A FLEXIBLE MODELING FRAMEWORK FOR COUPLED MATRIX AND TENSOR FACTORIZATIONS

Evrim Acar[*], *Mathias Nilsson*[*†], *Michael Saunders*[‡]

[*] University of Copenhagen, Faculty of Science, Frederiksberg C, Denmark
[†] University of Manchester, School of Chemistry, Manchester, UK
[‡] Stanford University, Dept. of Management Science and Engineering, CA, USA

ABSTRACT

Joint analysis of data from multiple sources has proved useful in many disciplines including metabolomics and social network analysis. However, data fusion remains a challenging task in need of data mining tools that can capture the underlying structures from multi-relational and heterogeneous data sources. In order to address this challenge, data fusion has been formulated as a coupled matrix and tensor factorization (CMTF) problem. Coupled factorization problems have commonly been solved using alternating methods and, recently, unconstrained all-at-once optimization algorithms. In this paper, unlike previous studies, in order to have a flexible modeling framework, we use a general-purpose optimization solver that solves for all factor matrices simultaneously and is capable of handling additional linear/nonlinear constraints with a nonlinear objective function. We formulate CMTF as a constrained optimization problem and develop accurate models more robust to overfactoring. The effectiveness of the proposed modeling/algorithmic framework is demonstrated on simulated and real data.

Index Terms— data fusion, tensor factorizations, nonlinear optimization, nonlinear constraints, SNOPT.

1. INTRODUCTION

In many domains, data from complementary information sources are collected. For instance, in metabolomics, with a goal of finding biomarkers, biofluids such as blood or urine are studied using complementary analytical techniques including LC-MS (Liquid Chromatography - Mass Spectrometry) and NMR (Nuclear Magnetic Resonance) spectroscopy [1]. In recommendation systems, in order to recommend activities at certain locations, in addition to users' previous activities, information such as points of interest at each location or user similarities based on social network data are also collected [2]. Joint analysis of data from such complementary sources has proved useful in terms of missing data estimation [2, 3, 4], clustering performance [4], and improving the understanding of the underlying biological processes [5].

However, analysis of data from multiple sources often needs to deal with multi-relational data of different orders, e.g., in the form of matrices and higher-order tensors (Figure 1), and with shared and unshared factors. Joint analysis of such data sets is challenging, especially when the goal is to capture the underlying structures. For instance, metabolomics applications require underlying factors to be captured accurately and uniquely so that extracted factors can be used further to identify biomarkers corresponding to a problem of interest, e.g., a specific type of diet or a disease.

Coupled factorization methods have emerged as an effective tool for joint analysis of multiple data sets. While earlier studies focused on factorization of multiple data sets in the form of matrices [5, 6], in order to deal with heterogeneous data sets (i.e., in the form of matrices and higher-order tensors), coupled matrix and tensor factorization (CMTF) methods have been developed [2, 3, 4, 7, 8]. Joint factorization of a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, coupled with a matrix $\mathbf{Y} \in \mathbb{R}^{I \times M}$, can be formulated as

$$f(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{V}) = \| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\mathbf{V}^{\mathsf{T}} \|^2, \quad (1)$$

where tensor $\mathcal{X}$ and matrix $\mathbf{Y}$ are factorized through the minimization of (1), which fits a CANDECOMP/PARAFAC (CP) [9, 10] model to $\mathcal{X}$ and factorizes $\mathbf{Y}$ in such a way that the factor matrix corresponding to the common mode, namely $\mathbf{A} \in \mathbb{R}^{I \times R}$, is the same. Here, $\| \,.\, \|$ denotes the Frobenius norm. $\mathbf{B} \in \mathbb{R}^{J \times R}$ and $\mathbf{C} \in \mathbb{R}^{K \times R}$ are factor matrices corresponding to the second and third modes of $\mathcal{X}$, and $\mathbf{V} \in \mathbb{R}^{M \times R}$ corresponds to the second mode of $\mathbf{Y}$. We use the notation $\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ to denote the CP model, which represents a higher-order tensor as a sum of rank-one tensors, i.e., $\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$, where $\circ$ denotes the vector outer product. Coupled matrix and tensor factorization models have been extended to tensor factorizations other than CP [3] and to different loss functions [3, 8]. Here, we focus on CMTF models using CP for modeling higher-order tensors and the squared Frobenius norm as the loss function.

Depending on the application, the CMTF formulation (1)

This work was supported by the Danish Council for Independent Research - Technology and Production Sciences [projects 11-116328, 11-120947] and the National Institute of General Medical Sciences of the National Institutes of Health [award U01GM102098].

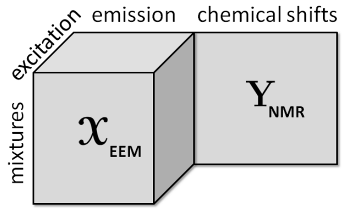


Fig. 1: A third-order tensor of fluorescence measurements coupled with a matrix representing the NMR measurements.

may need to incorporate constraints. For instance, nonnegativity constraints are used to build easily-interpretable models [8], and certain formulations of CMTF require unit norm constraints on the columns of the factor matrices to capture explicitly the weights of rank-one components in each data set [11]. The traditional approach for fitting CMTF models is to use alternating algorithms minimizing the objective function for one factor matrix at a time with the others fixed [6]. Alternating methods have been the workhorse for fitting tensor models as well because of the ease of implementation. However, nonlinear optimization methods solving for all factor matrices simultaneously have proved to outperform alternating least squares for fitting a CP model [12, 13, 14]. A potential limitation of all-at-once optimization approaches is that they are not flexible in terms of imposing constraints. Therefore, in this paper, we study the use of a general-purpose optimization solver SNOPT (Sparse Nonlinear OPTimizer) [15] to solve for all factor matrices simultaneously while handling a nonlinear objective function and additional linear/nonlinear constraints. We formulate CMTF as a constrained optimization problem and demonstrate that the optimizer can be used to develop accurate data fusion models. For instance, by imposing angular constraints, we develop CMTF models that are more robust to overfactoring, i.e., less likely to extract a spurious extra component that is not in any of the data sets. Using numerical experiments on both simulated and real data, we demonstrate the effectiveness of the proposed modeling/algorithmic framework.

2. COUPLED MATRIX AND TENSOR FACTORIZATIONS

Coupled factorizations of heterogeneous data have been originally formulated as in (1) [2, 3, 7]. Algorithmic approaches for minimizing (1) rely on either alternating algorithms or unconstrained first-order optimization methods.

While the CMTF formulation (1) can accurately capture the underlying structures in coupled data sets when all factors are shared across data sets [7], it may fail to capture the underlying factors uniquely in the presence of both shared and unshared components [11]. Therefore, in order to identify shared/unshared factors in coupled data sets, CMTF has been reformulated as follows [11]:

$$\begin{aligned} f(\boldsymbol{\lambda}, \boldsymbol{\Sigma}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{V}) \\ &= \| \mathcal{X} - [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T} \|^2 \\ &\quad + \gamma \| \boldsymbol{\lambda} \|_1 + \gamma \| \boldsymbol{\sigma} \|_1 , \end{aligned} \tag{2}$$

where the columns of factor matrices have unit norm, i.e., $\| \mathbf{a}_r \| = \| \mathbf{b}_r \| = \| \mathbf{c}_r \| = \| \mathbf{v}_r \| = 1$ for $r = 1, \ldots, R$, and $\boldsymbol{\lambda} \in \mathbb{R}^{R\times 1}$ and $\boldsymbol{\sigma} \in \mathbb{R}^{R\times 1}$ are the weights of rank-one components in the third-order tensor and the matrix, respectively. $\boldsymbol{\Sigma} \in \mathbb{R}^{R\times R}$ is a diagonal matrix with entries of $\boldsymbol{\sigma}$ on the diagonal. $\| . \|_1$ denotes the 1-norm of a vector, i.e., $\| \mathbf{x} \|_1 = \sum_{r=1}^{R} |x_r|$, and $\gamma > 0$ is a penalty parameter. This formulation sparsifies the weights through the 1-norm penalties so that unshared factors are expected to have weights equal to 0 in some data sets. In [11] the model is fitted to data by adding the norm constraints as quadratic penalty terms to the objective, replacing the 1-norm terms with differentiable approximations, i.e., for sufficiently small $\epsilon > 0$, $\sqrt{x_i^2 + \epsilon} = |x_i|$, and minimizing the objective function using a nonlinear conjugate gradient method. The reformulated CMTF model, called the *structure-revealing data fusion model*, has been effective in terms of identifying shared/unshared factors. However, the unconstrained optimization approach lacks the flexibility to incorporate additional constraints.

3. CMTF USING CONSTRAINED OPTIMIZATION

In this section, we formulate the structure-revealing CMTF model [11] as a constrained optimization problem and discuss its extensions by imposing various constraints. More specifically, we add (i) nonnegativity constraints on the factor matrices in order to extract interpretable factors, and (ii) angular constraints to build CMTF models robust to overfactoring.

The structure-revealing data fusion model in (2) can be formulated as a constrained optimization problem as in (3). Here, we make use of the fact that weights of rank-one components, i.e., λ_r and σ_r, are nonnegative; therefore, instead of the 1-norm penalty terms in (2), we have nonnegativity constraints on λ_r, σ_r for $r = 1, \ldots, R$, and upper bounds on the summations.

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{V},\Sigma,\lambda} \; & \| \mathcal{X} - [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T} \|^2 \\ \text{s.t.} \; & \|\mathbf{a}_r\|_2 = \|\mathbf{b}_r\|_2 = \|\mathbf{c}_r\|_2 = \|\mathbf{v}_r\|_2 = 1, \\ & \sum_{r=1}^{R} \lambda_r \le \beta, \quad \sum_{r=1}^{R} \sigma_r \le \beta, \\ & \sigma_r, \lambda_r \ge 0 \text{ for } r = 1, \ldots, R, \end{aligned} \tag{3}$$

where $\beta > 0$ is a user-defined parameter. This formulation jointly factorizes heterogeneous data sets, and the linear constraints sparsify the weights to reveal shared/unshared factors.

3.1. Additional Constraints

Additional constraints are needed in many data fusion problems. For instance, nonnegativity constraints are used to improve the interpretability of the models or incorporate prior knowledge. Coupled data sets in Figure 1 can be jointly analyzed using CMTF to capture the chemicals visible to each data set. Spectral profiles are nonnegative, and when captured accurately, they can be used to identify individual chemicals in mixtures. To incorporate such prior knowledge, we can enforce nonnegativity constraints on the factor matrices in (3), e.g., $b_{jr} \geq 0$, for $r = 1, \ldots, R, j = 1, \ldots, J$.

Another type of constraint also proves useful, for data fusion models in the case of overfactoring. When coupled data sets are overfactored, a shared factor may be represented by two closely-correlated factors; therefore, the structure-revealing CMTF model fails to identify shared factors accurately. To tackle this problem, we introduce constraints on the angle between the columns of the factor matrices, as follows:

$$\begin{aligned}
\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{V},\Sigma,\lambda} \; & \| \mathcal{X} - [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}} \|^2 \\
\text{s.t.} \quad & \|\mathbf{a}_r\|_2 = \|\mathbf{b}_r\|_2 = \|\mathbf{c}_r\|_2 = \|\mathbf{v}_r\|_2 = 1, \\
& |\mathbf{a}_r^{\mathsf{T}}\mathbf{a}_p| \leq \theta, |\mathbf{b}_r^{\mathsf{T}}\mathbf{b}_p| \leq \theta, |\mathbf{c}_r^{\mathsf{T}}\mathbf{c}_p| \leq \theta, |\mathbf{v}_r^{\mathsf{T}}\mathbf{v}_p| \leq \theta, \\
& \sum_{r=1}^{R} \lambda_r \leq \beta, \quad \sum_{r=1}^{R} \sigma_r \leq \beta, \\
& \sigma_r, \lambda_r \geq 0 \text{ for } r, p \in \{1, \ldots, R\}, r \neq p,
\end{aligned} \tag{4}$$

where $\beta > 0$ and $0 \leq \theta \leq 1$ are user-defined parameters. Angular constraints prevent two factors from being similar and cause the extra factor to get zero weights ($\lambda_r = \sigma_r = 0$), indicating that coupled data sets are overfactored. These types of angular constraints have also been used to deal with the non-existence of the CP model [16]. Previously, overfactoring has been studied for CP models, showing that all-at-once optimization methods are more robust to overfactoring than alternating least squares [12, 13]. Overfactoring has also been studied for the CMTF formulation (1) [7]. However, these studies focus on comparisons of algorithms rather than improving the models to make them robust to overfactoring.

3.2. SNOPT

We solve the CMTF problems in Section 3 using SNOPT [15], which is designed for large constrained optimization problems with smooth nonlinear functions in the objective and constraints. SNOPT uses a sequential quadratic programming (SQP) algorithm to minimize an augmented Lagrangian. The sequence of QP subproblems involve linearized constraints and limited-memory quasi-Newton approximations to the Hessian of the Lagrangian. In our experiments, we run SNOPT 7 on MATLAB 8.1. We provide function and gradient values for the objective and constraints as MATLAB functions.

4. EXPERIMENTS

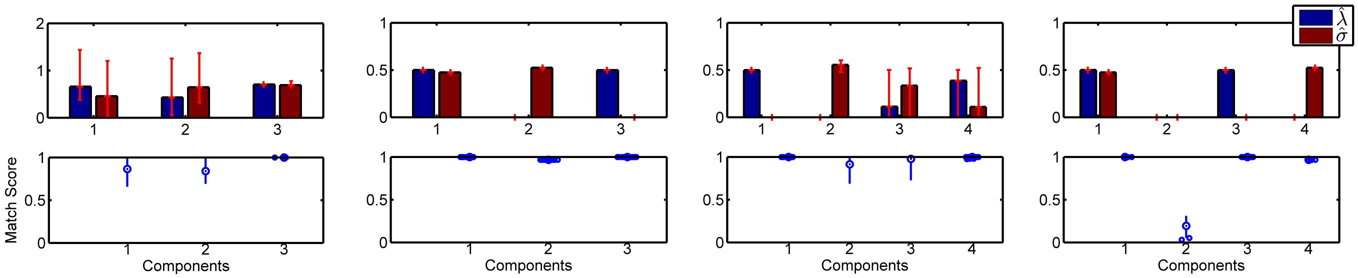
In this section, using experiments on simulated data, we demonstrate that the constrained optimization approach is successful in terms of capturing shared/unshared components in heterogeneous data sets. We also use the proposed CMTF models for jointly analyzing fluorescence spectroscopic and NMR measurements of mixtures with known chemical composition and show that individual chemicals in the mixtures can be accurately captured even in the case of overfactoring.

4.1. Simulated Data

We generate factor matrices $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}$, $\mathbf{V} \in \mathbb{R}^{M \times R}$ with entries randomly drawn from the standard normal distribution and columns normalized to unit norm. Here, we set $I = 50, J = 30, K = 40$ and $M = 20$. The factor matrices are used to construct a third-order tensor $\mathcal{X} = [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ coupled with a matrix $\mathbf{Y} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}$, where $\boldsymbol{\lambda}$ and diagonal entries of diagonal matrix $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\sigma}$, of length R, correspond to the weights of rank-one third-order tensors and matrices, respectively. As the test problem, we study the case with one shared and one unshared component in each data set, i.e., $\boldsymbol{\lambda} = [1\ \ 0\ \ 1]^{\mathsf{T}}$ and $\boldsymbol{\sigma} = [1\ \ 1\ \ 0]^{\mathsf{T}}$. This setting is a challenging case, for which the original CMTF formulation (1) fails to capture the underlying factors [11]. A small amount of Gaussian noise is added to data sets. We jointly factorize these coupled data sets using CMTF formulations with various constraints for different number of factors to study the effect of constraints. Similarly, we generate factor matrices with entries randomly drawn from the standard uniform distribution and construct coupled data using the same structure discussed above to show the effect of constraints together with nonnegativity constraints on the factors.

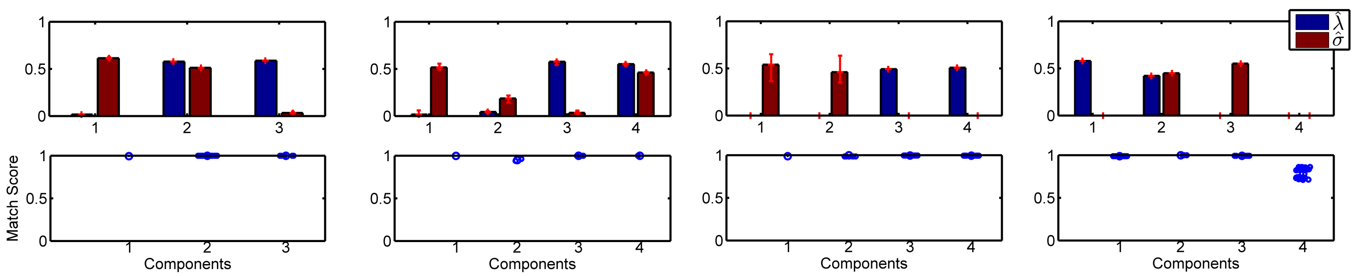
Experiments demonstrate that the CMTF model formulated as a constrained optimization problem (4) can identify the true underlying structures in coupled data with shared/unshared factors even in the case of overfactoring. Figure 2 demonstrates the performance of CMTF models for (a) $R = 3$ with no constraints on weight sums in (3), (b) $R = 3$ with $\beta = 1$ in (3), (c) $R = 4$ with $\beta = 1$ in (3), and (d) $R = 4$ with $\beta = 1$, $\theta = 0.25$ in (4). Top plots show the estimated weights of rank-one components in each data set, i.e., λ and σ, for the runs returning the same function value.[1] While the model fails to identify shared/unshared components without the constraints on sums of weights (Figure 2(a)), the true underlying structure can be captured using those linear constraints (Figure 2(b)). Bottom plots show how well the extracted factors match the true columns of $\mathbf{A}$. Let $\hat{\mathbf{a}}_r$ be the rth column of the factor matrix extracted from the shared mode. "Match score" corresponds to $\frac{\hat{\mathbf{a}}_r^{\mathsf{T}}\mathbf{a}_r}{\| \hat{\mathbf{a}}_r \| \| \mathbf{a}_r \|}$ for the best matching permutation of the columns. We observe that underlying factors can be captured accurately in Figure 2(b). In the case of overfactoring, i.e., $R = 4$, we expect to capture one component with $\lambda_r = \sigma_r = 0$. In Figure

[1]The minimum function value satisfying the constraints is obtained a number of times using random starts. Two function values are considered to be the same when the difference between them is less than 10^{-6}.



(a) $R = 3$ (b) $R = 3, \beta = 1$ (c) $R = 4, \beta = 1$ (d) $R = 4, \beta = 1, \theta = 0.25$

Fig. 2: Performance of CMTF models with various constraints.



(a) $R = 3$ (b) $R = 4$ (c) $R = 4, \beta = 1$ (d) $R = 4, \beta = 1, \theta = 0.8$

Fig. 3: Performance of CMTF models with various constraints in addition to nonnegativity constraints on factor matrices.

2(c), even with the constraints on sums of weights, the model fails to identify shared/unshared components. On the other hand, angular constraints enable recovery of the true structure (Figure 2(d)). Note that the extra component has a low match score in Figure 2(d). Here, θ is determined from the inner products of normalized vectors of given lengths with random entries from the standard normal distribution.

We also carry out experiments on coupled data sets with nonnegative factors. In Figure 3(a), we observe that shared/unshared factors can be successfully identified using only nonnegativity constraints on the factors. However, Figure 3(b) shows that nonnegativity constraints alone cannot deal with the overfactoring problem. Note that match scores for all extracted factors are close to 1, indicating that the extra component models one of the true factors. The CMTF formulation in (3) for $\beta = 1$ together with nonnegativity constraints also fails to identify the true underlying structure (Figure 3(c)). On the other hand, angular constraints enable identification of the true structure in Figure 3(d). Here, a higher θ value is used because entries of the factor matrices are drawn from the standard uniform distribution. The models depend heavily on the choice of β, especially in the case of overfactoring. Small β values sparsify the weights a lot, while higher values make the constraints ineffective. As we increase the component number, β is also increased (see next section), but determining the right β value remains a challenge.

4.2. Real Data

We use the proposed modeling framework for jointly analyzing fluorescence and NMR measurements of 12 mixtures containing four chemicals: Valine-Tyrosine-Valine (VTV), Tryptophan-Glycine (TG), Phenylalanine (Phe), and Propanol. These mixtures are from a larger data set described in detail on www.models.life.ku.dk/joda/prototype. Fluorescence measurements form a third-order tensor with modes: mixtures, emission and excitation wavelengths, and NMR is in the form of a mixtures by chemical shifts matrix[2] (Figure 1). While all chemicals are visible to NMR, propanol does not show up in fluorescence.

Figure 4 demonstrates that when these data sets are jointly factorized using CMTF with constraints (4), three shared and one unshared chemicals can be identified even in the case of overfactoring. A 5-component CMTF model reveals a component with zero weights in each data set, i.e., $\lambda_1 = \sigma_1 = 0$, indicating that data sets are overfactored. Furthermore, the component modeling propanol only shows up in NMR. Figure 4 also shows the factor matrix columns, which capture the relative concentrations of the chemicals in mixtures. We observe that extracted factors (in red) match the true design (blue). Here, a CMTF model with nonnegativity constraints on all factor matrices is used with $\beta = 1.25$ and $\theta = 0.8$. The extracted factors representing fluorescence (emission and ex-

[2]Single gradient level from diffusion NMR data is used in this study.

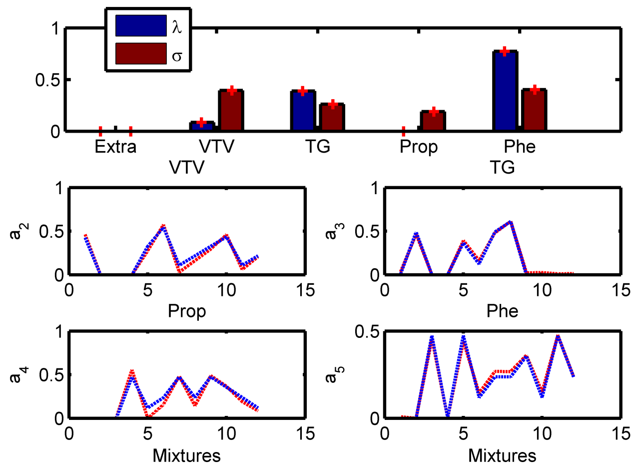


Fig. 4: Weights and scores captured by joint analysis of fluorescence and NMR using CMTF.

citation) and NMR spectra are very similar to the known spectra of the individual chemicals. Comfortingly small distortions were seen even for overfactored models, and for models with the correct number of factors the signatures were virtually identical to the known spectra.

5. CONCLUSIONS

Coupled factorization of heterogeneous data sets has been an effective approach for data fusion. In this paper, we have formulated coupled matrix and tensor factorizations as constrained optimization problems. In order to have a flexible modeling framework, we have used a general-purpose optimization solver SNOPT capable of handling both linear and nonlinear constraints with a nonlinear objective. Numerical results on simulated and real data demonstrate that the proposed modeling/algorithmic approach is effective in terms of building accurate data fusion models. Our focus has been limited to modeling flexibility and accuracy rather than computational efficiency, which we plan to address in future studies. Furthermore, we need a better understanding of the uniqueness properties of coupled matrix and tensor factorization models.

6. REFERENCES

[1] S. E. Richards, M.-E. Dumas, J. M. Fonville, T. M. D. Ebbels, E. Holmes, and J. K. Nicholson, "Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework," *Chemometr Intell Lab*, vol. 104, pp. 121–131, 2010.

[2] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang, "Collaborative filtering meets mobile recommendation: A user-centered approach," in *AAAI*, 2010.

[3] B. Ermis, E. Acar, and A. T. Cemgil, "Link prediction in heterogeneous data via generalized coupled tensor factorization," *Data Min Knowl Disc*, 2013.

[4] A. Banerjee, S. Basu, and S. Merugu, "Multi-way clustering on relation graphs," in *SDM*, 2007, pp. 145–156.

[5] 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*, vol. 100, pp. 3351–3356, 2003.

[6] A. Singh and G.Gordon, "Relational learning via collective matrix factorization," in *KDD*, 2008, pp. 650–658.

[7] E. Acar, T. G. Kolda, and D. M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," in *KDD MLG (arXiv:1105.3422)*, 2011.

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

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

[10] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," *UCLA working papers in phonetics*, vol. 16, pp. 1–84, 1970.

[11] E. Acar, A. J. Lawaetz, M. A. Rasmussen, and R. Bro, "Structure-revealing data fusion model with applications in metabolomics," in *EMBS*, 2013, pp. 6023–6026.

[12] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Comput Stat Data An*, vol. 50, pp. 1700–1734, 2006.

[13] E. Acar, D. M. Dunlavy, and T. G. Kolda, "A scalable optimization approach for fitting canonical tensor decompositions," *J Chemometr*, vol. 25, pp. 67–86, 2011.

[14] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(l_r, l_r, 1)$ terms, and new generalization," *SIAM J Optim*, vol. 23, pp. 695–720, 2013.

[15] P. E. Gill, W. Murray, and M. A. Saunders, "SNOPT: An SQP algorithm for large-scale constrained optimization," *SIAM Rev*, vol. 47, pp. 99–131, 2005.

[16] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans Inform Theory*, vol. 60, pp. 1260–1280, 2014.