MultiFacTV: Finding Modules from Higher-order Gene Expression Profiles with Time Dimension

Xutao Li, Yunming Ye and Qingyao Wu
Department of Computer Science, Shenzhen Graduate School
Harbin Institute of Technology, Shenzhen, 518055, China
xutaolee08@gmail.com, yeyunming@hit.edu.cn,
wuqingyao.china@gmail.com

Michael K. Ng
Department of Mathematics
Hong Kong Baptist University
Kowloon Tong, Hong Kong, China
mng@math.hkbu.edu.hk

***Abstract*—Module detection is an important task in bioinformatics which aims at finding a set of cells/genes that interact together to be responsible for some biological functionalities. In this paper, we propose a novel tensor factorization approach to finding modules from higher-order gene expression profiles with the time dimension, e.g., gene $\times$ condition $\times$ time data. The main idea is to incorporate a total variation regularization term for the time dimension during the tensor factorization, and then use the factorization results to identify the modules. Experimental results on two real gene $\times$ condition $\times$ time datasets have shown the effectiveness of the proposed method.**

***Keywords*-tensor factorization, total variation, module detection, regularization, alternating directions method**

I. Introduction

Module detection is an important task in bioinformatics which aims at finding a set of cells/genes that interact together to be responsible for some biological functionalities. A number of techniques have been developed and studied for accomplishing such objective. For example, traditional clustering algorithms like k-means clustering [1], hierarchical clustering [2] and spectral clustering [3] were used to explore interesting modules. In 2000, Cheng et al. introduced the biclustering approach [4] to clustering the rows and columns of a gene-sample data matrix simultaneously.

Besides these methods, some researchers proposed to find the biological modules based on the matrix factorization techniques. For example, Wall et al. showed how singular vector decomposition can be used to cluster gene expression data in [5]. In 2010, Lee et al. proposed to incorporate sparsity constraints on singular vector decomposition for finding interesting subspace modules [6]. In 2011, Sill et al. improved this method by using stability selection technique such that more stable modules can be discovered [7].

Recently, detecting modules by integrative analysis have also begun to attract researchers' attentions because integrative analysis is able to find modules from higher-order gene expression data and tends to provide a new understanding of biological processes. In this setting, data are usually represented as a higher-order matrix, namely a tensor. For example, the data may be in the form of gene $\times$ gene $\times$ condition tensor, or gene $\times$ condition $\times$ time tensor. Li et al. proposed to identify [8] *recurrent heavy subgraphs* from many gene co-expression networks by using a tensor computation method. In [9], Omberg et al. analyzed DNA microarray data from different biological studies interactively by applying singular value decomposition techniques to the corresponding tensor data. Narayanan et al. proposed JointCluster algorithm [10] to find sets of genes that cluster well in multiple networks of interest. Supper et al. extended the 2D *iterative signature algorithm* and proposed EDISA algorithm to extract bi-clusters from gene $\times$ condition $\times$ time data [11].

In this paper, we focus on identifying modules from higher-order gene expression data in which time dimension is involved, namely, gene $\times$ condition $\times$ time data. Such tensor data can be obtained by accumulating different experiments from time series microarray databases [12], [13] or studying one time series dataset under different conditions [14], [15]. For example, the *Arabidopsis thaliana* and *Homo sapiens* are two typical datasets of this kind of setting. Arabidopsis thaliana dataset is composed of the expression profiles from a set of genes in the root and shoot tissues over serval time-points when different abiotic stress stimulus are conducted [15]. Homo sapiens dataset is composed of expression profiles from a set of genes for 14 sclerosis patients over 9 time-points after IFN-β injection [14]. Finding modules composed of some genes, conditions, over some specific time intervals from such high-order gene expression data, enables us to understand and analyze their corresponding biological processes in depth.

The main aim of this paper is to propose an algorithm, MultiFacTV, to identify modules from higher-order gene expression profiles with the time dimension. The basic idea of the algorithm is to apply tensor factorization to the higher-order gene expression data and then use the factorization results to find the modules. Since the time dimension is involved, it is desired that the modules consistently exist during some consecutive time periods. To achieve this property, we introduce a total variation regularization term in the

978-1-4673-2560-8/12/$31.00 ©2012 IEEE

factorization process. Experimental results on Arabidopsis thaliana and Homo sapiens datasets have shown the effectiveness of the proposed method.

The outline of this paper is organized as follows: In Section 2, we present the notations that will be used in this paper. In Section 3, the proposed MultiFacTV method is presented. In Section 4, we present and discuss the experimental results on Arabidopsis thaliana and Homo sapiens datasets. Finally, some concluding remarks are given in Section 5.

II. Preliminaries

Let $\|\cdot\|_F$ be the Frobenius norm of a matrix or a tensor, i.e., given a $m \times n$ matrix X,

$$\|X\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j}^2,$$

or given a $m \times n \times p$ tensor $\mathcal{T}$,

$$\|\mathcal{T}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p} t_{i,j,k}^2.$$

Let Tr$(\cdot)$ be the trace of a square matrix, i.e., given a $m \times m$ matrix X,

$$\mathrm{Tr}(X) = \sum_{i=1}^{m} x_{i,i}.$$

Let $\mathrm{shrinkage}_{\alpha/\rho}(\cdot)$ denote an elementwise shrinkage operator for a matrix, i.e., given a $m \times n$ matrix X,

$$\mathrm{shrinkage}_{\alpha/\rho}(X)_{i,j} = x_{i,j} - \min(\alpha/\rho, |x_{i,j}|) \cdot \frac{x_{i,j}}{|x_{i,j}|},$$

where $\frac{x_{i,j}}{|x_{i,j}|}$ should be zero when $x_{i,j} = 0$.

Let $\otimes$ denote the Kronecker product. Given two matrices $X \in R^{m\times n}$, $Y \in R^{p\times q}$, $X \otimes Y$ is a $mp \times nq$ matrix and

$$X \otimes Y = \begin{bmatrix} x_{1,1}Y & x_{1,2}Y & ... & x_{1,n}Y \\ x_{2,1}Y & x_{2,2}Y & ... & x_{2,n}Y \\ ... & ... & ... & ... \\ x_{m,1}Y & x_{m,2}Y & ... & x_{m,n}Y \end{bmatrix}.$$

Let $\odot$ indicate the Khatri-Rao product. Given two matrices $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k] \in R^{m\times k}$, $Y = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_k] \in R^{n\times k}$, $X \odot Y$ is a $mn \times k$ matrix and

$$X \odot Y = [\mathbf{x}_1 \otimes \mathbf{y}_1, \mathbf{x}_2 \otimes \mathbf{y}_2, ..., \mathbf{x}_k \otimes \mathbf{y}_k].$$

Let $\|\cdot\|_1$ be the l_1-norm of a vector, i.e., given a $m \times 1$ vector $\mathbf{a} \in R^{m\times 1}$,

$$\|\mathbf{a}\|_1 = \sum_{i=1}^{m} |a_i|.$$

Let $\circ$ denote the outer product. Given three vectors $\mathbf{a} \in R^{m\times 1}$, $\mathbf{b} \in R^{n\times 1}$, $\mathbf{c} \in R^{p\times 1}$, $\mathbf{a}\circ\mathbf{b}\circ\mathbf{c}$ is a $m \times n \times p$ tensor and

$$(\mathbf{a}\circ\mathbf{b}\circ\mathbf{c})_{i,j,k} = a_i b_j c_k.$$

For a $m \times n$ matrix X, let $X_{[i,j]}$ refer the values in the entry of the i-th row and the j-th column, and $X_{[:,k]}$ denote the k-th column. Let X^T indicate the transpose of the matrix X. Let $vec(X)$ indicate a $mn \times 1$ vector that is constructed by stacking the columns of the matrix X. Let $I_{p\times p}$ indicate the identity matrix of size $p \times p$.

III. The Proposed Method

In the following, we assume that the gene expression data is in the form of gene $\times$ condition $\times$ time. Let m, n and p be the number of genes, conditions and time-points respectively. Let the third order $(m \times n \times p)$ tensor $\mathcal{T} = (t_{i,j,k})$ represent the corresponding gene-condition-time interactions, where $t_{i,j,k}$ is the expression value of the i-th gene under the j-th condition at the k-th time-point.

In order to identify L modules, we factorize the tensor $\mathcal{T}$ into L decompositions by employing the following objective function:

$$\begin{aligned} &\min \left\| \mathcal{T} - \sum_{l=1}^{L} X_{[:,l]} \circ Y_{[:,l]} \circ Z_{[:,l]} \right\|_F^2 + \alpha \sum_{l=1}^{L} \left\| F Z_{[:,l]} \right\|_1 \\ &\text{s.t.} \quad X \geq 0, Y \geq 0, Z \geq 0 \\ &\text{and} \quad \sum_{i=1}^{p} Z_{[i,l]} = 1 \quad \text{for} \quad l = 1, 2, ..., L, \end{aligned} \tag{1}$$

where X, Y and Z are the decomposition matrices for genes, conditions and time-points with sizes $m\times L$, $n\times L$ and $p\times L$ respectively, F is a $(p-1)\times p$ matrix such that $F_{[i,i]} = 1$ and $F_{[i,i+1]} = -1$ for $i = 1, 2, ..., p-1$ and the other entries are zeros, and α is a nonnegative regularization parameter. Note that the second term of this objective function is a total variation regularization on each decomposition of time-points.

Next we derive the updating formulas for matrices X, Y and Z respectively. Assume the unfolding matrices of tensor $\mathcal{T}$ with respect to gene dimension (mode-1), condition dimension (mode-2) and time dimension (mode-3) are $P^{(1)}$, $P^{(2)}$ and $P^{(3)}$ respectively. In this case, $P^{(1)}$, $P^{(2)}$ and $P^{(3)}$ are matrices of sizes $m\times np$, $n\times mp$ and $p\times mn$ respectively. Then we have the following three subproblems:

Subproblem 1: Minimize the objective function by finding X with Y and Z fixed.

The objective function for computing X is then changed into:

$$\min \left\| P^{(1)} - XH \right\|_F^2 \tag{2}$$

where $H = (Z \odot Y)^T$. We have

$$X_{[i,j]} = X_{[i,j]} \frac{\left(P^{(1)}H^T\right)_{[i,j]}}{\left(XHH^T\right)_{[i,j]}} \tag{3}$$

Subproblem 2: Minimize the objective function by finding Y with X and Z fixed.

The objective function for computing Y is then changed into:

$$\min \left\| P^{(2)} - YH \right\|_F^2 \tag{4}$$

where $H = (Z \odot X)^T$. We have

$$Y_{[i,j]} = Y_{[i,j]} \frac{\left(P^{(2)} H^T\right)_{[i,j]}}{\left(YHH^T\right)_{[i,j]}} \tag{5}$$

Subproblem 3: Minimize the objective function by finding Z with X and Y fixed.

The objective function for computing Z is then changed into:

$$\begin{aligned} &\min \left\| P^{(3)} - ZH \right\|_F^2 + \alpha \sum_{l=1}^{L} \left\| FZ_{[:,l]} \right\|_1 \\ &\text{s.t.} \quad \sum_{i=1}^{p} Z_{[i,l]} = 1 \quad \text{for} \quad l = 1, 2, ..., L, \end{aligned} \tag{6}$$

where $H = (Y \odot X)^T$.

By introducing a $(p-1) \times L$ matrix V as an auxiliary variable, the subproblem can be rewritten as:

$$\begin{aligned} &\min \left\| P^{(3)} - ZH \right\|_F^2 + \alpha \sum_{l=1}^{L} \left\| V_{[:,l]} \right\|_1 \\ &\text{s.t.} \quad FZ - V = 0 \\ &\text{and} \quad \sum_{i=1}^{p} Z_{[i,l]} = 1 \quad \text{for} \quad l = 1, 2, ..., L, \end{aligned} \tag{7}$$

where $H = (Y \odot X)^T$. By employing the Alternating Direction Method of Multipliers (ADMM) [16], [17], we have the following augmented Lagrangian problem:

$$\begin{aligned} J_\rho(Z, V, \Gamma) \quad &= \left\| P^{(3)} - ZH \right\|_F^2 + \alpha \sum_{l=1}^{L} \left\| V_{[:,l]} \right\|_1 \\ &+ \mathrm{Tr}(\Gamma^T (FZ - V)) + \frac{\rho}{2} \left\| FZ - V \right\|^2 \end{aligned} \tag{8}$$

where $\rho > 0$ and Γ is a $(p-1) \times L$ matrix representing the set of Lagrangian multipliers. Then ADMM consists of the three iterations:

$$\begin{aligned} Z &= \arg\min_Z J_\rho(Z, V, \Gamma), \\ V &= \arg\min_V J_\rho(Z, V, \Gamma), \\ \Gamma &= \Gamma + \rho \left(FZ - V \right). \end{aligned}$$

Therefore, we derive the following three updating formulas:

$$\begin{aligned} vec(Z) \quad = \quad & \left(\left(HH^T \otimes 2I_{p \times p} \right) + \left(\rho I_{L \times L} \otimes F^T F \right) \right)^{-1} \\ & vec\left(\rho F^T \left(V - \Gamma/\rho \right) + 2P^{(3)} H^T \right) \end{aligned} \tag{9}$$

$$V = \mathrm{shinkage}_{\alpha/\rho}(FZ + \Gamma/\rho) \tag{10}$$

$$\Gamma = \Gamma + \rho \left(FZ - V \right) \tag{11}$$

We note Subproblem 3 needs to be solved in an iterative manner, i.e., we update matrices Z, V and Γ as equations (9), (10) and (11) iteratively until its convergence (We need to normalize each column of Z after computing Z as equation (10) in each step such that the constraints in (1) hold).

Table I
THE MULTIFACTV ALGORITHM

Input: the number of modules L, the regularization parameter α, threshold factors θ_1, θ_2, and θ_3

Output: L modules composed of genes, conditions and time points

Procedure:

1: Randomly initialize X, Y and Z;
2: Update X using equation (3);
3: Update Y using equation (5);
4: Randomly initialize V and Γ;
5: Update Z, V and Γ using equations (9), (10) and (11) iteratively until converged;
6: Check the convergence for X, Y and Z; if it is not converged, goto Step 2; otherwise goto Step 7;
7: For $l = 1$ to L, put the i-th gene into the l-th module if $X_{[i,l]} >= 0.5\theta_1((\max(X_{[:,l]})+\min(X_{[:,l]}))$; put the i-th condition into the l-th module if $Y_{[i,l]} >= 0.5\theta_2((\max(Y_{[:,l]})+\min(Y_{[:,l]}))$; put the i-th time point into the l-th module if $Z_{[i,l]} >= 0.5\theta_3((\max(Z_{[:,l]})+\min(Z_{[:,l]}))$.

By iteratively solving Subproblems 1, 2 and 3 as shown as in the above formulas, we can minimize the objective function in (1) and solve the decomposition matrices X, Y and Z. Then we extract the modules based on these three matrices. The proposed algorithm is summarized in Algorithm 1.

The algorithm has five parameters: the number of modules to be identified L, the regularization parameter α, and three parameters θ_1, θ_2, and θ_3 for extracting modules from the factorization results. From Step 2 to Step 6, the algorithm updates the factorization matrices X, Y and Z iteratively. Note that Step 5 itself is an iterative procedure for updating the matrix Z. After the computation of decomposition matrices X, Y and Z, the algorithm extracts the modules based on them and the input parameters θ_1, θ_2, and θ_3.

IV. EXPERIMENTAL RESULTS

In this section, we present the experimental results of the proposed method on Arabidopsis thaliana and Homo sapiens datasets to show its performance and usefulness. In all the experiments, we set the regularization parameter α to be 10. In Arabidopsis thaliana dataset, we set $\theta_1 = \theta_2 = 1.0$ and $\theta_3 = 0.75$. In Homo sapiens dataset, we set $\theta_1 = \theta_2 = 0.5$ and $\theta_3 = 0.75$. These parameters are tuned after some trial runs of the proposed MultiFacTV algorithm. The MultiFacTV algorithm was implemented with MATLAB.

A. Results on Arabidopsis Thaliana Root/Shoot Datasets

In this experiment, we perform MultiFacTV on the Arabidopsis root and shoot datasets to identify some interesting modules. Arabidopsis thaliana datasets are composed of the expression profiles of a set of genes in the root and shoot tissue over serval time-points when different abiotic stress stimulus are conducted[1]. We consider different abiotic stress stimulus as different conditions. After preprocessing the datasets, we construct a tensor $\mathcal{T}$ consisting of 2395 genes, 9 conditions and 6 time-points for Arabidopsis root tissue, and construct a tensor $\mathcal{T}$ consisting of 3454 genes, 8 conditions and 6 time-points for Arabidopsis shoot tissue. Since the ground truth modules are unknown, we set L to be 40 when performing MultiFacTV algorithm. It took MultiFacTV 308 seconds and 454 seconds on root tissue and shoot tissue datasets respectively. In the following, we show and analyze some interesting modules identified on these two datasets.

Interesting modules on Arabidopsis thaliana root: In order to evaluate each module, we map the genes in the module to the Gene Ontology with DAVID [18] and present p-values calculated by DAVID.

- Modules responding to cold-osmotic stress. We obtained three modules associated to both cold and osmotic stresses. Figure 1 shows the expression profiles of the genes in these three modules respectively. It can be seen from Figure 1(a) that the genes in the first module have distinct expression shapes under cold and osmotic conditions compared to under the other conditions. This suggests that those genes co-regulate under cold and osmotic conditions. This module is assigned to "response to water deprivation" (p-value 3.9×10^{-4}), "response to water" (p-value: 4.7×10^{-4}) "cold acclimation" (p-value: 7.2×10^{-4}) and "response to cold" (p-value: 1.1×10^{-3}). From Figure 1(b), we see that the genes in the second module up-regulate from 6h to 24h under cold and osmotic conditions. This module is significantly mapped to "response to water" (p-value: 7.5×10^{-6}) and "response to osmotic stress" (p-value: 9.0×10^{-4}), and it is mapped to "response to temperature stimulus" (p-value: 8.7×10^{-6}) and "response to cold" (p-value: 1.3×10^{-2}). Different from the second module, the genes in the third module down-regulate from 6h to 24h under cold and osmotic conditions, shown as in Figure 1(c). This module is associated to "response to cold" (p-value: 2.7×10^{-3}) and "response to water" (p-value: 3.7×10^{-3}). These facts indicate those three modules is closely related and reacted to cold and osmotic stress stimulus.
- Module responding to heat stress. We found one module that is closely associated to heat stress. The expression profiles of genes in this module are shown as in Figure 2. Similarly, we see that the expression profiles of those genes under heat condition are quite different from the ones under other conditions. This module is assigned to "response to heat" (p-value: 1.0×10^{-55}) and "response to temperature stimulus" (p-value: 1.2×10^{-43}). This fact shows that the genes in this module react significantly to heat stress.

[1]The dataset is available on http://www.ra.cs.uni-tuebingen.de/software/EDISA/downloads/index.htm

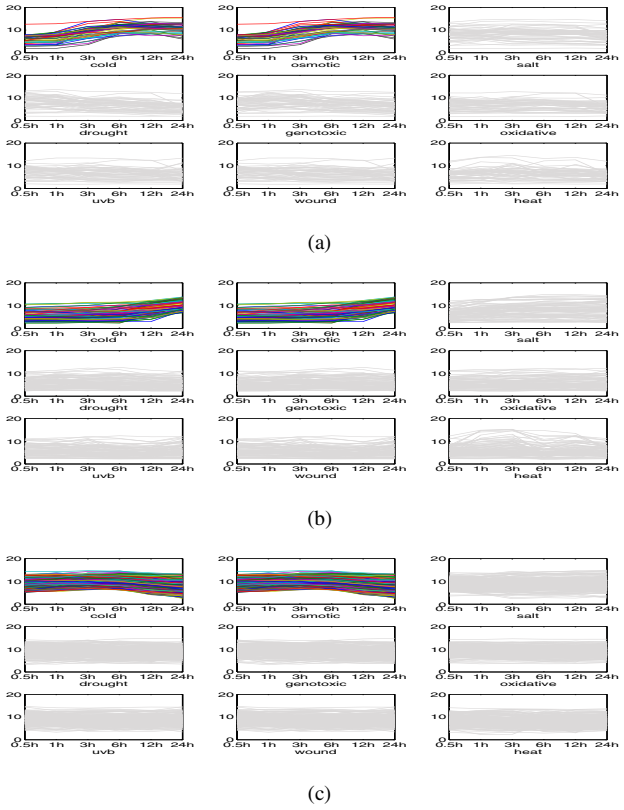

Figure 1. The expression profiles of genes in three cold-osmotic modules identified from root tissue respectively. (a) module 1; (b) module 2; (c) module 3.

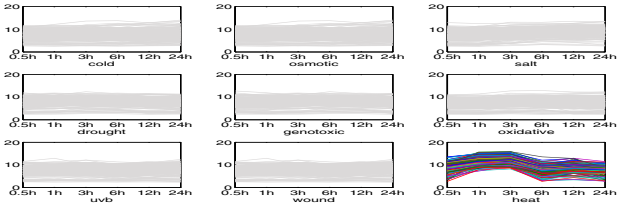

Figure 2. The expression profiles of genes in heat module identified from root tissue.

Interesting modules on Arabidopsis thaliana shoot: As some modules are similar to previous findings from root tissue, we only present some different ones below.

- Module responding to uvb-heat stress. We obtained one module that is associated to uvb and heat stress stimulus. Figure 3 shows the expression profiles of

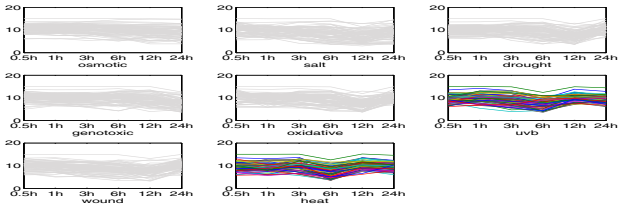

Figure 3. The expression profiles of genes in uvb-heat module identified from shoot tissue.

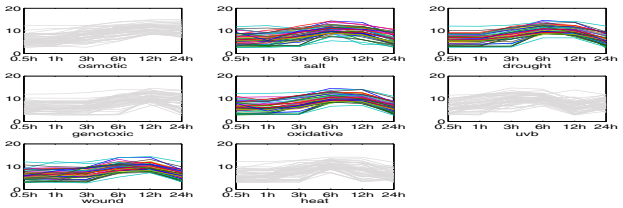

Figure 4. The expression profiles of genes in circadian rhythm module identified from shoot tissue. Note that the time-points 0.5h, 1h, 3h, 6h, 12h and 24h in the experiment correspond to 9:30, 10:00, 12:00, 15:00, 21:00 and 9:00 in the day time respectively.

genes in this module. We see from the figure that those genes down-regulate from 3h to 6h and then up-regulate from 6h to 12h under both uvb and heat conditions. This module is assigned to "response to light stimulus" (p-value: 4.9×10^{-4}) and "response to heat" (p-value: 5.8×10^{-3}) by DAVID. This fact indicates that the genes in this module co-regulate under uvb and heat stress stimulus.

- Module responding to circadian rhythm. We found a module responding to circadian rhythm, its expression profiles shown in Figure 4. We see from this figure that the set of genes in this module up-regulate from 12:00 to 21:00 (during the daytime) and down-regulate from 21:00 to 9:00 (during the night time) under salt, drought, oxidative and wound conditions. This module is mapped to "circadian rhythm" (p-value: 2.1×10^{-8}) and "rhythmic process" (p-value: 6.8×10^{-8}). These facts confirm that this module responses to circadian rhythm.

In summary, we have shown that MultiFacTV algorithm is able to identify interesting modules from Arabidopsis thaliana datasets. The functionalities of those modules have been partially verified by using Gene Ontology with DAVID.

B. Results on Homo Sapiens Dataset

In this experiment, we perform the MultiFacTV algorithm on Homo sapiens dataset to identify some interesting modules. Homo sapiens dataset is composed of expression profiles of 2920 genes for 14 sclerosis patients over 9 time-points after IFN-β treatment[2]. We consider different patients as different conditions and build a tensor $\mathcal{T}$ of size $2920 \times 14 \times 9$. The MultiFacTV algorithm was performed on such tensor with $L = 40$, and its running time was 257 seconds. In the following, we present three interesting modules identified.

- Module No. 1. The expression profiles of genes in the first module are shown as in Figure 5(a). We see from this figure this module is associated to patients A, B, C, D and E.
- Module No. 2. Figure 5(b) shows the expression profiles of the second module. It can be seen that this module is associated to patients E, F, G and H. The genes in this module down-regulate at time-points 2h and 24h for those patients.
- Module No. 3. Figure 5(c) shows the expression profiles of genes in the third module. Clearly, this module is associated to patients I, J, K, L and M. The genes in this module down-regulate from 0h to 1h, and then remain stable for those patients.

Interestingly, we found the obtained 40 modules can help us group the patients with their associated modules. For instance, the above three modules indicate we can partition the patients into four groups, where group 1 includes patients A to E, group 2 includes patients E to H, group 3 includes patients I to M, and group 4 includes the left patient N. Considering all those 40 modules, we construct a 14×40 membership matrix. Then we cluster 14 patients with k-means algorithm based on this matrix and the results are: $\{A, B, C, D\}$, $\{E, F, G, H\}$, $\{J, K, L, M\}$, $\{I, N\}$. As suggested in [11], this grouping result may be because of the differences of patients in disease history or progression, which would be helpful to design personalized treatment for different patients accordingly in the future.

V. Concluding Remarks

In this paper, we propose the MultiFacTV algorithm by introducing a total variation regularization term to the objective function of tensor factorization, to seek modules from higher-order gene expression profiles with time dimension. We have conducted experiments on Arabidopsis thaliana and Homo sapiens datasets. The results have demonstrated the effectiveness of the MultiFacTV.

In MultiFacTV, the input parameters are determined empirically. It would be interesting to study how the stability selection technique in [7] can be incorporated to determine those parameters adaptively in the future.

Acknowledgment

Y. Ye's research supported in part by NSFC under Grant no.61073195, Shenzhen Science and Technology Program under Grant no. CXB201005250024A

[2]The dataset is available on http://www.ra.cs.uni-tuebingen.de/software/EDISA/downloads/index.htm

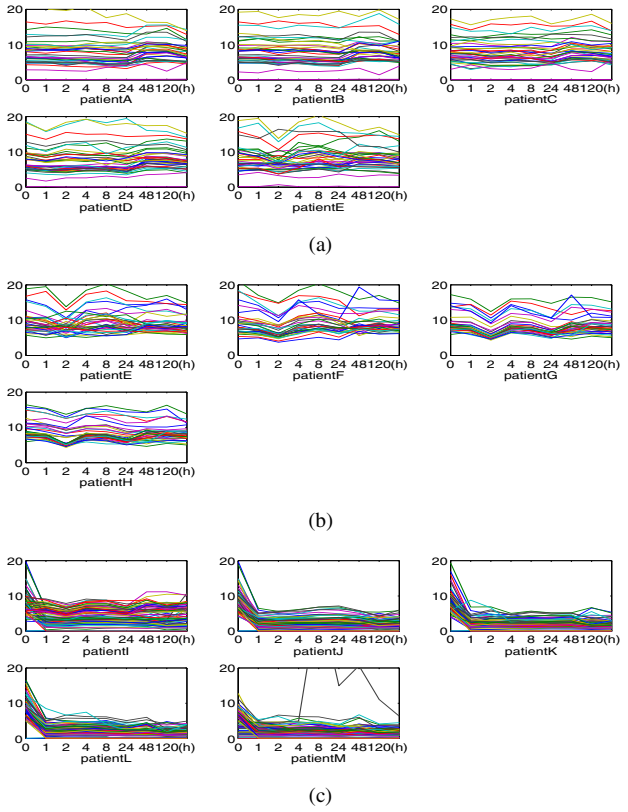

Figure 5. The expression profiles of three modules detected from Homo sapiens dataset. (a) module 1; (b) module 2; (c) module 3. (For space reason, we only show the expression profiles of these modules under their associated conditions/patients.)

and ZD201006100018A, and Natural Scientific Research Innovation Foundation in HIT under Grant no. HIT.NSFIR.2010128. M. Ng's research supported in part by Centre for Mathematical Imaging and Vision, HKRGC grant no. 201812.

REFERENCES

[1] S. Tavazoie, J. Hughes, M. Campbell, R. Cho, G. Church, et al. "Systematic determination of genetic network architecture." *Nature genetics*, 22:281–285, 1999.

[2] M. Eisen, P. Spellman, P. Brown, and D. Botstein. "Cluster analysis and display of genome-wide expression patterns." *Proceedings of the National Academy of Sciences*, 95(25):14863, 1998.

[3] W. Pentney and M. Meila. "Spectral clustering of biological sequence data." In *Proceedings of the 25th AAAI conference*, pages 845–850, 2005.

[4] Y. Cheng and G. Church. "Biclustering of expression data." In *Proceedings of the eighth international conference on intelligent systems for molecular biology*, 1:93–103, 2000.

[5] M. Wall, A. Rechtsteiner, and L. Rocha. "Singular value decomposition and principal component analysis." *A practical approach to microarray data analysis*, pages 91–109, 2003.

[6] M. Lee, H. Shen, J. Huang, and J. Marron. "Biclustering via sparse singular value decomposition." *Biometrics*, 66(4):1087–1095, 2010.

[7] M. Sill, S. Kaiser, A. Benner, and A. Kopp-Schneider. "Robust biclustering by sparse singular value decomposition incorporating stability selection." *Bioinformatics*, 27(15):2089–2097, 2011.

[8] W. Li, C. Liu, T. Zhang, H. Li, M. Waterman, and X. Zhou. "Integrative analysis of many weighted co-expression networks using tensor computation." *PLoS Computational Biology*, 7(6):e1001106, 2011.

[9] L. Omberg, G. Golub, and O. Alter. "A tensor higher-order singular value decomposition for integrative analysis of dna microarray data from different studies." *Proceedings of the National Academy of Sciences*, 104(47):18371, 2007.

[10] M. Narayanan, A. Vetta, E. Schadt, and J. Zhu. "Simultaneous clustering of multiple gene expression and physical interaction datasets." *PLoS computational biology*, 6(4):e1000742, 2010.

[11] J. Supper, M. Strauch, D. Wanke, K. Harter, and A. Zell. "Edisa: extracting biclusters from multiple time-series of gene expression profiles." *BMC bioinformatics*, 8(1):334, 2007.

[12] M. Garcia-Hernandez, T. Berardini, G. Chen, D. Crist, A. Doyle, E. Huala, E. Knee, M. Lambrecht, N. Miller, L. Mueller, et al. "Tair: a resource for integrated arabidopsis data." *Functional & integrative genomics*, 2(6):239–253, 2002.

[13] R. Edgar, M. Domrachev, and A. Lash. "Gene expression omnibus: Ncbi gene expression and hybridization array data repository." *Nucleic acids research*, 30(1):207–210, 2002.

[14] B. Weinstock-Guttman, D. Badgett, K. Patrick, L. Hartrich, R. Santos, D. Hall, M. Baier, J. Feichter, and M. Ramanathan. "Genomic effects of ifn-β in multiple sclerosis patients." *The Journal of Immunology*, 171(5):2694–2702, 2003.

[15] J. Kilian, D. Whitehead, J. Horak, D. Wanke, S. Weinl, O. Batistic, C. DAngelo, E. Bornberg-Bauer, J. Kudla, and K. Harter. "The atgenexpress global stress expression data set: protocols, evaluation and model data analysis of uv-b light, drought and cold stress responses." *The Plant Journal*, 50(2):347–363, 2007.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *"Distributed optimization and statistical learning via the alternating direction method of multipliers*." Now Publishers, 2011.

[17] M. Ng, P. Weiss, and X. Yuan. "Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods." *SIAM Journal on Scientific Computing*, 32(5):2710, 2010.

[18] G. Dennis Jr, B. Sherman, D. Hosack, J. Yang, W. Gao, H. Lane, R. Lempicki, et al. "David: database for annotation, visualization, and integrated discovery." *Genome Biol*, 4(5):P3, 2003.