

An Efficient Convex Nonnegative Network Component Analysis for Gene Regulatory Network Reconstruction

Jisheng Dai^{1,2}, Chunqi Chang¹, Zhongfu Ye², and Yeung Sam Hung¹

¹ Department of Electrical and Electronic Engineering

The University of Hong Kong, Pokfulam Road, Hong Kong

² Department of Electronic Engineering and Information Science

University of Science and Technology of China, Hefei, Anhui, P.R. China

jsdai@eee.hku.hk, cqchang@eee.hku.hk, yezf@ustc.edu.cn, yshung@eee.hku.hk

Abstract. A systems biology problem of reconstructing gene regulatory network from time-course gene expression microarray data via network component analysis (NCA) is investigated in this paper. Inspired by the idea that each column of the connectivity matrix can be estimated independently, we try to propose a fast and stable convex approach for nonnegative NCA (nnNCA). Compared with the existing method, our new method reduces the computational cost substantially, whereas maintains a reasonable accuracy. Both the simulation results and experimental results demonstrate the effectiveness of our method.

Keywords: Gene regulatory network, microarray, network component analysis, convex programming, positivity constraints.

1 Introduction

Data from high-throughput DNA microarrays and ChIP-chip binding assays have become the basis of transcriptional regulatory analysis in the post-genome period. The study of gene regulatory network is important to understand the information embedded in these data. To reveal the underlying inter-dependency and cause-and effect relationship between various cellular functions, much effort has been devoted to the research of gene regulatory network reconstruction from microarray data.

Developing tractable system identification techniques capable of reconstructing gene regulatory networks of a large size from small microarray datasets is a key challenge for gene regulatory network reconstruction. Several different approaches have been proposed to modeling gene regulatory network in early studies, including dynamic Bayesian networks (DBN) [1], probabilistic Boolean network (PBN) [2], and differential/difference equation [3]. More recently, with the assumption of linear instantaneous signal models, statistical techniques such as the principal component analysis (PCA) [4] and independent component analysis (ICA) [5] have been applied successfully to deduce biologically significant information from

microarray datasets. However, neither the reconstructed networks nor the reconstructed regulatory signals is consistent with real biological systems in general, because in these approaches the assumption of mutually orthogonality or statistical independence for the regulatory signals are not usually valid.

The network component analysis (NCA) has been shown to be a very effective approach because it incorporates helpful and biologically sound assumptions [6, 7, 8]. It is well known that a real gene regulatory network always has a sparse structure, and sometimes the non-zero entries of its (sparse) connectivity matrix may be obtained from ChIP-chip experiments [9]. Even when experimental data on the structure of the network are unavailable, the structural information sometimes may be extracted partially from the literature or predicted by bioinformatics methods [10]. The NCA approach can make use of this (sparse) structure information to fully reconstruct the network (including the connectivity matrix and the regulatory signals) in the case of noise-free, if some mild conditions, called NCA criteria, are satisfied. Most recently, the method was further improved in [11], where a closed-form solution to the NCA problem is obtained through fitting the model by a series of subspace projections.

So far, NCA is one of the most effective approaches to gene regulatory network reconstruction, existing algorithms, however, lack accuracy and consistency. To improve the performance of NCA and develop a more robust network reconstruction method, convex programming methods incorporating the nonnegativity constraints on the connectivity matrix are proposed in [12, 13]. Although this convex optimization problem can be solved by an inner-point algorithm, the required complexity is still very high, if the number of the entries in the connectivity matrix is large which is often happened in biological environments. This motivates us to find an efficient nonnegative network component analysis (nnNCA) approach for gene regulatory network reconstruction subject to the nonnegativity constraints. Inspired by the idea that each column of the connectivity matrix may be solved independently [11], we successfully derive a very fast and stable approach to nnNCA in this paper. Compared with the existing method, the computational cost is reduced substantially. Both the simulation results and experimental results demonstrate the effectiveness of our method.

2 Network Component Analysis

Figure 1 illustrates the instantaneous linear regulation model we are addressing, which is a bipartite network. Gene expressions are regulated by transcription factors. The upper layer in the figure represents the expression level of activated transcription factors (TF) or transcription factor activities (TFA), and the lower layer represents the microarray gene expression data. In general, gene regulation processes are dynamic and nonlinear, but are approximately log-linear [14]. Assuming that in the network we have N genes and M transcription factors, and the length of time series is K . Then if the gene expression are represented as log-ratios, the network can be modeled as

$$\mathbf{X} = \mathbf{AS} + \mathbf{\Gamma} \tag{1}$$

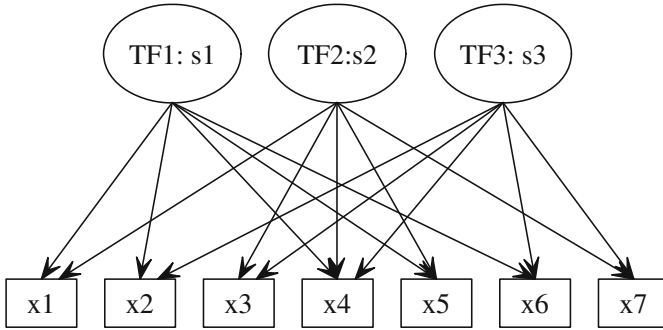


Fig. 1. Transcriptional regulatory model. The expression of genes (lower layer) are regulated by transcription factors (upper layer) through a bipartite network.

where $\mathbf{X} \in \mathbb{R}^{N \times K}$, $\mathbf{A} \in \mathbb{R}^{N \times M}$, $\mathbf{S} \in \mathbb{R}^{M \times K}$ and $\mathbf{\Gamma} \in \mathbb{R}^{N \times K}$ are the gene expression levels which can be measured by microarray data, the connectivity matrix of the network, the unknown TFAs, and the measurement noise, respectively. It should be noted that the connectivity matrix \mathbf{A} are assumed sparse, which agrees well with observations in biological environments.

Our goal is to recover the connectivity matrix \mathbf{A} and the TFAs \mathbf{S} from the time-course microarray data \mathbf{X} . It is proven in [6] that \mathbf{A} and \mathbf{S} can be estimated up to possible scaling factors, which is the inherent indeterminacy for any blind source separation problem, if the following *NCA* criteria are satisfied:

1. the connectivity matrix \mathbf{A} must be full-column rank;
2. when an element in the regulatory domain is removed together with the rows corresponding to the nonzero entries of this column, the connectivity matrix of the resulting network is still of full-column rank;
3. the TFAs \mathbf{S} must have full-row rank.

Under the constraint that the non-zero pattern of \mathbf{A} has the structure corresponding to the known a priori network topology, the \mathbf{A} and \mathbf{S} can be uniquely determined by solving the optimization problem as follows.

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{S}} \quad & \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F \\ \text{subject to:} \quad & \mathbf{A}(I) = 0 \end{aligned} \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, and I contains the indices where the entries of \mathbf{A} are zeros, which is induced by the network structure. Since the actual estimation of \mathbf{A} and \mathbf{S} is performed by a two-step alternating least-squares algorithm, very high computational complexity is required. To reduce the computational cost, [11] makes a effort to develop a fast, stable and globally convergent algorithm for reconstructing the gene-regulatory networks, referred as *FastNCA* algorithm, where a closed-form solution to the NCA problem is given. Interestingly, it was shown that the connectivity matrix \mathbf{A} can be estimated

first, especially each column of \mathbf{A} can be obtained independently, then the TFA can be obtained simply by

$$\mathbf{S} = \mathbf{A}^\dagger \mathbf{X} \quad (3)$$

where \mathbf{A}^\dagger is the pseudo inverse of \mathbf{A} .

However, all the NCA approaches only use the structure information about the connectivity matrix. In attempt to improve the performance of NCA method, as well as develop a more accurate and robust method for the network reconstruction, a convex algorithm incorporating another new prior information (the entries of connectivity matrix \mathbf{A} are all nonnegative, also referred as nnNCA) is proposed in [12]. It is reasonable to put the nonnegative constraint on the entries of the connectivity matrix \mathbf{A} , as long as any specific transcription factor has the same effect to all genes. The sound biological support for this assumption can be found in [15]. With this nonnegative constraint, in [12], it is shown that the connectivity matrix \mathbf{A} can be estimated by solving the following optimization problem

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{S}} \quad & \|\mathbf{C}^T \mathbf{A}\|_F \\ \text{subject to:} \quad & \mathbf{A}(I) = 0 \\ & \mathbf{A}(J) \geq c \end{aligned} \quad (4)$$

where \mathbf{C} is the null space of matrix \mathbf{X} , J contains the indices where the entries of \mathbf{A} are nonzero (positive), and c is a constant small positive value. Although the above convex optimization problem can be solved by inner-point algorithm [16], the required computational complexity is still very high, if the number of the entries in matrix \mathbf{A} is large which is often happened in real biological circumstance.

Inspired by the idea that each column of the connectivity matrix \mathbf{A} may be solved independently [11], we try to propose a very efficient approach to nnNCA based on individual estimation of column of \mathbf{A} in the next section.

3 Efficient Approach to nnNCA

It is well known that if there is noise in the model Eq. (1), the effect from noise can be coped with by applying the singular value decomposition (SVD) [17] to matrix \mathbf{X} . For simplicity of analysis, we first write \mathbf{X} (assuming no noise) in the standard SVD form as follows.

$$\mathbf{X} = \mathbf{U} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^T = [\mathbf{U}_s \ \mathbf{U}_n] \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^T = \mathbf{U}_s [\Sigma \ \mathbf{0}] \mathbf{V}^T \quad (5)$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$ and $\mathbf{V} \in \mathbb{R}^{K \times K}$ are the singular-vector matrices, $\mathbf{U}_s \in \mathbb{R}^{N \times M}$ and $\mathbf{U}_n \in \mathbb{R}^{N \times (N-M)}$ are the singular-vector matrices corresponding to M nonzero singular-values and $N-M$ zero singular-values, respectively, and $\Sigma \in \mathbb{R}^{M \times M}$ is the singular-value matrix containing nonzero singular-values arranged

in a decreasing order from top-left to bottom-right. From the subspace principle [18], we know that there always exists a matrix $\mathbf{T} \in \mathbb{R}^{M \times M}$ such that

$$\mathbf{U}_s = \mathbf{A}\mathbf{T} \quad (6)$$

Notice that the above equation only holds if there is no noise in Eq. (1). Nevertheless, $\mathbf{A}\mathbf{T}$ can provide a robust estimate of \mathbf{X} in the existence of noise. Therefore, we still assume that Eq. 6 holds in noise case. Since we can always re-order the rows of \mathbf{X} and \mathbf{A} , we assume without loss of generality that the first column of \mathbf{A} can be partitioned as

$$\mathbf{a}_1 = \begin{bmatrix} \tilde{\mathbf{a}} \\ \mathbf{0}_{L \times 1} \end{bmatrix} \quad (7)$$

where L is number of zero entries in the vector \mathbf{a}_1 , and $\tilde{\mathbf{a}} \in \mathbb{R}^{(N-L) \times 1}$.

To show that each column of the connectivity matrix \mathbf{A} can be solved independently for nnNCA, we need the following lemma, which forms the basis of our method.

Lemma 1. *If there is a vector $\mathbf{t} \in \mathbb{R}^{M \times M}$ satisfying*

$$\mathbf{U}_s \mathbf{t} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0}_{L \times 1} \end{bmatrix} \quad (8)$$

where $\mathbf{b} \in \mathbb{R}^{(N-L) \times 1}$, then \mathbf{b} must be an estimate of $\tilde{\mathbf{a}}$ up to a scaling ambiguity η , i.e.,

$$\mathbf{b} = \eta \tilde{\mathbf{a}} \quad (9)$$

Proof:

Using Eq. (6) and Eq. (8), we have

$$\mathbf{U}_s \mathbf{t} = \mathbf{A}\mathbf{T}\mathbf{t} = \mathbf{A}\hat{\mathbf{t}} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \quad (10)$$

where $\hat{\mathbf{t}} \triangleq \mathbf{T}\mathbf{t}$. We partition the above equation accordingly as

$$\mathbf{A}\hat{\mathbf{t}} = \begin{bmatrix} \tilde{\mathbf{a}} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \eta \\ \mathbf{t}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \quad (11)$$

or, equivalently,

$$\mathbf{b} = \eta \tilde{\mathbf{a}} + \mathbf{A}_{12} \mathbf{t}_2 \quad (12)$$

$$\mathbf{0} = \mathbf{A}_{22} \mathbf{t}_2 \quad (13)$$

Since \mathbf{A}_{22} is of full-column rank determined by NCA criterion (2), according to Eq. (13) we are able to obtain

$$\mathbf{t}_2 = \mathbf{0} \quad (14)$$

Then substituting the above equation into Eq. (12), we achieve the desired result

$$\mathbf{b} = \eta \tilde{\mathbf{a}} \quad (15)$$

End of proof

Relying on Lemma 1, we are able to derive a very simply approach to estimate $\tilde{\mathbf{a}}_1$. Since we can always re-order the columns of \mathbf{A} and the rows of \mathbf{T} accordingly while keeping the equality in Eq. (6), the other columns of \mathbf{A} can be estimated in the same way. Keeping these results in mind, we now proceed to solve the NCA problem with positivity constraints on the nonzero entries of the connectivity matrix. Unlike the strategy employed in [12], we follow a different approach where each column of the connectivity matrix \mathbf{A} is estimated independently, so that the computational cost will be reduced substantially. The optimization problem for estimating the i -th column (denoted as \mathbf{a}_i) with positivity constraints can be formulated as

$$\begin{aligned} \min_{\mathbf{t}_i} \quad & \|\mathbf{U}_s(I_i, :) \cdot \mathbf{t}_i\|_F \\ \text{subject to: } & \mathbf{U}_s(J_i, :) \cdot \mathbf{t}_i \geq c \end{aligned} \quad (16)$$

where I_i contains the indices where the entries of \mathbf{a}_i are zeros, J_i contains the indices where the entries of \mathbf{a}_i are nonzero, and $\mathbf{U}_s(I_i, :)$ denotes the rows of \mathbf{U}_s corresponding to indices I_i . Eq. (16) is convex and can be solved by effective algorithms [16]. Recall Lemma 1, it is not difficult to see that the estimate of \mathbf{a}_i can be given by

$$\mathbf{a}_i = \mathbf{U}_s \mathbf{t}_i \quad (17)$$

From the above discussions, now we summarize the proposed algorithm as follows.

Efficient Algorithm for nnNCA:

1. Perform a standard SVD on \mathbf{X} and obtain the \mathbf{U}_s as in Eq. (5);
2. Estimation of \mathbf{A}
 - For $i = 1, 2, \dots, M$
 - a) solve the optimization problem (16) and get the optimal \mathbf{t}_i ;
 - b) let $\mathbf{a}_i = \mathbf{U}_s \mathbf{t}_i$.
 - End
3. Estimate the TFA matrix by $\mathbf{S} = \mathbf{A}^\dagger \mathbf{X}$.

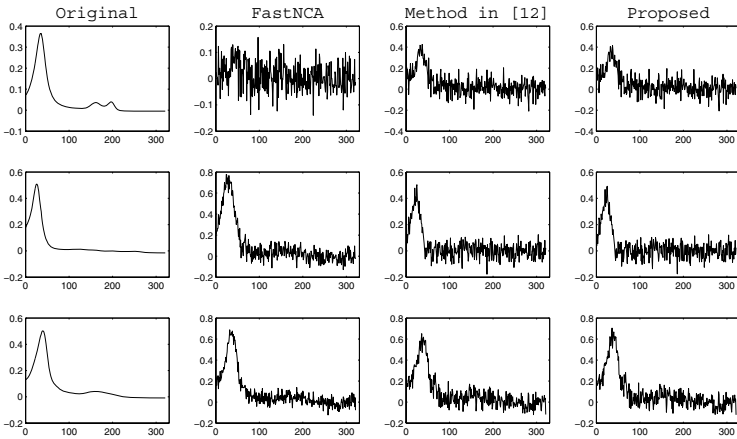
4 Results

4.1 Simulation Results

In order to test the proposed method, we use the simulation data described in [6, 12]. This is a hemoglobin spectroscopy data set with a 7×3 connectivity matrix and a 7×321 measurement data. If positivity constraints on the nonzero entries of the connectivity matrix and the noise are not considered, the same performances are achieved by the FastNCA, the method in [12], and our proposed

Table 1. True connectivity matrix and its estimation by FastNCA, the method in [12], and our proposed method without positivity constraints and noise

True			FastNCA			Method in [12]			Proposed		
0.417	1.184	0	0.435	1.196	0	0.435	1.196	0	0.435	1.196	0
2.083	0	1.25	2.12	0	1.267	2.12	0	1.267	2.12	0	1.267
0	1.184	0.25	0	1.164	0.202	0	1.164	0.202	0	1.164	0.202
0.417	1.053	0.25	0.429	1.056	0.236	0.429	1.056	0.236	0.429	1.056	0.236
1.25	0.921	0	1.176	0.904	0	1.176	0.904	0	1.176	0.904	0
0.833	0	2.00	0.841	0	2.015	0.841	0	2.015	0.841	0	2.015
0	0.658	1.25	0	0.681	1.281	0	0.681	1.281	0	0.681	1.281

**Fig. 2.** True sources and their corresponding estimates by FastNCA, the method in [12], and our proposed method subject to positivity constraints under noisy case

method. The normalized true (original) and estimated connectivity matrices are shown in Table 1.

In the next simulation, we study the performance of the proposed algorithm subject to positivity constraints in the noisy case, where the signal to noise ratio (SNR) is set to 9 dB. The true and the estimated source spectra signals are shown in Figure 2, and the normalized true and estimated connectivity matrices are shown in Table 3. The simulation results in the table shows that the connectivity matrix estimated by FastNCA contains both positive and negative entries. Since the true entries connectivity matrix are either nonnegative or nonpositive, such an estimate is unfavorable. The simulation results in the figure shows that the estimated connectivity matrix achieved by the method in [12] and our method are almost the same. This well demonstrates the optimality of our method.

Table 2. True connectivity matrix and its estimation by FastNCA, the method in [12], and our proposed method subject to positivity constraints under noisy case

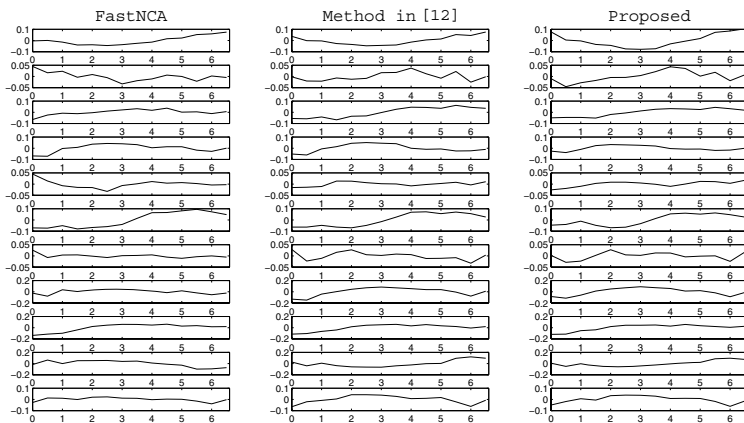
True			FastNCA			Method in [12]			Proposed		
0.417	1.184	0	-1.785	1.028	0	0.985	1.029	0	1.12	1.028	0
2.083	0	1.25	-0.202	0	2.065	1.505	0	1.236	1.485	0	1.358
0	1.184	0.25	0	1.244	-0.468	0	1.245	0.67	0	1.244	0.604
0.417	1.053	0.25	-0.947	0.966	0.096	1.027	0.965	0.67	0.917	0.966	0.604
1.25	0.921	0	-0.279	1.087	0	0.741	1.087	0	0.74	1.087	0
0.833	0	2.00	1.788	0	1.782	0.741	0	1.754	0.74	0	1.537
0	0.658	1.25	0	0.675	0.59	0	0.674	0.67	0	0.675	0.898

Table 3. The running time comparison between our method and the method in [12]

	Proposed	Method in [12]
Running Time	6.71 (seconds)	1144.62 (seconds)

4.2 Experimental Results

To test the proposed method in the real biological networks, we apply all the methods to analyze the time-course Yeast cell cycle microarray data used in [19]. Note that the network topology data required by the algorithm are from [9]. The microarray data contains 441 genes and 56 time points. In this analysis, we are interested in the 11 transcription factors that are known to regulate the expression of genes that are involved in the cell cycle process. Three experiments with different synchronization methods represented by alpha, *cdc15*, and elutriation

**Fig. 3.** Estimated yeast cell cycle related TFAs for alpha synchronization by FastNCA, the method in [12], and our proposed method

are studied. The estimated TFAs for these 11 cell cycle related transcription factors with different synchronization methods are shown in Figure 3, Figure 4 and Figure 5 for FastNCA, the method in [12], and our proposed method, respectively. It can be seen that the method in [12] and our proposed method produce almost the same results, and it is obvious that the estimated TFAs subject to positivity constraints show more cyclic behavior than that achieved by FastNCA. Cyclic behavior is expected in this case since all these transcription factors are cell cycle related. The results demonstrate the superior robustness of the nnNCA methods and the equivalence between our method and the existing method.

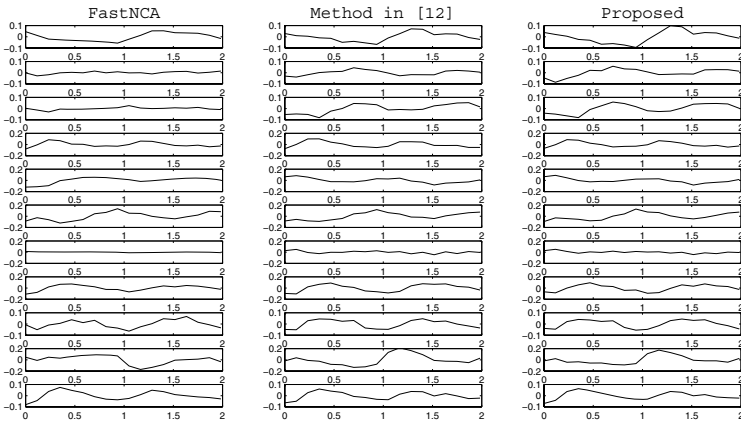


Fig. 4. Estimated yeast cell cycle related TFAs for *cdc15* synchronization by FastNCA, the method in [12], and our proposed method

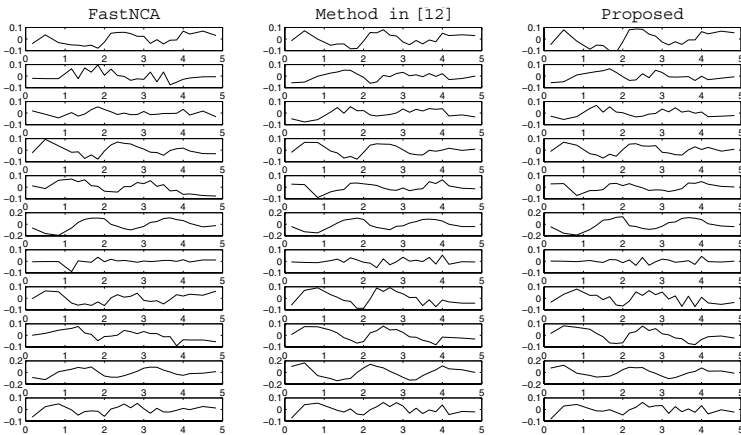


Fig. 5. Estimated yeast cell cycle related TFAs for elutriation synchronization by FastNCA, the method in [12], and our proposed method

Finally, to testify the improvement in computational cost, Table 3 gives the running time for our method and the method in [12]. All the computer codes are developed in MATLAB 7.7.0 on a Intel Core2 2.66GHz with 2GB of RAM. The result in the table shows that our method is much faster than the method in [12].

Both the simulation results and experimental results demonstrate the effectiveness of our proposed method. Compared with the existing method, our new method reduces the computational cost substantially, whereas maintains a reasonable accuracy.

5 Conclusion

Both the simulation results and experimental results shows the superior robustness of the nnNCA methods, and demonstrate that the performances of our method and the method in [12] are very similar. The contribution of this paper is to derive a powerful convex approach for nnNCA, which reduces the computational cost substantially, whereas maintains a reasonable accuracy.

Acknowledgements

This work is supported by a Seed Funding for Basic Research of the University of Hong Kong.

References

1. Friedman, N., Linial, M., Nachman, I., Peer, D.: Using bayesian networks to analyze expression data. *Journal of Computational Biology* 7(3), 601–620 (2000)
2. Shmulevich, I., Dougherty, E.R., Kim, S., Zhang, W.: Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18(2), 261–274 (2002)
3. Dougherty, E.R., Shmulevich, I., Chen, J., Wang, Z.J.: *Genomic Signal Processing and Statistics*. Hindawi Publishing Corporation (2005)
4. Alter, O., Brown, P.O.: Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences of the United States of America* 97(18), 10101–10106 (2000)
5. Lee, S.I., Batzoglou, S.: Application of independent component analysis to microarrays. *Genome Biology* 4(11), R76 (2003)
6. Liao, J.C., Boscolo, R., Yang, Y.L., Tran, L.M., Sabatti, C., Roychowdhury, V.P.: Network component analysis: Reconstruction of regulatory signals in biological systems. *Proceedings of the National Academy of Sciences of the United States of America* 100(26), 15522–15527 (2003)
7. Chang, C.Q., Hung, Y.S., Fung, P.C.W., Ding, Z.: Network component analysis for blind source separation. In: *Proc. 2006 International Conference on Communications, Circuits and Systems*, Guilin, China, pp. 323–326 (2006)
8. Chang, C.Q., Ding, Z., Hung, Y.S., Fung, P.C.W.: Fast network component analysis for gene regulation networks. In: *Proc. 2007 IEEE International Workshop on Machine Learning for Signal Processing*, Thessaloniki, Greece, pp. 21–26 (2007)

9. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A.: Transcriptional regulatory networks in *saccharomyces cerevisiae*. *Science* 298(5594), 799–804 (2002)
10. Wang, C., Xuan, J., Chen, L., Zhao, P., Wang, Y., Clarke, R., Hoffman, E.: Motif-directed network component analysis for regulatory network inference. *BMC Bioinformatics* 9(suppl. 1), S21 (2008)
11. Chang, C.Q., Ding, Z., Hung, Y.S., Fung, P.C.W.: Fast network component analysis (FastNCA) for gene regulatory network reconstruction from microarray data. *Bioinformatics* 24(11), 1349–1358 (2008)
12. Chang, C.Q., Ding, Z., Hung, Y.S.: A new optimization algorithm for network component analysis based on convex programming. In: Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan (2009)
13. Chang, C.Q., Ding, Z., Hung, Y.S.: Nonnegative network component analysis by linear programming for gene regulatory network reconstruction. In: Adali, T., et al. (eds.) ICA 2009. LNCS, vol. 5441, pp. 395–402. Springer, Heidelberg (2009)
14. Savageau, M.A.: *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading (1976)
15. Alon, U.: *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton (2007)
16. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
17. Golub, G.H., van Loan, C.F.: *Matrix Computation*, 3rd edn. The Johns Hopkins University Press (1996)
18. Lay, D.C.: *Linear Algebra and Its Applications*, 2nd edn. Addison-Wesley, New York (2000)
19. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., Futcher, B.: Comprehensive identification of cell cycleregulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell* 9(12), 3273–3297 (1998)