A Non-Gaussian Factor Analysis Approach To Transcription Network Component Analysis

Shikui Tu
Department of Computer Science and Engineering,
The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong
Email: sktu@cse.cuhk.edu.hk

Dingsheng Luo
Speech and Hearing Research Center,
Key Lab of Machine Perception (Ministry of Education),
Peking University, Beijing, 100871, China
Email: dsluo@cis.pku.edu.cn

Runsheng Chen[†]
Bioinformatics Laboratory and
National Laboratory of Biomacromolecules,
Institute of Biophysics, Chinese Academy of Sciences,
Beijing 100101, China,
Email: crs@sun5.ibp.ac.cn

Lei Xu[†,*]
Department of Computer Science and Engineering,
The Chinese University of Hong Kong, and
Speech and Hearing Research Center, Key Lab of
Machine Perception (Ministry of Education), Peking University
Email: lxu@cse.cuhk.edu.hk

Abstract—Transcription factor activities (TFAs), rather than expression levels, control gene expression and provide valuable information for investigating TF-gene regulations. Network Component Analysis (NCA) is a model based method to deduce TFAs and TF-gene control strengths from microarray data and *a priori* TF-gene connectivity data. We modify NCA to model gene expression regulation by non-Gaussian Factor Analysis (NFA), which assumes TFAs independently comes from Gaussian mixture densities. We properly incorporate *a priori* connectivity and/or sparsity on the mixing matrix of NFA, and derive, under Bayesian Ying-Yang (BYY) learning framework, a BYY-NFA algorithm that can not only uncover the latent TFA profile similar to NCA, but also is capable of automatically shutting off unnecessary connections. Simulation study demonstrates the effectiveness of BYY-NFA, and a preliminary application to two real world data sets shows that BYY-NFA improves NCA for the case when TF-gene connectivity is not available or not reliable, and may provide a preliminary set of candidate TF-gene interactions or double check unreliable connections for experimental verification.

Index Terms—transcription factor activity, network component analysis, non-Gaussian factor analysis, sparse learning, Bayesian Ying-Yang

I. INTRODUCTION

High-dimensional data from DNA microarray are typically controlled by low-dimensional regulatory signals through an interacting network [1]. The gene expression is controlled by one transcription factor (TF) alone or several TFs in combination on the target promoter regions. Transcription factor activities (TFA) rather than levels of transcription factor expression play roles in transcriptional regulations. It is a challenging problem in system biology [1] to reconstruct the dynamics of the hidden regulatory signals of TFs from the transcript levels of the genes they control.

† These authors contributed equally to this work.
∗ Corresponding Author: lxu@cse.cuhk.edu.hk

Network component analysis (NCA) [1], [2] is a popular method to deduce TFA and TF-gene regulation control strengths from transcriptome data and a priori network structure information which is usually constructed from ChIP-chip binding assays. In NCA, the (relative) gene expression is formulated as the product of each (relative) TFA to the power of the control strength from that TF to the gene, or equivalently, the relationship between TFA and gene expression is approximated as a log-linear model, where the log expression level of a gene is modeled by a mixture of log TFAs weighted by the control strengths. NCA takes the available network topology into account to compute the component TFAs, different from principal component analysis (PCA) or independent component analysis (ICA) which are based on statistical properties that may hinder the biological interpretations [1]. In NCA, the network topology needs to satisfy three criteria for a unique decomposition [3]. NCA has been successfully applied to determine the transcription regulatory activities, on microarray data generated from *Saccharamyces cerevisiae* during cell-cycle process or *Escherichia coli* carbon source transition from glucose to acetate [1], [4], and so on. Other related methods include REDUCE [5] that takes normalized motif binding copy number as the regulation strength of the TF for that gene, and obtains TFA profiles from gene expression by linear regression.

Based on the NCA framework, this paper modifies NCA to model gene transcriptional regulation by non-Gaussian Factor Analysis (NFA) [6], which assumes the latent TF activities come from independent Gaussian Mixture Models (GMM). From the perspective of NFA, the activity of a TF consists of two parts of information, i.e., a state to indicate the activity level and a Gaussian distribution to capture the stochasticity or noise variance of the state. NFA can be regarded as a generalization of ICA by relaxing the noise-free assumption.

NCA needs a priori knowledge about the connection topology of the TF-gene regulatory network. However, in most organisms, the connectivity information is largely unavailable. Due to the noise in experiments, the known connectivity data may not be reliable. A two-stage method that integrates NCA with stepwise regression was proposed to trim the network with the help of a modified Akaike information criterion [7]. To avoid such a repeated implementation, we incorporate a sparsity penalty on the mixing matrix of control strengths, so as to automatically push the extra entries to zero if there is not enough evidence for the existence of connections. Efforts on sparsity have been made on sparse learning or Lasso shrinkage by L1 norm penalty or a Laplacian prior, and so on [8], [9]. Within the Bayesian paradigm, we consider on each entry of the mixing matrix a joint Normal-Jeffreys prior distribution which has been shown to implement sparsity well without any hyper-parameters to be determined [10]. We propose to implement NFA under Bayesian Ying-Yang (BYY) [11], [6] learning framework. The derived BYY-NFA algorithm can utilize both a priori connectivity and sparsity on the matrix of control strengths.

To validate our proposed algorithm, experiments are conducted on both synthetic data and real data. Simulation study demonstrates the effectiveness of sparse BYY-NFA in recovering the hidden dynamics of TF regulatory signals, and in estimating the connectivity topology and control strengths. Moreover, the sparse BYY-NFA algorithm is applied on the yeast cell cycle data [12]. If using the available connectivity information but not considering sparsity, the reconstructed TFAs by BYY-NFA are similar to those by NCA. If sparsity is further incorporated, many connections are shut off as the their regulation strengths are pushed towards zeros, but without changing the cyclic patterns of TFAs. This suggest that some connections are not necessary or may be false positive. If considering the whole mixing matrix without knowledge about the network structure, the problem becomes difficult. Still, the connectivity and the regulatory dynamics can be inferred to some extent. Most connections are shut off, and approximately half the obtained connections are consistent with experimental ChIP-chip data. Therefore, our algorithm improves the NCA framework to the case when connectivity topology of the TF-gene network is not available or not reliable. The sparse BYY-NFA algorithm may provide a preliminary set of candidate TF-gene interactions, or double check the known connections for further experimental investigations. In addition, the proposed algorithm can also detect activations of involved TFAs during E. coli carbon source transition from glucose to acetate [4].

The rest of this paper is organized as follows. In Section II, we provide a brief review on NCA, based on which we proceed to introduce NFA for gene regulation modeling, with a BYY-NFA algorithm developed under BYY learning to implement sparse learning. Section III is devoted to experiments on synthetic data and real data, to validate the performance of BYY-NFA. Finally, conclusion is made in Section IV.

II. METHODS

A. A brief review on NCA

NCA approximates gene expression as the product of the contribution of each TF regulatory activity using a the following model [1]:

$$\frac{E_i(t)}{E_i(0)} = \prod_{j=1}^{m} \left(\frac{TFA_j(t)}{TFA_j(0)} \right)^{CS_{ij}}, \tag{1}$$

or equivalently a log-linear model in canonical matrix form:

$$X = AY + \Gamma, \tag{2}$$

where $E_i(t)$ is the gene expression level, $TFA_j(t)$ is the activity of the TF j, and CS_{ij} represents the control strength of TF j on gene i, and $X = [\log(E_i(t)/E_i(0))]_{n \times N}$, $Y = [\log(TFA_j(t)/TFA_j(0))]_{m \times N}$, $A = [CS_{ij}]_{n \times m}$, and Γ is the residual.

NCA is to minimize the residual Γ to get a decomposition $X \approx \hat{A}\hat{Y}$, from the observed gene expression profile X and the known connectivity for A, i.e., C_{ij} is fixed at zero if TF j does not bind to the promoter region of gene i, otherwise the control strength is non-zero. To guarantee a unique decomposition up to some normalization factors, NCA requires A and the resultant connectivity matrix by removing a regulatory node together with its connected neighbor genes to have full-column rank, and requires Y to have full-row rank. It should be noted that the NCA solution does not assume any relationship between the TFAs.

B. non-Gaussian Factor Analysis

The non-Gaussian Factor Analysis (NFA) [13] generalizes classical Factor Analysis (FA) by assuming that each hidden factor follows from a non-Gaussian distribution, and it also relaxes the noise-free assumption of Independent Component Analysis (ICA) [14], for blind separation and deconvolution of noisy signals [15].

The NFA model assumes that an n-dimensional observation $\boldsymbol{x}$ is formed by n mixtures of m independent non-Gaussian factors $y_1, \ldots, y_m$ and added with a Gaussian noise $\mathbf{e}$, i.e.,

$$\boldsymbol{x} = A\boldsymbol{y} + \mathbf{a}_0 + \mathbf{e}, \tag{3}$$

where A is an $n \times m$ mixing matrix, and

$$q(\boldsymbol{y}) = \prod_{j=1}^{m} q(y_j), \tag{4}$$

$$q(\boldsymbol{x}|\boldsymbol{y}) = G(\boldsymbol{x}|A\boldsymbol{y} + \mathbf{a}_0, \Sigma_e), \tag{5}$$

and Σ_e is a diagonal covariance matrix, and $G(z|\mu, \Sigma)$ denotes a Gaussian probability density with mean μ and covariance Σ. Each non-Gaussian factor y_j is modeled by a Gaussian mixture:

$$q(y_j) = \sum_{\ell=1}^{k_j} \alpha_{j\ell} G(y_j|\mu_{j\ell}, \sigma_{j\ell}^2), \tag{6}$$

where $0 \leq \alpha_{j\ell} \leq 1$, and $\sum_{\ell=1}^{k_j} \alpha_{j\ell} = 1$.

According to Eq.(2)&(3), we use NFA to generalize the framework of network component analysis (NCA) [1] by considering the mixing matrix A as the connectivity matrix between the transcription factors (TF) and the genes, the latent factor $\boldsymbol{y}$ to represent the TF activity, and the observation $\boldsymbol{x}$ being the gene expression.

C. Sparsity on the mixing matrix

As indicated by the literature [1], the number of target genes of a TF is usually small, and thus the connectivity matrix A is sparse, i.e., with many zero entries. From some TFs, the topology of connectivity is known based on a priori knowledge from biological experiments or computational predictions. Then, this case falls in the general framework of semi-blind factor analysis, e.g., see Sect.4.3 in [16]. However, the connectivity information may be not available or at least partially unknown in practice. One alternative way is to assume a full connectivity matrix, and then deactivate those false positive connections by automatically shrinking extra entries to zero during learning with the help of a sparseness constraint. Efforts on sparsity have been made on sparse learning or Lasso shrinkage by L1 norm penalty or a Laplacian prior, and so on [8], [9]. This paper adopts the following joint Normal-Jeffreys probability density on each entry of the mixing matrix [10]:

$$q(A|\boldsymbol{\gamma})q(\boldsymbol{\gamma}) \propto \prod_{i,j} G(a_{ij}|0, \gamma_{ij}) \cdot \prod_{i,j} \frac{1}{\gamma_{ij}}, \qquad (7)$$

where γ_{ij} is the variance of a_{ij} and is further controlled by a Jeffreys prior without any hyper-parameters.

D. Implementing sparse NFA under BYY framework

Systematically developed over a decade [11], [17], Bayesian Ying-Yang (BYY) harmony learning is a general statistical learning framework for parameter learning and model selection under a best harmony principle. For the above NFA model, the harmony measure is as follows:

$$H(p\|q) = \int \sum_L p(X)p(\boldsymbol{\Theta}|X)p(Y,L|X,\boldsymbol{\Theta}) \ln[$$
$$q(X|Y,L,\boldsymbol{\Theta})q(Y|L,\boldsymbol{\Theta})q(L|\boldsymbol{\Theta})q(\boldsymbol{\Theta})]dYdXd\boldsymbol{\Theta}, \quad (8)$$

where $q(\cdot)$ gives the Ying representation, and $p(\cdot)$ gives the Yang representation. All components in Ying representation follow from the above specifications in Eq.(4)-(7). In Yang representation, the empirical density $p(X) = \delta(X - X_N)$ is adopted with $X_N = \{\boldsymbol{x}_t\}_{t=1}^N$, and the other components are free, i.e., no constraints on their probability density functions.

Such a derived algorithm to maximize $H(p\|q)$ is called BYY-NFA, and the details are referred to Tab. VI in the Appendix. Sparse learning on the mixing matrix A is activated by $q(\boldsymbol{\Theta}) = q(A|\boldsymbol{\gamma})q(\boldsymbol{\gamma})$, or is shut off by $q(\boldsymbol{\Theta}) = 1$. Moreover, the available connectivity data can be utilized by fixing the corresponding entries of A at zero if there is no given connection, or by a confidence probability for a flexible incorporation [16]. The obtained combinations of implementations of BYY-NFA are summarized in Tab. I. With

TABLE I
IMPLEMENTATION OF ALGORITHMS. "C": CONSTRAINING A WITH CONNECTIVITY DATA FROM E.G., CHIP-CHIP ASSAY; "F": USING A FULL A (NO A PRIORI CONNECTIVITY); "N"/"S": IMPLEMENTING BYY-NFA WITHOUT/WITH SPARSE PRIOR.

algorithm	explanation
NCA	implemented by NCA toolbox [1]
BYY-NFA(n+c)	constraining A as NCA, $\tau = 0$ in Tab. VI
BYY-NFA(s+c)	constraining A as NCA, $\tau = 1$ in Tab. VI
BYY-NFA(s+f)	$\widetilde{A} = \mathbf{1}_{n \times m}$ (a matrix of all ones), $\tau = 1$ in Tab. VI
BYY-NFA(n+f)	$\widetilde{A} = \mathbf{1}_{n \times m}$, $\tau = 0$ in Tab. VI

the matrix A completely unknown to be estimated and also no a priori $q(\boldsymbol{\Theta})$ in consideration, the problem becomes learning a standard NFA, and the last one BYY-NFA(n+f) is actually the BYY-NFA proposed in Sect. 5 of [18] and Sect. IV(C) in [6], which is extended here by considering either or both of a priori connectivity and a priori distribution $q(\boldsymbol{\Theta})$.

The above is a brief introduction to BYY based on NFA. Readers are referred to not only a summary of nine aspects on the novelty and favorable natures of BYY best harmony learning, made at the end of Sect. 4.1 in [17], but also the roadmap shown in Fig. A2 of in [17], as well as to a systematic outline on the 13 topics about best harmony learning in Sect. 7 of [19].

III. RESULTS

A. On simulated data

First, we demonstrate the effectiveness of the proposed algorithm by simulated data sets. We set $n = 6, m = 2, k_1 = k_2 = 2, \Sigma_e = 0.1\mathbf{I}_n, \mathbf{a}_0 = \mathbf{0}$. The mixing matrix A is randomly generated and then randomly set $p_s = 40\%$ percentage of the entries to be zero. The final mixing matrix is given by A_o in Eq.(9). The factor distributions $q(y_r)$ are given in Fig. 1. Then, a synthetic data set X_N of sample size $N = 100$ is randomly generated according to the NFA model given in Eq.(3)-(5).

Given the data set X_N, the learning task consists of inferencing the activities of the factors $\{\mathbf{y}_t\}$, estimating the parameters $\boldsymbol{\Theta} = \{A, \Sigma_e, \mathbf{a}_0, \alpha_{rj_r}, \mu_{rj_r}, \sigma^2_{rj_r}\}$. The BYY-NFA(s+f) is implemented by randomly initializing $\boldsymbol{\Theta}$. The latent factor activities $\{\mathbf{y}_t\}$ are accurately estimated as shown in Fig. 2. Given in Eq.(11), the estimated mixing/connectivity matrix $\hat{A}_{sf}$ not only approximates non-zero connection strengths of the original one A_o well, but also pushes the elements to zero where there are no connections between the factors and the observation dimensions with the help of the imposed sparsity constraint by Eq.(7). If Eq.(7) is not imposed on A, the obtained $\hat{A}_{nf}$ in Eq.(10) by BYY-NFA(n+f) is not so sparse as $\hat{A}_{sf}$.
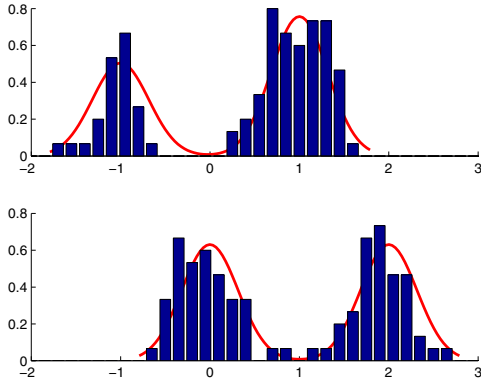
Fig. 1. The true 2-component GMM distributions (in red curves) and the histograms of the synthetic data of sample size $N = 100$ (in blue bars) of the two factors, $q(y_1)$ in the above and $q(y_2)$ at the bottom.
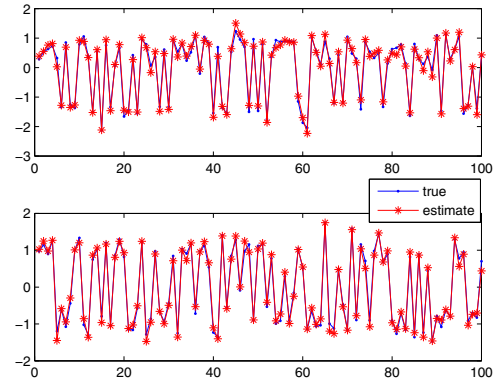

Fig. 2. The true activities (in blue) and the estimated ones (in red) of the latent factors. (Noted that some blue points are now shown evidently because they are covered by the red points.)

$$A_o = \begin{pmatrix} -0.6440 & 0 \\ -1.1045 & -1.2155 \\ -1.2183 & 0 \\ 2.1293 & -0.0369 \\ 0 & -1.0647 \\ 0 & -1.0775 \end{pmatrix} \quad (9)$$

$$\hat{A}_{nf} = \begin{pmatrix} -0.6113 & 0.1076 \\ -1.1072 & -1.0784 \\ -1.1792 & 0.1413 \\ 2.0782 & -0.2382 \\ 0.0195 & -1.0181 \\ -0.0027 & -1.0197 \end{pmatrix} \quad (10)$$

$$\hat{A}_{sf} = \begin{pmatrix} -0.5954 & 0.0096 \\ -1.1003 & -1.1069 \\ -1.1676 & 0.0001 \\ 2.0578 & 0.0019 \\ 0.0092 & -0.9290 \\ -0.0010 & -0.9436 \end{pmatrix} \quad (11)$$

B. On real data

1) Yeast cell-cycle data: In this study, we apply our algorithm to microarray data sets that are about yeast cell-cycle regulation. The data were taken from wild-type S. cerevisiae cultures synchronized by three independent methods, α-factor arrest, elutriation, and arrest of a *cdc15* and temperature-sensitive mutant [12], as well as *cdc28* data [20]. The connectivity information between transcription factors and their regulated genes comes from the genomewide location or ChIP-chip assay [21].

We focus on 6 transcription factors (TF) that are known to be related to cell-cycle regulation [21], [1]. Based on the "NCA Toolbox" [1], 137 genes regulated by these TFs were selected, with the connectivity information obtained from [21].

TABLE II
YEAST CELL CYCLE DATA POINTS IN COMBINED DATASET TIME POINTS
(MIN)

data	time-points	experiment
1-4	cln3-1, cln3-2, clb2-2, clb2-1	Cln3 and Clb2
5-22	0,7,14,21,...,119	α-factor
23-46	10,30,50,70,80,90,...,240,250,270,290	*cdc15*
47-63	0,10,20,...,150,160	*cdc28*
64-77	0,30,60,...,360,390	elutriation

First, we implement the BYY-NFA(n+c) (i.e., without utilizing the sparse prior). The elements of the mixing matrix A are fixed at zeros where there is no connection indicated by the connectivity data, in the same way as the NCA algorithm does. The reconstruction mean square errors (MSE), i.e.,

$$MSE = \frac{1}{nN} \sum_{t=1}^{N} \|\boldsymbol{x}_t - \hat{A}\hat{\boldsymbol{y}}_t - \hat{\mathbf{a}}_0\|^2 \quad (12)$$

are given in Tab. V, where $\boldsymbol{x}_t$ is the gene expression vector at time t, and $\hat{A}, \hat{\boldsymbol{y}}_t, \hat{\mathbf{a}}_0$ are connectivity matrix, TF activities and mean vector, estimated by NCA or BYY-NFA(n+c). The two algorithms both reconstruct the microarray expression data with small errors. The estimated regulatory activities of TFs are presented in Fig. 3&4. It should be noted that both NCA and BYY-NFA can only reconstruct TFA up to a normalization factor, and thus for an easy comparison, the TFA profile by BYY-NFA is normalized in the same way as NCA does in "NCA Toolbox" [1], [6]. Similar to NCA results, the dynamics of the TF activities estimated by the BYY-NFA(n+c) also show cyclic behavior, which reveals the role of each TF during cell cycle regulation. The reconstructed regulatory signals by NCA and by BYY-NFA(n+c) are very similar with a high average correlation coefficient (CC) 0.9781 of the six TFs.

Second, we implement BYY-NFA(s+c), still constraining A according to the connectivity data. The estimated TFA profile is given in Fig. 5, which resembles Fig. 3 for most TFs except FKH1. According to Tab. IV, 130 of 203 connections from the

TABLE III
RECONSTRUCTION MEAN SQUARE ERRORS (MSE) OF NCA AND
BYY-NFA ON YEAST CELL-CYCLE DATA. NCA AND BYY-NFA ARE
IMPLEMENTED ACCORDING TO TAB. I.

algorithm	MSE(n+c)	MSE(s+c)	MSE(s+f)
NCA	0.1320	-	-
BYY-NFA	0.1311	0.1373	0.0951

Fig. 3. The estimated TFA profile by NCA

Fig. 4. The estimated TFA profile by BYY-NFA(n+c). The average of the correlation coefficients of the TFAs in this profile to the corresponding ones in Fig. 3 is 0.9781.

Fig. 5. The estimated TFA profile by BYY-NFA(s+c). The average of the correlation coefficients of the TFAs in this profile to the corresponding ones in Fig. 3 is 0.8875.

Fig. 6. The estimated TFA profile by BYY-NFA(s+f). The average of the correlation coefficients of the TFAs in this profile to the corresponding ones in Fig. 3 is 0.7240.

ChIP-chip assay were shut off by the sparsity during learning, which indicates the 130 connections might be false positive or a mechanism of redundancy to ensure the robustness [22].

Third, we implement BYY-NFA(s+f) with all entries of A being variables. Table IV shows that a large part of the relaxed connections are switched off, and about half of the remaining connections are consistent with the ChIP-chip experiment. As in Fig. 6, most of TFAs still preserve the cyclic pattern and resemble to the corresponding one in Fig. 6, except ACE2 which does not have evident cyclic pattern as in Fig. 6 and thus decreases the average correlation coefficient to 0.72240.

Finally, we test the sensitivity of TFA profile to random insertions in the connectivity from ChIP-chip data, as in [23]. The available connectivity is altered by randomly adding an extra percentage P_{in} of connections to the mixing matrix A. Using the TFA profile by NCA in Fig. 3 as a reference, the correlation coefficients (CC) between the estimated TFA profiles to the reference are calculated and averaged in Tab. V for 10 independent runs for each $P_{in} \in \{0\%, 10\%, 30\%, 50\%, 100\%\}$. A high CC value indicates a strong resemblance to the reference TFA profile in Fig. 3. Both NCA and BYY-NFA(n+c) are sensitive to $P_{in} > 50\%$, while BYY-NFA(s+c) shows differences for no insertions (in Fig. 5) but keeps the CC values more constantly around 80% with the cyclic patterns preserved.

2) E. coli carbon source transition data: We further apply BYY-NFA to temporal gene expression profiles of *E. coli* during transition from glucose to acetate, with samples taken at $5, 15, 30, 60$ min and every hour until 6 h after transition [4]. Similar to [4], the repeated data points are averaged, and we demonstrate the effectiveness of BYY-NFA on 5 of 16
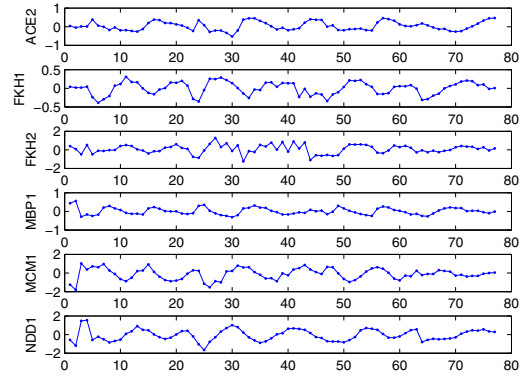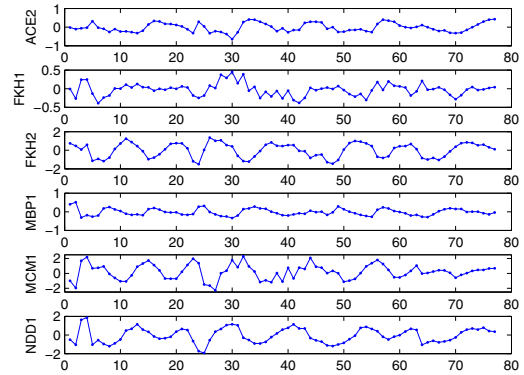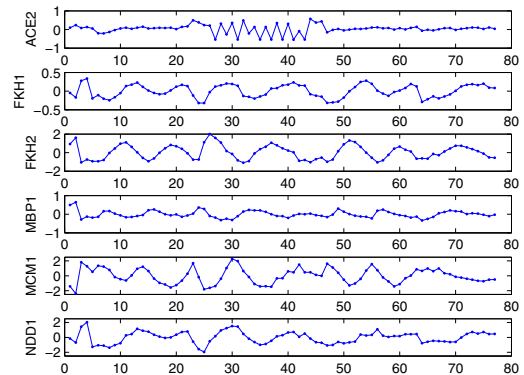
TFs that were analyzed by NCA in [4]. Based on the available connectivity data and the applicability for NCA, 22 genes were

TABLE IV
CONFUSION MATRICES OF THE RECONSTRUCTED CONNECTIVITY BY
SPARSE BYY-NFA ON YEAST CELL-CYCLE DATA AGAINST THE KNOWN
CONNECTIVITY FROM AVAILABLE EXPERIMENTS. FOR SPARSE BYY-NFA,
A CONNECTION IS DETERMINED TO BE PRESENT IF THE ABSOLUTE VALUE
OF ITS ESTIMATED CONTROL STRENGTH IS LARGER THAN A THRESHOLD
0.02. NOTATIONS: $\hat{0}$ OR $\hat{1}$ DENOTES RECONSTRUCTION, 0^* OR 1^*
DENOTES THE KNOWN CONNECTIVITY.

		0^*	1^*
BYY-NFA(s+c)	$\hat{0}$	0	130
	$\hat{1}$	0	73
		0^*	1^*
BYY-NFA(s+f)	$\hat{0}$	572	154
	$\hat{1}$	47	49

TABLE V
SENSITIVITY OF TFA PROFILES TO RANDOMLY INSERTING CONNECTIONS
IN CONNECTIVITY FROM CHIP-CHIP DATA. THE AVERAGE CORRELATION
COEFFICIENT FOR EACH CASE IS CALCULATED.

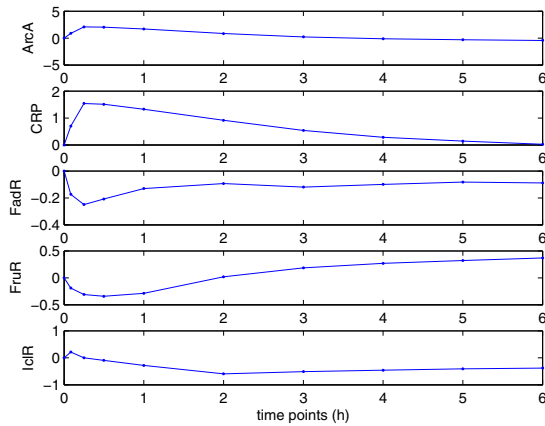| algorithm	0%	10%	30%	50%	70%	100%
NCA	1.0	0.975	0.937	0.913	0.778	0.665
BYY-NFA(n+c)	0.978	0.954	0.943	0.904	0.794	0.781
BYY-NFA(s+c)	0.888	0.899	0.873	0.833	0.765	0.732

Fig. 7. The estimated TFA profiles on E.Coli data by NCA with MSE 0.0117.

selected by the NCA toolbox [1], and there were 30 regulations from TFs to genes.

Similar to the patterns of the reconstructed TFA profile by NCA in Fig. 7, the TFAs by BYY-NFA(s+c) in Fig. 8 show activation immediate after transition and then gradually become stable, corresponding to the adaptation of cells to the new environmental condition. Moreover, 2 of 30 regulation strengths are pushed to zero, i.e., the corresponding regulations are shut off by BYY-NFA. Similar to Fig. 4-6, the results by other implementations of BYY-NFA can also be computed.

IV. CONCLUSION

Based on the NCA framework, we modify NCA to model the gene transcriptional regulation by NFA, which assumes the latent regulatory dynamics independently come from GMM
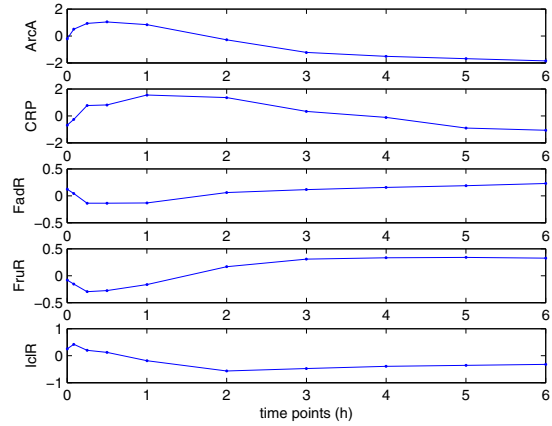

Fig. 8. The estimated TFA profiles on E.Coli data by BYY-NFA(s+c) with MSE 0.0077 and average CC 0.8926, which is calculated with respect to the corresponding ones in Fig. 7.

densities. Formulated within Bayesian paradigm, NFA can properly incorporate the *a priori* knowledge from experiments and the sparseness feature on the connectivity matrix via a Normal-Jeffreys prior. Synthetic experiments have demonstrated the effectiveness of our derived sparse BYY-NFA algorithm in uncovering the latent TFA profile, estimating the control strengths. The extra connections are shut off by pushing the corresponding control strengths to zeros due to the sparseness feature. Moreover, a preliminary application to *Saccharomyces cerevisiae* cell cycle data and *Escherichia coli* carbon source transition data shows that BYY-NFA not only reconstructs the hidden TF regulatory signals comparably as NCA does, but also improves NCA to be capable of shutting off unnecessary or unreliable connections. This property may be useful to provide candidate TF-gene interactions for biological experiments or to double check those unreliable interactions for further experimental investigations.

APPENDIX

In Yang representation, the empirical density $p(X) = \delta(X - X_N)$ is adopted with $X_N = \{x_t\}$, and all the other components are free, i.e., no constraints on their probability functions. In such a setting, the best harmony, i.e, maximizing $H(p\|q)$, leads the unknown Yang components to be Dirac delta functions, $p(\Theta|X) = \delta(\Theta - \Theta^*)$ and $p(y_t, \mathbf{j}|x_t) = \delta([y, \mathbf{j}] - [y_t^*, \mathbf{j}_t^*])$. It follows that Eq.(8) becomes

$$H(p\|q) \approx \sum_{t=1}^{N} \mathcal{L}(x_t, y_t, \mathbf{j}|\Theta) + \tau \ln q(A|\gamma)q(\gamma), \quad (13)$$

$$\mathcal{L} = \ln[G(x_t|Ay_t + \mathbf{a}_0, \Sigma_e)\alpha_{\mathbf{j}^*}G(y_t^*|\mu_{\mathbf{j}^*}, \Lambda_{\mathbf{j}^*})],$$

$$[y_t^*, \mathbf{j}_t^*] = \arg\max_{y_t, \mathbf{j}_t} \mathcal{L}(x_t, y_t, \mathbf{j}|\Theta), \quad (14)$$

where $\mathbf{j} = [j_1, \ldots, j_m]$ with $q(y_r|j_r) = G(y_r|\mu_{rj_r}, \sigma_{rj_r}^2)$, and $q(y_t|\mathbf{j})$ is a multivariate Gaussian density with mean $\mu_{\mathbf{j}} = [\mu_{1,j_1}, \ldots, \mu_{m,j_m}]$ and covariance matrix $\Lambda = diag[\sigma_{1,j_1}^2, \ldots, \sigma_{m,j_m}^2]$, and $diag[\mathbf{u}]$ denotes a diagonal matrix

with the vector $\mathbf{u}$ as its diagonal, and $q(\mathbf{j}) = \boldsymbol{\alpha}_{\mathbf{j}} = \prod_{r=1}^{m} \alpha_{rj_r}$, and τ is an indicator of whether the last term in Eq.(13) takes effects. Specifically, $\tau = 0$ shuts off the effects from the last term.

A Ying-Yang alternative procedure is implemented between "Step Y" and "Step $\boldsymbol{\theta}$ with details given in Tab. VI. The obtained algorithm is called BYY-NFA. It should be noted that $\tau = 0$ removes the sparsity constraint in BYY-NFA, which degenerates back to the one in Sect. 5 of [18] and Sect. IV(C) in [6]. The binary matrix $\widetilde{A}$ is used to incorporate known connectivity data, and it is set to be a matrix of all ones when no connectivity is available.

TABLE VI
THE BYY-NFA ALGORITHM

Objective: maximize $H(p\|q)$ by Eq.(13)
Initialization: initialize Θ, m, $k_1, \ldots, k_m$.
Step Y: get $\boldsymbol{y}_t^*$ and $\boldsymbol{j}_t^* = [j_{1t}^*, \ldots, j_{mt}^*]$, for $t = 1, \ldots, N$. $\quad \boldsymbol{y}_t^*(\boldsymbol{x}_t\|\mathbf{j}_t) = Q_{\mathbf{j}_t}^{-1}[A^T \Sigma_e^{-1}(\boldsymbol{x}_t - \boldsymbol{a}_0) + \Lambda_{\mathbf{j}_t}^{-1}\boldsymbol{\mu}_{\mathbf{j}_t}]$, $\quad \boldsymbol{j}_t^* = \arg\max_{\mathbf{j} \in \mathcal{M}} \ln[G(\boldsymbol{x}_t\|A\boldsymbol{y}_t^* + \boldsymbol{a}_0, \Sigma_e)\alpha_{\mathbf{j}}G(\boldsymbol{y}_t^*\|\boldsymbol{\mu}_{\mathbf{j}}, \Lambda_{\mathbf{j}})]$, $\quad$ where $\mathbf{j} = [j_1, \ldots, j_m]$, $\boldsymbol{\alpha}_{\mathbf{j}} = \prod_{r=1}^{m}\alpha_{r,j_r}$, $\quad$ and $\boldsymbol{\mu}_{\mathbf{j}} = [\mu_{1,j_1}, \ldots, \mu_{m,j_m}]$, $Q_{\mathbf{j}_t} = \Lambda_{\mathbf{j}_t}^{-1} + A^T\Sigma_e^{-1}A$ $\quad$ and $\Lambda_{\mathbf{j}} = diag[\sigma_{1,j_1}^2, \ldots, \sigma_{m,j_m}^2]$, $\quad$ and $\mathcal{M} = \{\mathbf{j} \| j_r = 1, \ldots, k_r; 1 \le r \le m\}$.
Step $\boldsymbol{\theta}$: by gradient method $\boldsymbol{\theta}^{new} \leftarrow \boldsymbol{\theta}^{old} + \eta \cdot \partial\boldsymbol{\theta}$, $\quad \partial\boldsymbol{\theta} = \frac{\partial H(p\|q)}{\partial\boldsymbol{\theta}}\Big\|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{old}}, \forall \boldsymbol{\theta} \in \{A, \Sigma_e, \boldsymbol{a}_0, \alpha_{rj_r}, \mu_{rj_r}, \sigma_{rj_r}^2\}$. $\quad \boldsymbol{a}_0 \leftarrow \boldsymbol{a}_0 + \eta \cdot \sum_{t=1}^{N}\{\Sigma_e^{-1}\mathbf{e}_t\}, \quad \mathbf{e} = \boldsymbol{x}_t - A\boldsymbol{y}_t - \boldsymbol{a}_0$, $\quad A \leftarrow \left[A + \eta \cdot \left\{\sum_{t=1}^{N}\left\{\Sigma_e^{-1}\mathbf{e}_t\boldsymbol{y}_t^T\right\} + \tau \cdot B\right\}\right] \circ \widetilde{A}$, $\quad \Sigma_e \leftarrow \Sigma_e + \eta \cdot \sum_{t=1}^{N}\left\{-\frac{1}{2}\Sigma_e^{-1} - \frac{1}{2}\Sigma_e^{-1}\mathbf{e}_t\mathbf{e}_t^T\Sigma_e^{-1}\right\}$, $\quad \gamma_{ij} \leftarrow \gamma_{ij} + \eta \cdot \tau \cdot \left\{-\frac{3}{2\gamma_{ij}} + \frac{a_{ij}^2}{2\gamma_{ij}^3}\right\}, \quad \gamma_{ij} > 0$, $\quad \beta_{r\ell'} \leftarrow \beta_{r\ell'} + \eta \cdot \{\sum_{t=1}^{N} z_{r\ell't} - \alpha_{r\ell'}N\}$ $\quad \alpha_{rj_r} \leftarrow \exp\{\beta_{rj_r}\}/\sum_{\ell'=1}^{k_r}\exp\{\beta_{r\ell'}\}, \quad B = -a_{ij}/\gamma_{ij}$; $\quad \mu_{j\ell} \leftarrow \mu_{j\ell} + \eta \cdot \sum_{t=1}^{N}\left(y_{j\ell t}^z/\sum_{t=1}^{N} y_{j\ell t}^z\right)y_{jt}^r$, $\quad \sigma_{r\ell}^2 \leftarrow \sigma_{r\ell}^2 - \eta \cdot \frac{1}{2}\lambda_{r\ell}^{-2}\left[\lambda_{r\ell}(\sum_t z_{r\ell t}) - \sum_t z_{r\ell t}(y_{rt} - \mu_{r\ell})^2\right]$, $\quad$ where $z_{r\ell t} = 1$ if $\ell = j_r^*$, otherwise $z_{r\ell t} = 0, \forall \ell \in \{1, \ldots, k_r\}$, $\quad \alpha_{rj_r}$ is relaxed to updating βrj_r, and η is a small learning step size, and $\tau = 1, 0$ indicate whether or not sparsity is taken into account respectively, and $\widetilde{A} = [\widetilde{a}_{ij}]_{n \times m} \in \{0, 1\}^{n \times m}$ is specified according to the connectivity constraint, and the operator "$\circ$" denotes an element-wise product.
Convergence: repeat Step Y and Step Θ until $\mathcal{H} = H(p\|q)$ converges, i.e., $\|\mathcal{H}^{new} - \mathcal{H}^{old}\|/\|\mathcal{H}^{old}\| < 10^{-4}$.

ACKNOWLEDGMENT

REFERENCES

[1] J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury, "Network component analysis: Reconstruction of regulatory signals in biological systems," *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15 522–15 527, 2003.

[2] L. M. Tran, M. P. Brynildsen, K. C. Kao, J. K. Suen, and J. C. Liao, "gNCA: A framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation," *Metabolic Engineering*, vol. 7, no. 2, pp. 128 – 141, 2005.

[3] S. J. Galbraith, L. M. Tran, and J. C. Liao, "Transcriptome network component analysis with limited microarray data," *Bioinformatics*, vol. 22, no. 15, pp. 1886–1894, 2006.

[4] K. C. Kao, Y.-L. Yang, R. Boscolo, C. Sabatti, V. Roychowdhury, and J. C. Liao, "Transcriptome-based determination of multiple transcription regulator activities in Escherichia coli by using network component analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 2, pp. 641–646, 2004.

[5] H. J. Bussemaker, H. Li, and E. D. Siggia, "Regulatory element detection using correlation with expression," *Nature Genetics*, vol. 27, pp. 167–174, 2001.

[6] L. Xu, "Advances on BYY harmony learning: information theoretic perspective, generalized projection geometry, and independent factor autodetermination," *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 885–902, 2004.

[7] L. Tran, D. Hyduke, and J. Liao, "Trimming of mammalian transcriptional networks using network component analysis," *BMC Bioinformatics*, vol. 11, no. 1, pp. 511–521, 2010.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[9] P. M. Williams, "Bayesian regularisation and pruning using a laplace prior," *Neural Computation*, vol. 7, pp. 117–143, 1994.

[10] Y. Guan and J. G. Dy, "Sparse probabilistic principal component analysis," in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS'09)*, vol. 5, 2009, pp. 185–192.

[11] L. Xu, "Bayesian-Kullback coupled YING-YANG machines: unified learning and new results on vector quantization," in *Proceedings of International Conference on Neural Information Processing, Oct 30-Nov.3, Beijing, China*, 1995, pp. 977–988, (A further version in NIPS8, D.S. Touretzky, et al (Eds.), MIT press, 444-450).

[12] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher, "Comprehensive identification of cell cycleregulated genes of the yeast saccharomyces cerevisiae by microarray hybridization," *Molecular Biology of the Cell*, vol. 9, no. 12, pp. 3273–3297, 1998.

[13] L. Xu, "BYY harmony learning, independent state space, and generalized APT financial analyses," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 822–849, 2001.

[14] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, Wiley, 2001.

[15] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. ICASSP'97*, 1997, pp. 3617–3620.

[16] L. Xu, "Codimensional matrix pairing perspective of BYY harmony learning: hierarchy of bilinear systems, joint decomposition of datacovariance, and applications of network biology," *Frontiers of Electrical and Electronic Engineering in China*, vol. 6, pp. 86–119, 2011.

[17] ——, "Bayesian Ying-Yang System, Best Harmony Learning, and Five Action Circling," *A special issue on Emerging Themes on Information Theory and Bayesian Approach, Frontiers of Electrical and Electronic Engineering in China*, vol. 5, no. 3, pp. 281–328, 2010.

[18] ——, "Independent component analysis and extensions with noise and time: A Bayesian Ying-Yang learning perspective," in *Neural Information Processing Letters and Reviews*, 2003, pp. 1–52.

[19] ——, "On essential topics of BYY harmony learning: Current status, challenging issues, and gene analysis applications," *Frontiers of Electrical and Electronic Engineering in China (A special issue on Machine learning and intelligence science: IScIDE (C))*, vol. 7, pp. 147–196, 2012.

[20] R. J. Cho, M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. J. Lockhart, and R. W. Davis, "A genome-wide transcriptional analysis of the mitotic cell cycle," *Molecular Cell*, vol. 2, pp. 65–73, 1998.

[21] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young, "Transcriptional regulatory networks in saccharomyces cerevisiae," *Science*, vol. 298, no. 5594, pp. 799–804, 2002.

[22] L. MacNeil and A. J. Walhout, "Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression," *Genome Research*, vol. 21, pp. 645–657, 2011.

[23] Y.-L. Yang, J. Suen, M. P. Brynildsen, S. J. Galbraith, and J. C. Liao, "Inferring yeast cell cycle regulators and interactions using transcription factor activities," *BMC Genomics*, vol. 6, no. 90, 2005.