27 genes and 6 time points shown in table 6, see Appendix A [8]. The content of the dataset contains gene expression data of surgical samples, including both breast cancer primary tissue and metastasis tissue, collected from 30 patients in different clinical staging. The oligonucleotide microarray technique was used to identify the gene expression profiling and screen the differential expression genes in breast cancer samples with a special emphasis on metastasis factors [8]. 27 genes were identified, 14 of which are up-regulation genes whose Ratio is as large as 3, and the rest are down-regulation gene whose Ratio is as small as 0.33 [8].

For saving the page space, we only give the computational result of the first ten genes regulated by other genes and display the results at step 2, step 3 and step 4 in the table 1 separately to show the improvement process of the solution. The computational outcome by SVD technique (step 1) listed in table 2 and table 3 (with the values filtered by the threshold 0.094) display no sparseness, which shows that the least L_2-norm solution does not possess biological plausibility. The threshold ε is taken as follows:

$$\varepsilon = 2/3(\sum_{i=1}^{N}\sum_{j=1}^{N}|A(i,j)|)/N^2, \tag{7}$$

i.e. the threshold is two third of the average expression of all genes.

Then starting from the special least norm solution at step 2, we solve the optimization problem (6) at step 3. And its computational outcome is shown in table 4. To show that the solution is as expected as we plan to posses the sparseness property, we take the same formula to determine a threshold to filter the data. Table 5 lists the outcomes with 52 interactions after step 4, where ε approximates 1 according to formula (7), which shows that the last solution has better biological plausibility than the initial solution at step 1.

Table 2: A computational result of the SVD technique without using the optimization process for sparseness. It shows the first ten genes regulated by all the genes

	gene1	gene2	gene3	gene4	gene5	gene6	gene7	gene8	gene9	gene10
gene1	-1.57	1.31	0.14	0.13	-0.22	-0.07	-0.34	-0.03	0.07	-0.75
gene2	1.09	-0.70	-0.31	-0.07	0.15	0.02	0.28	-0.19	-0.27	0.37
gene3	-1.52	-0.68	-0.60	-0.12	-0.14	-0.15	-0.06	-0.14	-0.17	-0.62
gene4	0.37	0.47	-0.22	-0.23	-0.01	-0.09	0.20	0.27	0.16	0.07
gene5	0.49	0.20	0.23	-0.01	0.00	0.00	-0.01	0.13	0.15	0.26
gene6	-0.26	-0.08	0.21	0.02	-0.06	-0.02	-0.16	0.06	0.13	0.03
gene7	1.06	0.14	0.56	0.12	0.07	0.09	-0.06	0.06	0.15	0.55
gene8	0.68	-1.53	0.27	0.21	0.13	0.12	-0.10	-0.36	-0.23	0.51
gene9	1.55	-1.06	-0.21	-0.05	0.21	0.06	0.31	-0.19	-0.26	0.63
gene10	0.39	0.41	-0.33	-0.13	0.03	-0.04	0.23	0.02	-0.08	-0.03
gene11	-1.11	-0.69	0.59	0.35	-0.11	0.06	-0.54	-0.26	0.00	-0.14
gene12	0.86	0.18	-0.23	-0.16	0.07	-0.02	0.27	0.10	-0.01	0.24
gene13	-0.33	1.21	-0.01	-0.15	-0.12	-0.10	0.00	0.34	0.29	-0.22
gene14	0.61	-1.64	-0.44	-0.00	0.16	0.03	0.20	-0.43	-0.47	0.26
gene15	0.16	0.29	-0.08	-0.06	0.00	-0.01	0.08	0.06	0.02	-0.00
gene16	0.24	0.00	0.04	-0.00	0.02	0.00	0.02	0.02	0.02	0.10
gene17	0.20	-0.02	-0.02	-0.01	0.02	0.00	0.04	-0.00	-0.01	0.06
gene18	0.18	0.42	-0.02	-0.06	-0.00	-0.01	0.07	0.11	0.07	0.01
gene19	0.35	1.12	0.00	-0.13	-0.02	-0.04	0.13	0.29	0.21	0.00
gene20	0.37	0.17	-0.06	-0.05	0.03	-0.00	0.10	0.04	-0.00	0.09
gene21	0.14	0.69	-0.01	-0.08	-0.02	-0.02	0.07	0.17	0.12	-0.02
gene22	0.07	0.32	-0.09	-0.06	-0.00	-0.02	0.07	0.06	0.02	-0.04
gene23	0.35	0.50	-0.10	-0.10	0.01	-0.02	0.14	0.12	0.05	0.03
gene24	0.98	0.42	-0.02	-0.12	0.07	0.00	0.22	0.17	0.09	0.31
gene25	0.42	0.46	0.05	-0.06	0.01	-0.00	0.08	0.15	0.12	0.12
gene26	0.58	0.73	-0.06	-0.14	0.02	-0.02	0.18	0.22	0.13	0.12
gene27	0.63	-0.34	-0.10	-0.03	0.08	0.02	0.14	-0.06	-0.10	0.23

The computational outcome indicates that our model is efficient and effective in searching gene regulatory networks with sparsity property. Without incorporating more information of a given biological system, the solution found by our model maybe inaccurate and far from the real solution. Therefore, we will enhance the accuracy of our model in the future's research by integrating more additional known information in literature or public database into the model.

5 Outlook and Future Work

Grounded on linear ordinary differential equation, we proposed a simple optimization model for achieving the sparseness of GRN. The model is applied to a microarray data related to breast cancer metastasis, and the computational outcome shows that this model can efficiently find solution with biological plausibility. The merit of the model is linear and takes less time for computation.

Due to the ambiguity in the data and there are many local solutions for the op-

Table 3: The table lists the outcome after the data in table 2 being filtered by threshold $\varepsilon = 0.094$. The total number of interactions between genes is 153. Therefore the network is not sparse since the upper bound of connections is 270.

	gene1	gene2	gene3	gene4	gene5	gene6	gene7	gene8	gene9	gene10
gene1	-1.57	1.31	0.14	0.13	-0.22	0	-0.34	0	0	-0.75
gene2	1.09	-0.70	-0.31		0.15	0	0.28	-0.19	-0.27	0.37
gene3	-1.52	-0.68	-0.60	-0.12	-0.14	-0.15	0	-0.14	-0.17	-0.62
gene4	0.37	0.47	-0.22	-0.23	0	0	0.20	0.27	0.16	0
gene5	0.49	0.20	0.23	0	0	0	0	0.13	0.15	0.26
gene6	-0.26		0.21	0	0	0	-0.16	0	0.13	0
gene7	1.06	0.14	0.56	0.12	0	0	0	0	0.15	0.55
gene8	0.68	-1.53	0.27	0.21	0.13	0.12	-0.10	-0.36	-0.23	0.51
gene9	1.55	-1.06	-0.21	0	0.21	0	0.31	-0.19	-0.26	0.63
gene10	0.39	0.41	-0.33	-0.13	0	0	0.23	0	0	0
gene11	-1.11	-0.69	0.59	0.35	-0.11	0	-0.54	-0.26	0	-0.14
gene12	0.86	0.18	-0.23	-0.16	0	0	0.27	0.10	0	0.24
gene13	-0.33	1.21		-0.15	-0.12	-0.10	0	0.34	0.29	-0.22
gene14	0.61	-1.64	-0.44	0	0.16	0	0.20	-0.43	-0.47	0.26
gene15	0.16	0.29	0	0	0	0	0	0	0	0
gene16	0.24	0	0	0	0	0	0	0	0	0.10
gene17	0.20	0	0	0	0	0	0	0	0	0
gene18	0.18	0.42	0	0	0	0	0	0.11	0	0
gene19	0.35	1.12	0	-0.13	0	0	0.13	0.29	0.21	0
gene20	0.37	0.17	0	0	0	0	0.10	0	0	0
gene21	0.14	0.69	0	0	0	0	0	0.17	0.12	0
gene22	0	0.32	0	0	0	0	0	0	0	0
gene23	0.35	0.50	-0.10	-0.10	0	0	0.14	0.12	0	0
gene24	0.98	0.42	0	-0.12	0	0	0.22	0.17	0	0.31
gene25	0.42	0.46	0	0	0	0	0	0.15	0.12	0.12
gene26	0.58	0.73	0	-0.14	0	0	0.18	0.22	0.13	0.12
gene27	0.63	-0.34	-0.10	0	0	0	0.14	0	-0.10	0.23

timization problem, it is difficult for the proposed model to find the true solution. As one future enhancement of the proposed method, we plan to incorporate some additional information to identify the correct network. As the initial results showed, precautions have to be taken to prevent the model from finding the trivial solution without biological significance. In future's work we plan to take the concept of time delay into our model, which is essential in biological networks. More *a priori* information can be imported into the inference process of real microarray data. These information include partially known pathways and information about co-regulated genes, which can be found in literature or in public databases. This would enable our method to search for models consistent with current biological knowledge, but would also allow for alternative solutions where biological information is missing or faulty. Furthermore, non-linear interaction could appear in our model for enhancing the precise of gene regulatory networks to overcome the insufficiency of the proposed model.

Table 4: This table lists the solutions of relations between the first ten genes and the total 27 genes after the step 3.

	gene1	gene2	gene3	gene4	gene5	gene6	gene7	gene8	gene9	gene10
gene1	-0.08	0.01	0.97	0.01	0.00	-0.00	0	-1.54	0.00	-0.00
gene2	0	-7.44	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	0.66
gene3	0	0.00	0.00	0.00	-0.00	0.00	-0.11	-1.76	-0.48	-0.02
gene4	6.75	2.62	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00
gene5	5.46	-0.00	0.61	-4.38	0.10	2.60	-5.37	-0.00	-0.00	0.00
gene6	0.00	-4.41	0.00	2.36	0.00	-2.16	-0.00	1.95	0.00	-0.00
gene7	-0.00	3.20	0.00	2.20	-0.00	-0.00	-0.00	0.18	-0.00	-0.00
gene8	0	-0.00	-0.00	0.18	-0.00	0.00	-0.00	0.00	-0.00	0.00
gene9	0	0.00	-0.16	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00
gene10	-0.00	13.75	0.00	0.00	0.00	0.00	0	7.94	0.00	-0.00
gene11	0	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.69	0.00
gene12	0	-0.00	0.00	0.00	0.00	-0.00	0	-0.00	-0.82	-0.00
gene13	16.09	0.00	-1.56	-0.00	0.00	0.00	2.89	-0.00	0.00	-1.16
gene14	-0.00	-0.00	-0.00	0.00	-0.00	0.00	1.56	-0.00	0	0.00
gene15	0.00	-0.00	0.00	-0.00	17.14	0.00	0.00	-0.00	16.73	-21.77
gene16	8.92	28.36	6.59	-2.50	0.00	0.00	23.35	7.53	10.92	-0.00
gene17	-0.00	-62.20	-0.00	-0.00	0.00	0.84	0	-47.18	-0.00	2.18
gene18	0.76	-0.00	-17.53	-0.00	-13.01	-0.00	0.00	-0.00	-0.00	19.58
gene19	0.00	0	38.80	0.00	0.00	-0.00	0	-0.00	0.00	11.59
gene20	-0.00	-0.00	0.00	-0.00	0.00	-3.63	0	-0.00	-14.22	-0.00
gene21	13.51	0.00	-47.56	-0.00	0.00	0.00	-0.00	-0.00	0.00	-31.51
gene22	8.63	0.00	0.00	0.00	-9.05	0.00	-0.00	-0.00	2.14	-0.00
gene23	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	12.77
gene24	0	0.17	-4.95	0.00	0.00	0.88	0	-0.00	-0.00	-0.00
gene25	0.00	0.10	0.05	0.00	3.85	-2.37	-0.00	-0.00	0.00	-0.00
gene26	0	-0.00	-0.17	0.00	0.76	0.00	0.00	-0.00	0.00	-0.00
gene27	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00

References

[1] Losifina Pournara and Lorenz Wernisch. Reconstruction of gene netwoeks using Bayesian learning and manipulation experiments. *Bioinformatics*, 20(17), 2934–2942, 2004.

[2] M. Schena, D. Shalon, R. W. Davis and P. Q. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270, 467–470, 1995.

[3] Y. Wang, T. Joshi, X. S. Zhang, D. Xu and L. N. Chen. Inferring gene regulatory networks from multiple microarray datasets. In submission.

[4] M. Gustafsson, M. Hornquist and A. Lombardi. Constructing and analyzing a large-scale gene-to-gene regulatory network-lasso-constrained inference and biological validation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(3), 254–261, 2005.

Table 5: The table lists the computational outcome after a threshold filtering. The threshold is taken to 1. This table makes us more clear that the solution has biological plausibility since there are totally 52 interactions.

	gene1	gene2	gene3	gene4	gene5	gene6	gene7	gene8	gene9	gene10
gene1	0	0	0	0	0	0	0	-1.5	0	0
gene2	0	-7.4	0	0	0	0	0	0	0	0
gene3	0	0	0	0	0	0	0	-1.7	0	0
gene4	6.7	2.6	0	0	0	0	0	0	0	0
gene5	5.4	0	0	-4.3	0	2.6	-5.3	0	0	0
gene6	0	-4.4	0	2.3	0	-2.1	0	1.9	0	0
gene7	0	3.2	0	2.2	0	0	0	0	0	0
gene8	0	0	0	0	0	0	0	0	0	0
gene9	0	0	0	0	0	0	0	0	0	0
gene10	0	13.7	0	0	0	0	0	7.9	0	0
gene11	0	0	0	0	0	0	0	0	0	0
gene12	0	0	0	0	0	0	0	0	0	0
gene13	-16.0	0	-1.5	0	0	0	2.9	0	0	-1.2
gene14	0	0	0	0	0	0	1.5	0	0	0
gene15	0	0	0	0	17.1	0	0	0	16.7	-21.7
gene16	8.9	28.3	6.6	-2.5	0	0	23.3	7.5	10.9	0
gene17	0	-62.2	0	0	0	0	0	-47.2	0	2.2
gene18	0	0	-17.5	0	-13.0	0	0	0	0	19.6
gene19	0	0	38.8	0	0	0	0	0	0	11.6
gene20	0	0	0	0	0	-3.6	0	0	-14.2	0
gene21	13.5	0	-47.5	0	0	0	0	0	0	-31.5
gene22	8.6	0	0	0	-9.0	0	0	0	2.1	0
gene23	0	0	0	0	0	0	0	0	0	12.8
gene24	0	0	-4.9	0	0	0	0	0	0	0
gene25	0	0	0	0	3.8	-2.3	0	0	0	0
gene26	0	0	0	0	0	0	0	0	0	0
gene27	0	0	0	0	0	0	0	0	0	0

[5] H. D. Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1), 67–103, 2002.

[6] Christian Spieth, Felix Streichert, Nora Speer and Andreas Zell. Multi-objective model optimization for inferring gene regulatory networks. C. A. Coello *et al.* (eds.), *EMO 2005*, LNCS 3410, 607–620, Springer-Verlag Berlin Heidelberg, 2005.

[7] B. Lewin. *Gene VII*, Oxford University Press, Oxford, 1999.

[8] C. Ni, B. Sun, Y. Feng, D. Zhang, X. Li, L. Dong and L. Zhang. Primary research of linear differential model for the genetic regulatory networks of gene related to breast cancer metastasis. *Progress in Biochemistry and Biophysics*, 33(12), 1165–1172, 2005. (in Chinese)

[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, 90(11), 1778–1792, 2002.

[10] E. R. Dougherty and I. Shmulevich. Mappings between probabilistic boolean networks. *Signal Processing*, 83(4), 799–809, 2003.

[11] R. F. Hashimoto, S. Kim, I. Shmulevich, W. Zhang, M. L. Bittner and E. R. Dougherty. Growing genetic regulatory networks from seed genes. *Bioinformatics*, 20(8), 1241–1247, 2004.

[12] I. Shmulevich and S. A. Kauffman. Activities and sensitivities in boolean network models. *Physical Review Letters*, 93(4), 1–4, 2004.

[13] N. Friedman, M. Linial, I. Nachman and D. Peér. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4), 601–620, 2000.

[14] S. Y. Kim, S. Imoto and S. Miyano. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Briefings in Bioinformatics*, 4(3), 228–235, September 2003.

[15] Min Zou and Suzanne D. Conzen. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1), 71–79, 2005.

[16] T. Chen, H. L. He and G. M. Church. Modeling gene expression with differential equations. *Proceedings of the Pacific Symposium on Bio-computing*, 4, 29–40, 1999.

[17] Michiel de Hoon, Seiya Imoto and Satoru Miyano. Inferring gene regulatory networks from time-ordered gene expression data using differential equations. *Discovery Science: 5th International Conference*, 2534, 267–274, 2002.

[18] P. D'haeseleer, X. Wen, S. Fuhrman and R. Somogyi. Linear modeling of mRNA levels during CNS development and injury. *Proceedings of the Pacific Symposium on Biocomputing*, 4, 41–52, 1999.

[19] T. S. Gardner, D. di Bernardo, D. Lorentz and J. J. Collins. Reverse engineering gene networks and identifying compound mode of action via expression profiling. *Science*, 301, 102–105, 2003.

[20] J. A. Jensen, R. E. Carroll, R. V. Benya, *et al.* The case for gastrin-releasing peptide acting as a morphogen when it and its receptor are aberrantly expressed in cancer. *Peptides*, 22(4), 689–699, 2001.

Appendix A Time-course gene expression data related to breast cancer metastasis

The following table contains breast-cancer-metastasis gene expression data with 27 genes and 6 time points [8].

Table 6: Time-course gene expression data related to breast cancer metastasis with 27 genes and 6 time points [8].

ID	Name of gene	GB.accession	t0	t1	t2	t3	t4	t5
1	C7	NM_000587	3.34121	2.52946	1.14962	2.60146	3.26723	2.74518
2	MS4A1	NM_021950	2.20179	1.92798	0.85207	1.79855	1.71894	4.23766
3	TCL1A	NM_021966	2.45153	2.90927	1.50698	1.21093	0.81390	0.74240
4	PDE6H	NM_006205	2.10673	2.40264	1.71421	1.05099	0.96903	0.34485
5	C8	AL049265	2.46396	2.00801	1.55623	1.43118	1.39612	0.18583
6	CR2	NM_001877	2.59425	2.33566	1.67739	1.2959	1.00653	-0.2989
7	EPHA3	AF213459	2.30812	0.77683	1.06821	1.42411	1.46094	1.10292
8	P2RX5	NM_002561	2.19251	1.21243	0.59190	1.28985	0.00001	1.06793
9	MAL	NM_002371	1.58432	1.38382	0.65715	1.24737	0.33575	1.30913
10	RGS1	NM_112922	1.81627	1.59915	0.30823	1.51197	2.01035	-0.1520
11	CD69	NM_001781	2.53768	1.39257	0.54790	1.14642	0.95727	0.83420
12	CD19	NM_001770	1.44899	1.52744	0.81232	0.98293	1.07732	0.02047
13	NTS	NM_00613	2.16028	2.10508	1.41663	1.06212	1.67674	4.48989
14	VLCS-H1	NM_014031	1.58027	1.65533	-0.1275	1.22935	0.20798	0.52290
15	COL11A1	NM_001854	-3.3161	-2.0804	-3.0678	-1.7776	-0.5645	-0.1640
16	GRP	NM_002091	-1.3116	-2.0925	-1.6290	-2.1934	-3.1304	-0.3964
17	MMP13	NM_002427	-1.1922	-1.6903	-2.3104	-1.4250	-1.3724	-1.9142
18	SFRP2	AF156100	-1.7357	-1.7517	-1.7018	-1.6943	-0.4419	-1.4481
19	C9	AK026320	-1.7078	-1.9841	-0.8269	-1.3158	0.55405	-1.5373
20	FIGL6	AF156100	-1.2134	-1.7702	-2.1249	-0.8500	-0.2737	-1.5152
21	BPAG1	NM_001723	-1.9883	-2.1649	-1.6438	-1.7930	0.00001	-0.2337
22	MMP1	NM_002421	-2.0907	-1.4412	-2.6885	-1.5775	-0.3662	1.52346
23	C10	AK022342	-3.0795	-1.8282	-1.9676	-1.2781	-0.0433	-0.8540
24	C11	AK022198	-2.3102	-2.1909	-0.5531	-1.1064	-0.7373	-0.9374
25	C12	AK016784	-0.8538	-1.5268	-0.7189	-1.4255	-0.5974	-0.2140
26	KRT17	NM_000422	-2.5797	-1.2558	-0.6813	-1.4386	-0.1230	-0.1660
27	MMP3	NM_002422	-1.4914	-1.5166	-2.2186	-0.7820	-1.9914	-1.9525