35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

# Structure-Revealing Data Fusion Model with Applications in Metabolomics

Evrim Acar, Anders J. Lawaetz, Morten A. Rasmussen and Rasmus Bro

*Abstract*— **In many disciplines, data from multiple sources are acquired and jointly analyzed for enhanced knowledge discovery. For instance, in metabolomics, different analytical techniques are used to measure biological fluids in order to identify the chemicals related to certain diseases. It is widely-known that, some of these analytical methods, e.g., LC-MS (Liquid Chromatography - Mass Spectrometry) and NMR (Nuclear Magnetic Resonance) spectroscopy, provide complementary data sets and their joint analysis may enable us to capture a larger proportion of the complete metabolome belonging to a specific biological system. Fusing data from multiple sources has proved useful in many fields including bioinformatics, signal processing and social network analysis. However, identification of *common (shared)* and *individual* (*unshared)* structures across multiple data sets remains a major challenge in data fusion studies. With a goal of addressing this challenge, we propose a novel unsupervised data fusion model. Our contributions are two-fold: (i) We formulate a data fusion model based on joint factorization of matrices and higher-order tensors, which can automatically reveal common and individual components. (ii) We demonstrate that the proposed approach provides promising results in joint analysis of metabolomics data sets consisting of fluorescence and NMR measurements of plasma samples in terms of separation of colorectal cancer patients from controls.**

## I. Introduction

Data fusion, in other words, joint analysis of data from multiple sources, has been shown to enhance knowledge discovery in many disciplines. For instance, in bioinformatics, jointly analyzing multiple data sets representing different organisms [1] or different tissue types [2, 3] improves the understanding of the underlying biological processes. Similarly, in metabolomics, biological fluids such as blood, are measured using different analytical techniques, e.g., LC-MS and NMR, and their fusion has the potential for more accurate biomarker identification [4].

An effective way of jointly analyzing data from multiple sources is to represent data sets as a collection of matrices, and jointly analyze those matrices using collective matrix factorization [5]. Matrix factorization-based data fusion studies have been successfully applied in bioinformatics [1, 2]. Recently, joint matrix factorization approaches have been extended to joint analysis of heterogeneous data sets, i.e., data in the form of matrices and higher-order tensors [6, 7]. For instance, chemical mixtures measured using fluorescence spectroscopy can be represented as a third-order tensor with modes: *mixtures*, *emission* and *excitation wavelengths* while NMR measurements of the same mixtures can be represented using a *mixtures* by *chemical shifts* matrix (Figure 1). Joint factorization of such heterogeneous data has been commonly used to analyze multi-relational data in social networks [8, 9].

While there are many successful data fusion applications, identification of common and individual factors across multiple data sets is still a major challenge. The traditional formulation of joint factorization of data sets is based on modeling the common factors. However, data from multiple sources often have both common and individual factors. Ignoring unshared factors may affect the shared factors as well. In this paper, we develop a new data fusion model for joint factorization of heterogeneous data in order to identify common and individual components. Using numerical experiments, we demonstrate that while the traditional formulation modeling only common factors fails to capture the underlying structures, the proposed approach achieves to identify shared and individual components accurately. Several studies have recently discussed methods revealing common and distinctive components [1, 10, 11]. However, these studies focus on coupled matrix factorizations. Our contributions can be summarized as follows: (i) Introducing a new data fusion model for joint factorization of matrices and higher-order tensors, which can identify the common and individual components across multiple data sets automatically. (ii) Demonstrating the effectiveness of the proposed approach on simulated data and a novel metabolomics application.

We survey the related work in Section II. In Section III, we introduce our data fusion model and the algorithmic approach. Section IV demonstrates the performance of the proposed approach on simulated data and real metabolomics data. Section V concludes with future research directions.

## II. Related work

Data fusion within the context of joint factorization of matrices has been studied for years [2, 5, 12]. The problem is typically formulated as:

$$f(\mathrm{U},\mathrm{V},\mathrm{W}) = \left\|\mathrm{X}-\mathrm{U}\mathrm{V}^{\mathrm{T}}\right\|^2 + \left\|\mathrm{Y}-\mathrm{U}\mathrm{W}^{\mathrm{T}}\right\|^2 \quad (1)$$

where $\mathrm{X} \in \mathbb{R}^{I\times J}$ and $\mathrm{Y} \in \mathbb{R}^{I\times K}$ are matrices coupled in the first mode and the factor matrix corresponding to the shared mode, $\mathrm{U} \in \mathbb{R}^{I\times R}$, is common in both factorizations.

* Research supported by the Danish Council for Independent Research (DFF) - Technology and Production Sciences (FTP) Program under the projects 11-116328 and 11-120947

Evrim Acar, Anders J. Lawaetz, Morten A. Rasmussen and Rasmus Bro are with the Faculty of Science, University of Copenhagen, DK-1958 Frederiksberg C, Denmark (e-mail: {evrim, ajla, mortenr, rb}@life.ku.dk).

Evrim Acar is the corresponding author (e-mail: evrim@life.ku.dk).

As an extension of (1), joint factorization of heterogeneous data, e.g., a third-order tensor, $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, coupled with a matrix, $\mathrm{Y} \in \mathbb{R}^{I \times M}$, can be formulated as:

$$f_1(\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{V}) = \left\| \mathcal{X} - [\![ \mathrm{A}, \mathrm{B}, \mathrm{C} ]\!] \right\|^2 + \left\| \mathrm{Y} - \mathrm{A}\mathrm{V}^{\mathrm{T}} \right\|^2 \qquad (2)$$

where tensor $\mathcal{X}$ and matrix Y are simultaneously factorized through the minimization of (2), which fits a CANDECOMP/PARAFAC (CP) [13, 14] model to $\mathcal{X}$ and factorizes Y in such a way that the factor matrix corresponding the common mode, i.e., $\mathrm{A} \in \mathbb{R}^{I \times R}$, is the same. $\mathrm{B} \in \mathbb{R}^{J \times R}$ and $\mathrm{C} \in \mathbb{R}^{K \times R}$ are the factor matrices of $\mathcal{X}$ corresponding to the second and third modes, respectively. We use the notation $\mathcal{X} = [\![ \mathrm{A}, \mathrm{B}, \mathrm{C} ]\!]$ to denote the CP model. $\mathrm{V} \in \mathbb{R}^{M \times R}$ is the factor matrix that corresponds to the second mode of Y. This formulation of coupled matrix and tensor factorization (CMTF) model, dating back to the studies of Harshman and Lundy [15] and Smilde et al. [6], has recently been a topic of interest in many studies [3, 8, 9, 16, 17].

## III. Our Approach

### A. Model

The formulation in (2) makes an implicit assumption that all columns of factor matrix A, i.e., $\mathrm{a}_r$ for *r=1,...R*, are shared by the matrix and the tensor. However, in general, there are both common and individual factors in coupled data sets. Therefore, we reformulate the problem in such a way that through modeling constraints, we identify the common and individual components in CMTF. We modify $f_1$ and rewrite the optimization problem as follows:

$$\begin{aligned} &\min f_2(\lambda, \sigma, \mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{V}) \\ &= \left\| \mathcal{X} - [\![ \lambda; \mathrm{A}, \mathrm{B}, \mathrm{C} ]\!] \right\|^2 + \left\| \mathrm{Y} - \mathrm{A}\Sigma\mathrm{V}^{\mathrm{T}} \right\|^2 + \beta \|\lambda\|_1 + \beta \|\sigma\|_1 \\ &\text{s.t. } \|\mathrm{a}_r\| = \|\mathrm{b}_r\| = \|\mathrm{c}_r\| = \|\mathrm{v}_r\| = 1, \text{ for } r = 1, \ldots, R. \end{aligned}$$

where $\lambda \in \mathbb{R}^{R \times 1}$ and $\sigma \in \mathbb{R}^{R \times 1}$ correspond to the weights of rank-one components in the third-order tensor and the matrix, respectively. $\Sigma \in \mathbb{R}^{R \times R}$ is a diagonal matrix with entries of $\sigma$ on the diagonal. $\|.\|$ denotes the Frobenius norm for higher-order tensors/matrices and the 2-norm for vectors while $\|.\|_1$ denotes the 1-norm of a vector, i.e., $\|\mathrm{x}\|_1 = \sum_{r=1}^{R} |x_r|$. $\beta \geq 0$ is a penalty parameter. In this formulation, our goal is to sparsify the weights $\lambda$ and $\sigma$ using the 1-norm penalties so that unshared components have norms equal to or close to 0 in one of the data sets.

In order to solve this constrained optimization problem, we first convert it into a differentiable unconstrained optimization problem and then use a first-order optimization algorithm. Using the quadratic penalty method [18], we convert the constraints into penalty terms. In order to make the objective function differentiable, we also replace the 1-norm terms with differentiable approximations, e.g., for sufficiently small $\varepsilon > 0, \sqrt{x_i^2 + \varepsilon} = |x_i|$ [19]. Our objective

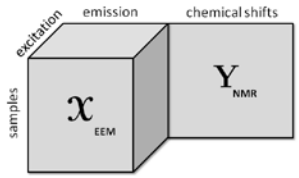



Figure 1. A third-order tensor coupled with a matrix.

function can be finally formulated as follows, for $\alpha \geq 0$:

$$\begin{aligned} f_3(\lambda, \sigma, \mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{V}) = &\left\| \mathcal{X} - [\![ \lambda; \mathrm{A}, \mathrm{B}, \mathrm{C} ]\!] \right\|^2 + \left\| \mathrm{Y} - \mathrm{A}\Sigma\mathrm{V}^{\mathrm{T}} \right\|^2 \\ &+ \beta \sum_{r=1}^{R} \sqrt{\lambda_r^2 + \varepsilon} + \beta \sum_{r=1}^{R} \sqrt{\sigma_r^2 + \varepsilon} \\ &+ \alpha \sum_{r=1}^{R} (\|\mathrm{a}_r\| - 1)^2 + \alpha \sum_{r=1}^{R} (\|\mathrm{b}_r\| - 1)^2 \\ &+ \alpha \sum_{r=1}^{R} (\|\mathrm{c}_r\| - 1)^2 + \alpha \sum_{r=1}^{R} (\|\mathrm{d}_r\| - 1)^2 \end{aligned} \qquad (3)$$

### B. Algorithm

We minimize the objective function $f_3$ in (3) using a gradient-based optimization approach [18]. The gradient can be computed by taking the partial derivates of $f_3$ with respect to the factor matrices and the vectors $\lambda$ and $\sigma$. The gradient $\nabla f_3$ of length *R(I+J+K+M+2)* can then be formed by vectorizing the partials with respect to the factor matrices and concatenating them with the partials with respect to vectors $\lambda$ and $\sigma$, as follows:

$$\nabla f_3 = \left[ \mathrm{vec}(\frac{\partial f_3}{\partial \mathrm{A}})^{\mathrm{T}} \quad \mathrm{vec}(\frac{\partial f_3}{\partial \mathrm{B}})^{\mathrm{T}} \quad \mathrm{vec}(\frac{\partial f_3}{\partial \mathrm{C}})^{\mathrm{T}} \quad \mathrm{vec}(\frac{\partial f_3}{\partial \mathrm{V}})^{\mathrm{T}} \quad (\frac{\partial f_3}{\partial \lambda})^{\mathrm{T}} \quad (\frac{\partial f_3}{\partial \sigma})^{\mathrm{T}} \right]^{\mathrm{T}}$$

Let $\mathcal{T} = [\![ \lambda; \mathrm{A}, \mathrm{B}, \mathrm{C} ]\!]$ and $\mathrm{Z} = \mathrm{A}\Sigma\mathrm{V}^{\mathrm{T}}$. Assuming that each term in $f_3$ is multiplied by 0.5 for the ease of computation, the partial derivatives can be computed as:

$$\begin{aligned} \frac{\partial f_3}{\partial \mathrm{A}} &= (\mathrm{T}_{(1)} - \mathrm{X}_{(1)})(\lambda \odot \mathrm{C} \odot \mathrm{B}) + (\mathrm{Z} - \mathrm{Y})\mathrm{V}\Sigma + \alpha(\mathrm{A} - \bar{\mathrm{A}}) \\ \frac{\partial f_3}{\partial \mathrm{B}} &= (\mathrm{T}_{(2)} - \mathrm{X}_{(2)})(\lambda \odot \mathrm{C} \odot \mathrm{A}) + \alpha(\mathrm{B} - \bar{\mathrm{B}}) \\ \frac{\partial f_3}{\partial \mathrm{C}} &= (\mathrm{T}_{(3)} - \mathrm{X}_{(3)})(\lambda \odot \mathrm{B} \odot \mathrm{A}) + \alpha(\mathrm{C} - \bar{\mathrm{C}}) \\ \frac{\partial f_3}{\partial \mathrm{V}} &= (\mathrm{Z} - \mathrm{Y})^{\mathrm{T}}\mathrm{A}\Sigma + \alpha(\mathrm{V} - \bar{\mathrm{V}}) \\ \frac{\partial f_3}{\partial \lambda_r} &= (\mathcal{T} - \mathcal{X}) \times_1 \mathrm{a}_r \times_2 \mathrm{b}_r \times_3 \mathrm{c}_r + \frac{\beta}{2} \frac{\lambda_r}{\sqrt{\lambda_r^2 + \varepsilon}} \\ \frac{\partial f_3}{\partial \sigma_r} &= \mathrm{a}_r^{\mathrm{T}}(\mathrm{Z} - \mathrm{Y})\mathrm{v}_r + \frac{\beta}{2} \frac{\sigma_r}{\sqrt{\sigma_r^2 + \varepsilon}} \end{aligned}$$

where $\mathrm{X}_{(n)}$ denotes tensor $\mathcal{X}$ unfolded in the *n*th mode; $\times_n$ denotes the tensor-vector product in the *n*th mode, and $\odot$ denotes the Khatri-Rao product (See [20, 21] for details). $\bar{\mathrm{A}}$ corresponds to A with columns divided by their 2-norm.

Once the gradient is computed, to minimize the objective in (3), we use the Nonlinear Conjugate Gradient method

[18] with the Moré-Thuente line search as implemented in the Poblano Toolbox [22]. Function and gradient computations are available with the CMTF Toolbox [23].

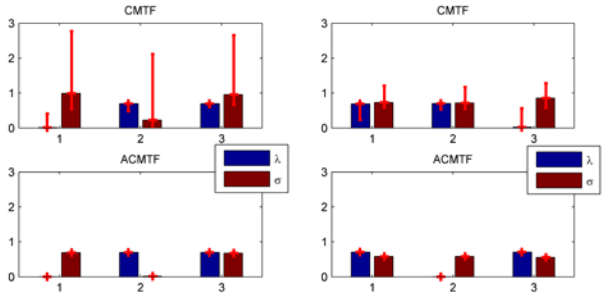
## IV. Experiments and Results

In this section, we first show the benefits of formulating coupled matrix and tensor factorization as in (3) using simulated data. Then we demonstrate the usefulness of the proposed approach in a novel metabolomics application.

### A. Simulated Data

We generate factor matrices $A \in \mathbb{R}^{I \times R}, B \in \mathbb{R}^{J \times R}, C \in \mathbb{R}^{K \times R}$ and $V \in \mathbb{R}^{M \times R}$ with entries drawn from the standard normal. The columns of factor matrices are normalized to unit norm. Here, we set $I$=50, $J$=30, $K$=40, $M$=20 and $R$=3. The factor matrices are then used to construct a third-order tensor $\mathcal{X} = [\![\lambda; A, B, C]\!]$ coupled with matrix $Y = A\Sigma V^T$, where $\lambda$ and diagonal entries of diagonal matrix $\Sigma$, i.e., $\sigma$, of length $R$, correspond to the weights of rank-one tensors and matrices, respectively. Small amount of noise is added to each data set. Using different sets of weights, we generate cases where $R$ components are shared differently among data sets: (i) **Case 1:** One common and one individual component in each data set, i.e., $\lambda=[1\ 0\ 1]^T$, $\sigma=[1\ 1\ 0]^T$. (ii) **Case 2:** One individual component in the matrix, i.e., $\lambda=[1\ 1\ 0]^T$, $\sigma=[1\ 1\ 1]^T$. (iii) **Case 3:** One individual component in the tensor, i.e., $\lambda=[1\ 1\ 1]^T$, $\sigma=[1\ 1\ 0]^T$.

Coupled data sets are then jointly factorized using the traditional CMTF model in (2) and our proposed approach in (3) (referred to as ACMTF). We use $\beta = 0.001$ and $\alpha = 1$ in our experiments. Figure 2 and 3 demonstrate the weights $(\lambda, \sigma)$ estimated using both models for 100 runs returning the same function value (Multiple random starts are used and the minimum function value is obtained 100 times). When we use CMTF, weights are estimated by normalizing the columns of the extracted factor matrices. In Figure 2(a), we expect to recover $\lambda=[1\ 0\ 1]^T$, $\sigma=[1\ 1\ 0]^T$; however, we observe that weights captured by CMTF widely vary hiding the true underlying structure of the data sets. On the other hand, ACMTF reveals the exact structure indicating that there is one common and one individual component in each data set. Similarly, in Figure 2(b), we expect to see three non-zero weights for the matrix and two non-zero weights for the tensor. However, there is too much variation for the same function value hiding the true structure of the data sets. ACMTF, on the other hand, can identify common and individual components accurately. Unlike Case 1 and 2, CMTF performs well for Case 3, where the tensor has all three components and two of them are shared with the matrix (Figure 3). This is as a result of the uniqueness properties of the CP model [20, 21]. Numerical experiments demonstrate that building a model by taking into account individual components as well as common components can be beneficial in terms of capturing the true underlying structures.




(a) Case 1 (b) Case 2

Figure 2. Weights estimated by CMTF and ACMTF.

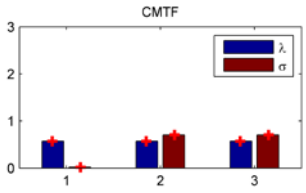


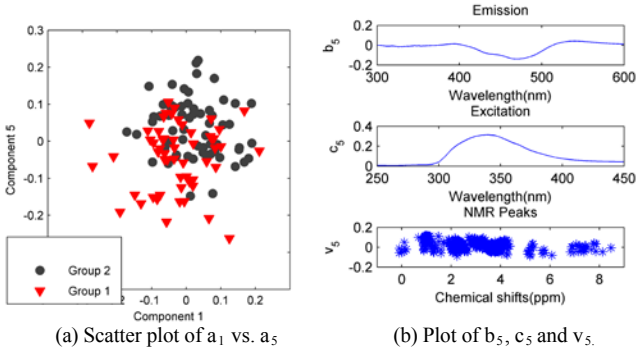

Figure 3. Weights estimated by CMTF for Case 3.

### B. Metabolomics Data:

Next, we use the proposed data fusion model to jointly analyze metabolomics data sets. The data consists of human plasma samples, which are part of a study conducted on patients undergoing large bowel endoscopy due to symptoms which could be associated to colorectal cancer (CRC) [24]. In this paper, we use the samples from verified CRC group (group 1) and the group with other nonmalignant findings (group 2). We have 55 and 64 samples in group 1 and group 2, respectively. These samples are measured using both fluorescence and NMR spectroscopy. Fluorescence measurements are represented as a third-order tensor with modes: *samples*, *emission* and *excitation wavelengths* (Figure 1). We have used the undiluted samples measured in the spectral region of emission wavelengths 300nm to 600nm and excitation wavelengths 250nm to 450nm. For the same samples, $^1$H-NMR spectra have also been collected and the data has been preprocessed by identifying the peaks as described in [24]. NMR measurements can be represented as a *samples* by *peaks (chemical shifts)* matrix. In summary, we have a third-order tensor of size $119 \times 301 \times 41$ coupled with a matrix of size $119 \times 455$ (Figure 1). Based on the prior chemical knowledge, Beer's law suggests that the fluorescence data will follow a CP model. This makes the CMTF model adequate for modeling these coupled data sets.

After centering both data sets across the sample mode and scaling the peaks in NMR with their standard deviation, we jointly factorize the matrix and the tensor by extracting the same factor matrix from the sample mode using the model in (3). We set $\beta = 0.001$ and $\alpha = 1$. The 8-component model reveals the following weights: $\lambda$=[0.82, 0.72, 0.30, 0.39, 0.10, 0.06, 0.09, 0.00] and $\sigma$=[0.14, 0.20, 0.35, 0.14, 0.42, 0.59, 0.46, 0.39] indicating that, out of 8 components, 7 of them are common while the last component is only available

in NMR. Among the common components, first and fifth components play a role in the separation of CRC samples from the control group (Figure 4(a)). When samples are clustered using *k*-means into two clusters based on $a_1$ and $a_5$, we achieve 71.4% accuracy (with 63.6 % sensitivity, 78.1% specificity) in terms of separation of CRC samples. Figure 4(b) illustrates the factor vectors corresponding to the emission and excitation modes as well as the NMR peaks, which are mainly responsible for the separation in Figure 4(a). The factor vectors extracted from the fluorescence data with excitation and emission maximum at 340nm/460nm, respectively, can be assigned to NAD(P)H. Increased levels of NAD(P)H have also previously been associated with cancer, as it is one of the factors affected by the altered metabolic functions in cancer cells compared to healthy cells [25]. Further research is needed for identification of the chemicals represented by the factor vector corresponding to the NMR peaks. When we explore the eighth component that is only available in NMR, we observe that it can be related to gender separation. When samples are clustered based on $a_8$ into two clusters, we can separate females and males with 63.0% accuracy (There are almost equal number of males and females, i.e., 58 vs. 61). While these findings need to be further validated biologically, this example illustrates that the proposed model is useful in terms of capturing common and individual components in data fusion studies.




(a) Scatter plot of $a_1$ vs. $a_5$ (b) Plot of $b_5$, $c_5$ and $v_5$.

Figure 4. Factors separating the group of CRC samples from the control group captured by the data fusion model in (3).

## V. Conclusion

In this paper, we have introduced a structure-revealing data fusion model that can identify common and individual factors across multiple data sets in the form of matrices and higher-order tensors. Numerical experiments demonstrate the benefit of modeling common and individual components by incorporating sparsity penalties on component weights. We have also demonstrated the applicability of the proposed model on a novel metabolomics application. We plan to study its sensitivity to penalty parameters and better understand its uniqueness properties. Furthermore, extension of the proposed idea for identification of common and individual components to joint factorization of data sets with different noise models as well as alternative optimization approaches are future topics of interest.

## References


[1] O. Alter, P. O Brown, and D. Botstein, "Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms", *PNAS,* 100:3351-3356, 2003.

[2] L. Badea, "Extracting gene expression profiles common to colon and pancreatic adenocarcinoma using simultaneous nonnegative matrix factorization", *In Pacific Symposium on Biocomputing*, pp. 279-290, 2008.

[3] E. Acar, G. E. Plopper, and B. Yener, "Coupled analysis of in vitro and histology tissue samples to quantify structure-function relationship", *PLoS One*, 7(3): e32227, 2012.

[4] S. E. Richards, M. E. Dumas, J. M. Fonville et al., "Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework", *Chemometr. Intell.*, 104(1), 121–131, 2010.

[5] A. P. Singh and G. J. Gordon, "Relational Learning via collective matrix factorization ", *In KDD'08,* pp. 650-658, 2008.

[6] A. K. Smilde, J. A. Westerhuis, and R. Boque, "Multiway multiblock component and covariates regression models", *Journal of Chemometrics,* 14:301-331, 2000.

[7] A. Banerjee, S. Basu and S. Merugu, "Multi-way clustering on relation graphs", *In SDM'07,* pp. 145-156, 2007.

[8] Y. R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, and A. Kelliher, " Metafac: community discovery via relational hypergraph factorization", *In KDD'09,* pp. 527-536, 2009.

[9] B. Ermis, E. Acar, and T. Cemgil, "Link Prediction via Generalized Coupled Tensor Factorisation", *In ECML/PKDD Workshop on Collective Learning and Inference on Structured Data*, 2012.

[10] K. Van Deun, I. Van Mechelen, L. Thorrez, M. Schouteden, B. De Moor, M. J. van der Werf, et al., "DISCO-SCA and Properly Applied GSVD as Swinging Methods to Find Common and Distinctive Processes", *PLoS One,*7(5): e37840, 2012.

[11] S. K. Gupta, D. Phung, B. Adams, T. Tran and S. Venkatesh, "Nonnegative Shared Subspace Learning and Its Application to Social Media Retrieval", *In KDD'10*, pp. 1169-1178, 2010.

[12] J. A. Westerhuis, T. Kourti, and J. F. Macgregor, "Analysis of multiblock and hierarchical pca and pls models", *Journal of Chemometrics.*, 12: 301-321, 1998.

[13] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multi-model factor analysis", *UCLA Working Papers in Phonetics,* 16:1-84, 1970.

[14] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition", *Psychometrika*, 35: 283-319, 1970.

[15] R. A. Harshman and M. E. Lundy, "Parafac: Parallel factor analysis", *Computational Statistics and Data Analysis,* 18: 39-72, 1994.

[16] T. F. Wilderjans, E. Ceulemans, H. A. L. Kiers, and K. Meers, "The LMPCA program: A graphical user interface for fitting the linked-mode parafac-pca model to coupled real-valued data", *Behavior Research Methods*, 41: 1073-1082, 2009.

[17] E. Acar, T. G. Kolda, and D. M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations", *In KDD Workshop on Mining and Learning with Graphs*, 2011.

[18] J. Nocedal, and S.J. Wright, *Numerical Optimization*, Springer, 2006.

[19] S. Lee, H. Lee, P. Abbeel, and A. Y. Ng, " Efficient l1 regularized logistic regression", *In AAAI'06*, pp. 401-408, 2006.

[20] E. Acar and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey", *IEEE Transactions on Knowledge and Data Engineering*, 21(1): 1-15, 2009.

[21] T. G. Kolda and B. Bader, "Tensor Decompositions and Applications", *SIAM Review*, 51(3): 455-500, 2009.

[22] D. M. Dunlavy, T. G. Kolda and E. Acar, "Poblano v1.0:A Matlab toolbox for gradient-based optimization", Sandia National Labs, Tech. Rep. SAND2010-1422, 2010.

[23] The MATLAB CMTF Toolbox, http://www.models.life.ku.dk/joda/CMTF_Toolbox [April 10, 2013].

[24] R. Bro, H. J. Nielsen, F. Savorani, K. Kjeldahl et al., "Data Fusion in Metabolomic Cancer Diagnostics", *Metabolomics*, 9(1):3-8, 2013.

[25] I. Georgakoudi, B. C. Jacobson, M. G. Muller, E. E. Sheets, K. Badizadegan et al., "NAD(P)H and Collagen as in Vivo Quantitative Fluorescent Biomarkers of Epithelial Precancerous Changes", *Cancer Research,* 62(3), 682-7, 2002.