Chemometrics and Intelligent Laboratory Systems 129 (2013) 53–63

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Understanding data fusion within the framework of coupled matrix and tensor factorizations

Evrim Acar [a,*], Morten Arendt Rasmussen [a], Francesco Savorani [a], Tormod Næs [a,b], Rasmus Bro [a]

[a] Faculty of Science, University of Copenhagen, Frederiksberg C, Denmark
[b] Nofima Mat, Ås, Norway

ARTICLE INFO

Article history:
Received 3 May 2012
Received in revised form 10 June 2013
Accepted 14 June 2013
Available online 27 June 2013

Keywords:
Data fusion
Matrix factorization
Tensor factorization
Missing data
CANDECOMP
PARAFAC

ABSTRACT

Recent technological advances enable us to collect huge amounts of data from multiple sources. Jointly analyzing such multi-relational data from different sources, i.e., data fusion (also called multi-block, multi-view or multi-set data analysis), often enhances knowledge discovery. For instance, in metabolomics, biological fluids are measured using a variety of analytical techniques such as Liquid Chromatography–Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy. Data measured using different analytical methods may be complementary and their fusion may help in the identification of chemicals related to certain diseases. Data fusion has proved useful in many fields including social network analysis, collaborative filtering, neuroscience and bioinformatics.

In this paper, unlike many studies demonstrating the success of data fusion, we explore the limitations as well as the advantages of data fusion. We formulate data fusion as a coupled matrix and tensor factorization (CMTF) problem, which jointly factorizes multiple data sets in the form of higher-order tensors and matrices by extracting a common latent structure from the shared mode. Using numerical experiments on simulated and real data sets, we assess the performance of coupled analysis compared to the analysis of a single data set in terms of missing data estimation and demonstrate cases where coupled analysis outperforms analysis of a single data set and vice versa.

1. Introduction

Technological advances, e.g., the Internet, communication and multi-media devices, and new medical diagnostic techniques, provide abundance of relational data. Consequently, nowadays, many fields are faced with the challenge of extracting information from data sets from multiple sources. For instance, metabolomics studies involve the detection of a wide range of chemical substances in biological fluids using a number of analytical techniques including Liquid Chromatography–Mass Spectrometry (LC–MS) and proton Nuclear Magnetic Resonance (NMR). These techniques generate data sets often complementary to each other [1]. Similarly, in sensory data analysis, both descriptive sensory data, i.e., data showing the assessment of products based on various attributes such as visual appearance, smell and taste, and instrumental measurements are collected. Such data may need to be jointly analyzed to capture groups of attributes and understand how they relate to consumer liking [2].

Analysis of data from multiple sources has been a topic of interest for decades dating back to one of the earliest models capturing the common variation in two data sets, i.e., Canonical Correlation Analysis [3]. Earlier approaches in data fusion have focused on joint analysis of multiple matrices [3–10] and formulated data fusion problem as collective factorization of matrices. However, in many applications, analysis of data from multiple sources requires handling of data sets of different orders (also referred to as heterogeneous data sets), i.e., matrices and/or higher-order tensors [11–13]. For instance, Banerjee et al. [11] discuss the problem of analyzing heterogeneous data sets with a goal of simultaneously clustering different classes of entities based on multiple relations, where each relation is represented as a matrix (e.g., *movies* by *review words* matrix showing movie reviews) or a higher-order tensor (e.g., *movies* by *viewers* by *actors* tensor showing viewers' ratings). Similarly, coupled analysis of matrices and higher-order tensors has been a topic of interest in the areas of community detection [14], collaborative filtering [15,16], bioinformatics [17], psychometrics [18,19] and chemometrics [13].

Despite many successful applications demonstrating the usefulness of data fusion, it is still difficult to assess when joint analysis of data from multiple sources enhances knowledge discovery. In this paper, our goal is to understand when and to which extent data fusion is useful. In order to do that, we study the performance of joint analysis of data sets with different underlying structures, i.e., with/without or partially common factors, and assess the knowledge discovery performance in terms of missing data estimation. We formulate data fusion as coupled matrix and tensor factorization (CMTF)

* Corresponding author. Tel.: +90 5326780120.
E-mail addresses: evrim@life.ku.dk (E. Acar), mortenr@life.ku.dk (M.A. Rasmussen), frsa@life.ku.dk (F. Savorani), tormod.naes@nofima.no (T. Næs), rb@life.ku.dk (R. Bro).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.06.006

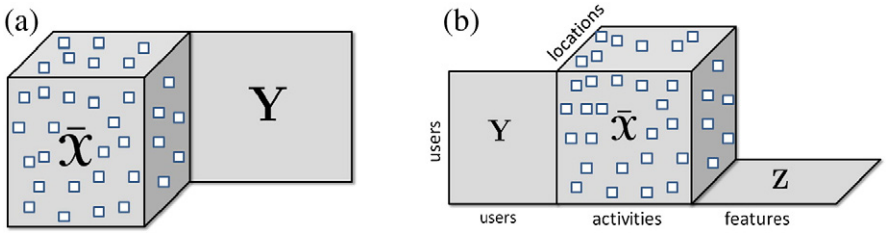

Fig. 1. (a)A third-order tensor with randomly-scattered missing entries coupled with a matrix. (b)A third-order tensor with randomly-scattered missing entries coupled with two matrices in two different modes.

since it is the common formulation used for joint analysis of heterogeneous data [14–20]. Suppose we have a third-order tensor, $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, and a matrix, $\mathbf{Y} \in \mathbb{R}^{I \times M}$, coupled in the first dimension (mode) of each. An R-component CMTF model of a tensor $\mathcal{X}$ and a matrix $\mathbf{Y}$, for Euclidean loss, is defined as [18–20]:

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

Here, tensor $\mathcal{X}$ and matrix $\mathbf{Y}$ are jointly factorized through the minimization of (1), which fits a CANDECOMP/PARAFAC(CP) [21–23] to $\mathcal{X}$ and factorizes matrix $\mathbf{Y}$ in such a way that the factor matrix corresponding to the common mode, i.e., $\mathbf{A} \in \mathbb{R}^{I \times R}$, is the same. $\mathbf{B} \in \mathbb{R}^{J \times R}$ and $\mathbf{C} \in \mathbb{R}^{K \times R}$ are the factor matrices of $\mathcal{X}$ corresponding to the second and third modes, respectively. We use the notation $\mathcal{X} = [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]$ to denote the CP model. Matrix $\mathbf{V} \in \mathbb{R}^{M \times R}$ is the factor matrix that corresponds to the second mode of matrix $\mathbf{Y}$. In CMTF formulation, CP model has been used to model the tensor since it has numerous successful applications in various disciplines [24–26]. The formulation in (1) can be considered as a special case of *Linked mode* PARAFAC models mentioned by Harshman and Lundy [27] (p.65) as well as multi-way multi-block models studied by Smilde et al. [13].

In this paper, we focus on the CMTF formulation (1) and explore the limitations and advantages of data fusion within the framework of CMTF using numerical experiments. We study coupled data sets with different underlying structures, i.e., common, partially common, no common components, and demonstrate that performance of coupled analysis, in terms of missing data estimation, depends on the underlying structures of coupled data sets. Our experiments focus on the case where an incomplete third-order tensor is coupled with a matrix in one mode (Fig. 1a) and we illustrate (i) cases where coupled analysis outperforms analysis of a single data set as well as (ii) cases where coupled analysis performs worse than the analysis of a single data set in terms of estimation of the missing entries in the third-order tensor. Note that the model for CMTF generalizes to multiple higher-order tensors and matrices coupled in several modes, e.g., Fig. 1b, as well.

The paper is organized as follows. In Section 2, we discuss coupled matrix and tensor factorization formulation and its extension to incomplete data sets, i.e., data with missing entries, as well as an all-at-once optimization algorithm for fitting a CMTF model. Section 3 describes numerical experiments on both simulated data sets and real data from sensory data analysis and metabolomics. Finally, we conclude with discussions in Section 4.

2. Coupled matrix and tensor factorization

In this section, we consider joint analysis of a matrix and an *nth*-order tensor with one mode in common, where the tensor is factorized using an R-component CP model and the matrix is factorized by extracting R factors using matrix factorization. Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\mathbf{Y} \in \mathbb{R}^{I_n \times M}$ have the *nth* mode in common, where $n \in \{1,\ldots,N\}$. The objective function for coupled analysis of these two data sets is defined by

$$f\left(\mathbf{A}^{(1)},\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)},\mathbf{V}\right) = \left\|\mathcal{X} - [\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}]\!]\right\|^2 + \left\|\mathbf{Y} - \mathbf{A}^{(n)}\mathbf{V}^{\mathsf{T}}\right\|^2 \tag{2}$$

where $\|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2}$. Similarly, $\|.\|$ refers to Frobenius norm for matrices. Our goal is to find matrices $\mathbf{A}^{(i)} \in \mathbb{R}^{I_i \times R}$, for $i = 1,2,\ldots N$, and matrix $\mathbf{V} \in \mathbb{R}^{M \times R}$ that minimize the objective in (2). This formulation extends to multiple matrices and higher-order tensors coupled in various modes [20]. As an alternative to CMTF, unfolding the higher-order tensor and concatenating with the matrix can also be used to jointly analyze $\mathcal{X}$ and $\mathbf{Y}$; however, this scheme does not generalize to higher-order tensors coupled with matrices in various modes, e.g., Fig. 1b. For instance, in collaborative filtering, if we were to recommend users certain activities based on their locations, we might analyze a *users* × *activities* × *locations* tensor showing how often users perform activities at various locations, coupled with a *users* by *users* similarity matrix as well as a matrix in the form of *locations* by *features* indicating the number of points of interest at various locations [15,16]. In this case, simply unfolding and concatenating data sets will not work. Furthermore, unfolding a higher-order tensor ignores the multi-way structure of the data and, recently, it has been shown that using a tensor model outperforms a matrix model on the unfolded data in terms of missing data recovery [28] (chapter 11).

In the presence of missing data, we can still jointly factorize heterogeneous data sets by ignoring missing entries and fitting the tensor and/or the matrix model to the known data entries. Here, we study the case where tensor $\mathcal{X}$ has missing entries as in Fig. 1a and modify the objective function (2) as[1]

$$f_{\mathcal{W}}\left(\mathbf{A}^{(1)},\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)},\mathbf{V}\right) = \underbrace{\frac{1}{2}\left\|\mathcal{W} * \left(\mathcal{X} - [\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}]\!]\right)\right\|^2}_{f_{\mathcal{W}_1}\left(\mathbf{A}^{(1)},\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)}\right)} + \underbrace{\frac{1}{2}\left\|\mathbf{Y} - \mathbf{A}^{(n)}\mathbf{V}^{\mathsf{T}}\right\|^2}_{f_{\mathcal{W}_2}} \tag{3}$$

where $*$ denotes the Hadamard product. Tensor $\mathcal{W} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ indicates the missing entries of $\mathcal{X}$ such that

$$w_{i_1 i_2 \ldots i_N} = \begin{cases} 1 & \text{if } x_{i_1 i_2 \ldots i_N} \text{ is known,} \\ 0 & \text{if } x_{i_1 i_2 \ldots i_N} \text{ is missing,} \end{cases} \tag{4}$$

for all $i_n \in \{1,\ldots,I_n\}$ and $n \in \{1,\ldots,N\}$.

2.1. All-at-once optimization for CMTF

In order to minimize functions (2) and (3), we use an all-at-once optimization approach called CMTF-OPT [20] that solves the

[1] We multiply with $\frac{1}{2}$ for ease of computation of the gradient.

optimization problem for CMTF for all factor matrices simultaneously. CMTF problem can also be solved using an alternating least squares (ALS) approach, which solves the problem for one factor matrix at a time by fixing the others. However, while ALS methods are easy to implement and flexible in terms of incorporating constraints, direct nonlinear optimization methods solving for all factor matrices simultaneously have shown to be more accurate in the case of CP factorization [29,30] and CMTF [20]. The computational cost *per iteration* is the same for both ALS and CMTF-OPT. Both approaches compute the partial derivatives of the objective with respect to each factor matrix; while ALS solves the partials for one factor matrix at a time fixing the others, CMTF-OPT concatenates the partials, forms the gradient and uses a gradient-based optimization algorithm.

Here we briefly describe the all-at-once algorithm used to minimize $f_{\mathcal{W}}$ in (3). Initially, we compute the partial derivatives of $f_{\mathcal{W}}$ with respect to $\mathbf{A}^{(i)}$, for $i = 1,\ldots,N$. The partial of the first part, $f_{\mathcal{W}_1}$, with respect to $\mathbf{A}^{(i)}$, for $i = 1,\ldots,N$, is as follows:

$$\frac{\partial f_{\mathcal{W}_1}}{\partial \mathbf{A}^{(i)}} = \left(\mathbf{W}_{(i)} * \mathbf{Z}_{(i)} - \mathbf{W}_{(i)} * \mathbf{X}_{(i)}\right)\mathbf{A}^{(-i)},$$

where $\mathcal{Z} = [\![\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(N)}]\!], \mathbf{Z}_{(i)}$ indicates the matricized version of tensor $\mathcal{Z}$ in the *ith* mode (see [25] for matricization/unfolding), and $\mathbf{A}^{(-i)} = \mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(i+1)} \odot \mathbf{A}^{(i-1)} \odot \cdots \odot \mathbf{A}^{(1)}$, for $i = 1,\ldots,N$. $\odot$ denotes the Khatri-Rao product.

The partial derivatives of the second part, $f_{\mathcal{W}_2}$, with respect to $\mathbf{A}^{(i)}$ and $\mathbf{V}$ can be computed as

$$\frac{\partial f_{\mathcal{W}_2}}{\partial \mathbf{A}^{(i)}} = \begin{cases} -\mathbf{Y}\mathbf{V} + \mathbf{A}^{(i)}\mathbf{V}^{\mathsf{T}}\mathbf{V}, & \text{for } i = n, \\ 0 & \text{for } i \neq n, \end{cases}$$

$$\frac{\partial f_{\mathcal{W}_2}}{\partial \mathbf{V}} = -\mathbf{Y}^{\mathsf{T}}\mathbf{A}^{(i)} + \mathbf{V}\mathbf{A}^{(i)\mathsf{T}}\mathbf{A}^{(i)}.$$

Combining the results, we write the partial of $f_{\mathcal{W}}$ with respect to $\mathbf{A}^{(i)}$ as:

$$\frac{\partial f_{\mathcal{W}}}{\partial \mathbf{A}^{(i)}} = \begin{cases} \dfrac{\partial f_{\mathcal{W}_1}}{\partial \mathbf{A}^{(i)}} & \text{for } i \in \{1,\ldots,N\}/\{n\}, \\ \dfrac{\partial f_{\mathcal{W}_1}}{\partial \mathbf{A}^{(i)}} + \dfrac{\partial f_{\mathcal{W}_2}}{\partial \mathbf{A}^{(i)}} & \text{for } i = n. \end{cases}$$

$$\frac{\partial f_{\mathcal{W}}}{\partial \mathbf{V}} = \frac{\partial f_{\mathcal{W}_2}}{\partial \mathbf{V}}$$

The gradient of $f_{\mathcal{W}}$, $\nabla f_{\mathcal{W}}$, which is a vector of size $P = R(\sum_{n=1}^{N} I_n + M)$, can finally be formed by vectorizing the partial derivatives with respect to each factor matrix and concatenating them all.

Once we have the function, $f_{\mathcal{W}}$, and gradient, $\nabla f_{\mathcal{W}}$, we use a gradient-based optimization algorithm, i.e., Nonlinear Conjugate Gradient (NCG) [31] with the Moré–Thuente line search, from the Poblano Toolbox [32], to compute the factor matrices. In this section, we have studied the case where there are missing entries only in tensor $\mathcal{X}$ but note that both the CMTF model and the all-at-once algorithm generalize to coupled factorization of incomplete data sets in the form of matrices or higher-order tensors (See CMTF Toolbox available from http://www.models.life.ku.dk).

In our formulations in (2) and (3), we have used equal weights for different parts of the objective function and in the experiments, we divide each data set by its Frobenius norm so that the model does not favor one part of the objective. However, determining the right weighting scheme remains to be an open research question [18,33].

3. Experiments

3.1. Simulated coupled data sets

Our goal with simulated experiments is to understand when and to which extent CMTF is useful. In order to observe the behavior of CMTF on different types of coupled data sets, we generate coupled data, i.e., a third-order tensor $\mathcal{X}$ of size $I \times J \times K$ and a matrix $\mathbf{Y}$ of size $I \times M$, with different characteristics using the following formulation:

$$\mathcal{X} = \alpha[\![\mathbf{A}^{(1)},\mathbf{A}^{(2)},\mathbf{A}^{(3)}]\!] + (1-\alpha)[\![\mathbf{B}^{(1)},\mathbf{B}^{(2)},\mathbf{B}^{(3)}]\!], \tag{5}$$

$$\mathbf{Y} = \alpha\mathbf{A}^{(1)}\mathbf{V}^{\mathsf{T}} + (1-\alpha)\mathbf{C}^{(1)}\mathbf{C}^{(2)\mathsf{T}},$$

where $0 \le \alpha \le 1$. $\mathbf{A}^{(1)} \in \mathbb{R}^{I \times R_o}$ represents the factor matrix common to both data sets; $\mathbf{A}^{(2)} \in \mathbb{R}^{J \times R_o}$ and $\mathbf{A}^{(3)} \in \mathbb{R}^{K \times R_o}$ are the factor matrices corresponding to the other modes of the tensor while $\mathbf{V} \in \mathbb{R}^{M \times R_o}$ is the factor matrix corresponding to the second mode of matrix $\mathbf{Y}$. R_o indicates the number of common (overlapping) components. Factors that are only present in tensor $\mathcal{X}$ are represented by $\mathbf{B}^{(1)} \in \mathbb{R}^{I \times R_{no}}$, $\mathbf{B}^{(2)} \in \mathbb{R}^{J \times R_{no}}$ and $\mathbf{B}^{(3)} \in \mathbb{R}^{K \times R_{no}}$; factors that are only present in the matrix $\mathbf{Y}$ are represented by $\mathbf{C}^{(1)} \in \mathbb{R}^{I \times R_{no}}$ and $\mathbf{C}^{(2)} \in \mathbb{R}^{M \times R_{no}}$. R_{no} indicates the number of non-overlapping components.

Tensor $\mathcal{X}$ and matrix $\mathbf{Y}$ are constructed based on the above formulation for a set of α values using factor matrices with entries randomly drawn from the standard normal distribution and columns normalized to unit norm.[2] In our experiments, we add noise to tensor $\mathcal{X}$ and construct $\mathcal{X}_{\text{noisy}}$ as $\mathcal{X}_{\text{noisy}} = \mathcal{X} + \eta\frac{\mathcal{N}}{\|\mathcal{N}\|}\|\mathcal{X}\|$, where $\mathcal{N}$ is a tensor of same size as $\mathcal{X}$ with entries drawn from the standard normal distribution. Similarly, $\mathbf{Y}_{\text{noisy}}$ is constructed as $\mathbf{Y}_{\text{noisy}} = \mathbf{Y} + \eta\frac{\mathbf{N}}{\|\mathbf{N}\|}\|\mathbf{Y}\|$, where $\mathbf{N}$ is a matrix of same size as $\mathbf{Y}$ with entries drawn from the standard normal.

3.1.1. Performance measure

In order to assess the performance of data fusion, we randomly set some tensor entries to missing and use missing data recovery error as a performance measure. To set entries to missing, a binary indicator tensor $\mathcal{W}$, defined in (4), is generated such that out of its *IJK* entries, ⌊*SIJK*⌋ entries are set to zero uniformly at random, where *S* indicates the percentage of missing entries.

We estimate missing tensor entries using two different approaches: (i) *Analysis of only the third-order tensor:* modeling the third-order tensor with missing entries using a CP model and then using the CP factor matrices to reconstruct the tensor and recover missing values and (ii) *Coupled Analysis of the third-order tensor and the matrix*: modeling the incomplete third-order tensor jointly with matrix $\mathbf{Y}$ using a CMTF model and then using the CMTF factor matrices to reconstruct the tensor and estimate missing entries. Let the extracted factors by either model from $\mathcal{X}_{\text{noisy}}$ be $\hat{\mathbf{U}}^{(1)} \in \mathbb{R}^{I \times R}$, $\hat{\mathbf{U}}^{(2)} \in \mathbb{R}^{J \times R}$ and $\hat{\mathbf{U}}^{(3)} \in \mathbb{R}^{K \times R}$. In order to quantify the missing data recovery error, we define tensor completion score (TCS), that computes the error between the original and estimated values of missing entries as follows:

$$\text{TCS} = \frac{\left\|(1-\mathcal{W}) * \left(\mathcal{X}_{\text{noisy}} - \hat{\mathcal{X}}\right)\right\|}{\left\|(1-\mathcal{W}) * \mathcal{X}_{\text{noisy}}\right\|}, \tag{6}$$

where $\hat{\mathcal{X}}$ corresponds to the data reconstructed using the extracted factor matrices, i.e., $\hat{\mathcal{X}} = [\![\hat{\mathbf{U}}^{(1)},\hat{\mathbf{U}}^{(2)},\hat{\mathbf{U}}^{(3)}]\!]$.

[2] We do not assume orthogonality among the columns of each factor matrix; however, note that when entries are randomly drawn from standard normal, the data generation scheme may result in nearly orthogonal columns in the factor matrices corresponding to the modes with large dimensions, e.g., 100.

3.1.2. Experimental settings

In our simulations, we generate coupled data sets by varying the following parameters: (i) data set sizes, i.e., $(I,J,K,M) = \{(10,10,10,100),(60,10,6,116),(10,20,10,100)\}$, (ii) noise levels, i.e., $\eta = \{0,0.2,0.4\}$, (iii) α values, i.e., totally common factors ($\alpha = 1$), partially common factors ($\alpha = \{0.75,0.25\}$) and no common factors ($\alpha = 0$), and (iv) the amount of missing data, i.e., $S\% = \{50,70,80,85,90\}$. The number of overlapping components, i.e., R_o, is set to $R_o = 2$, while the number of non-overlapping components, i.e., R_{no}, is set to $R_{no} = 1$. In order to see the performance of the models with different number of components, the number of extracted components, i.e., R, is set to $R = \{1,2,3,4\}$.

We create our test problems as follows. For each data set size, α and noise level η, we generate factor matrices as described in Section 3.1 and construct a third-order tensor coupled with a matrix. Twenty such independent test problems (replicates) are generated for each data set size, α and η value. In each replicate, $S\%$ of the entries in the third-order tensor is then set to missing by generating a binary indicator tensor $\mathcal{W}$ (as described in Section 3.1). Finally, performance of CP and CMTF is assessed based on the estimation of missing entries.

For fitting the CMTF model, we use the all-at-once optimization algorithm described in Section 2. Similarly, for fitting the CP model to an incomplete tensor, we use the CP-WOPT algorithm [34], which is also an all-at-once optimization method solving for all factor matrices simultaneously by fitting the model only to the known entries. As stopping conditions, both algorithms use relative change in function value (set to 10^{-10}) and the two-norm of the gradient divided by the number of entries in the gradient (set to 10^{-10}). Also, maximum number of iterations and function evaluations are set to 10^4 and 10^5, respectively. As we are dealing with a problem with potentially many local minima, we use multiple starts (i.e., svd-based initialization as well as 19 random starts) for both algorithms and choose the starting point that gives the minimum function value.[3]

3.1.3. Results

Fig. 2 shows the missing data recovery error for CP and CMTF for different values of α and different amounts of missing data in terms of tensor completion score for a third-order tensor of size $10 \times 10 \times 10$ coupled with a matrix of size 10×100. We observe that coupled analysis can more accurately recover missing entries for higher α values and high amounts of missing data. For $\alpha = 0$, since there are no common components, it is expected that CMTF is not the right model; therefore, it cannot recover missing entries as accurately as CP. However, we should also note that CMTF performs worse than CP even for $\alpha = 0.25$, i.e., partially common factors, indicating that if common components contribute little to the coupled data sets, we do not see the advantage of coupling in terms of missing data recovery. The reported tensor completion scores in each boxplot correspond to 20 replicates. The plots show the median values as black dots inside a colored circle, the 25th and 75th percentile values are indicated by the top and bottom edges of the solid bars extending from the median values, and the outliers are shown as smaller colored circles. A lack of solid bars extending from the median indicates that the 25th and 75th percentile values are equal to the median. In each replicate, different number of components, i.e., $R = 1,2,3,4$, is tested for each model and we report the best performance of each model (See Appendix A for more details on the number of components used for performance comparison). Here, the noise level is set to $\eta = 0.2$.

[3] We pick the minimum function value only if the algorithm stops due to the relative change in fit (with a sufficiently small gradient) or the gradient condition; otherwise, we consider that the algorithm has not converged.

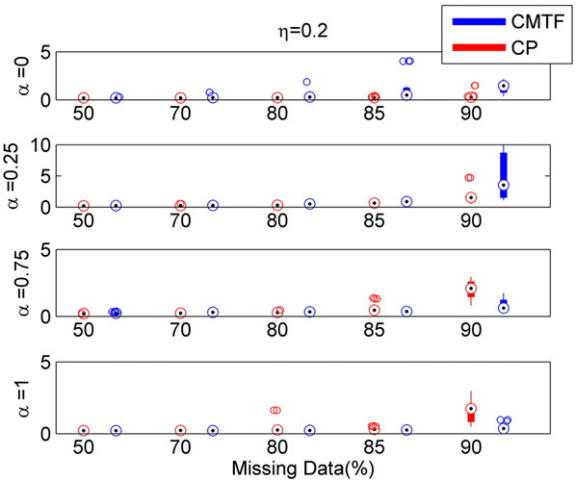

Fig. 2. Tensor completion score of CP and CMTF for different values of α and different amounts of missing data, S, for coupled data sets of size $(I,J,K,M) = (10,10,10,100)$ and noise level $\eta = 0.2$.

3.1.3.1. Noise levels (η)

We observe similar performance comparison for CMTF and CP at different levels of noise. For instance, Fig. 3 shows the tensor completion scores for CP and CMTF for noise-free data sets and noise level $\eta = 0.4$. We observe that jointly analyzing data sets can handle higher amounts of missing data more accurately compared to the CP model only for higher α values, i.e., $\alpha = 0.75$ and $\alpha = 1$, just like in Fig. 2.

Based on the results reported so far, it is important to note that when there are common components in coupled data sets, their joint factorization does not necessarily improve the missing data estimation performance; in other words, there may be common components but the missing data recovery error achieved by coupled analysis can still be worse than the CP model as we have observed in Figs. 2 and 3 for $\alpha = 0.25$.

3.1.3.2. Data set sizes (I,J,K,M)

Relative sizes of the data sets have an impact on the performance of CMTF compared to a CP model. So far we have considered the performance of CP and CMTF in terms of recovering the missing entries in a third-order tensor coupled with a matrix, where the third-order tensor and the matrix have the same number of entries. In Figs. 4 and 5, we illustrate the missing data recovery performance for coupled data sets of different sizes.

Fig. 4 compares the performance of CP and CMTF when the third-order tensor has twice the number entries than the coupled matrix. Just like in the case where the tensor has as many entries as in the matrix, we observe that CMTF performs better than CP at high amounts of missing data for higher α values only.

On the other hand, Fig. 5 demonstrates that when the coupled matrix has almost twice the number of entries than the third-order tensor, i.e., 60×116 matrix coupled with a third-order tensor of size $60 \times 10 \times 6$, the benefit of coupled factorization is even visible for low α values. We observe that CMTF outperforms CP in terms of estimation of missing entries in the third-order tensor at high amounts of missing data even for $\alpha = 0.25$ (Fig. 5(a)–(b)). For high amounts of noise, CMTF outperforms CP only for high α values, though.

3.1.3.3. Number of extracted components (R)

Finally, we illustrate the missing data recovery performance of each model for different number of components and show that the best

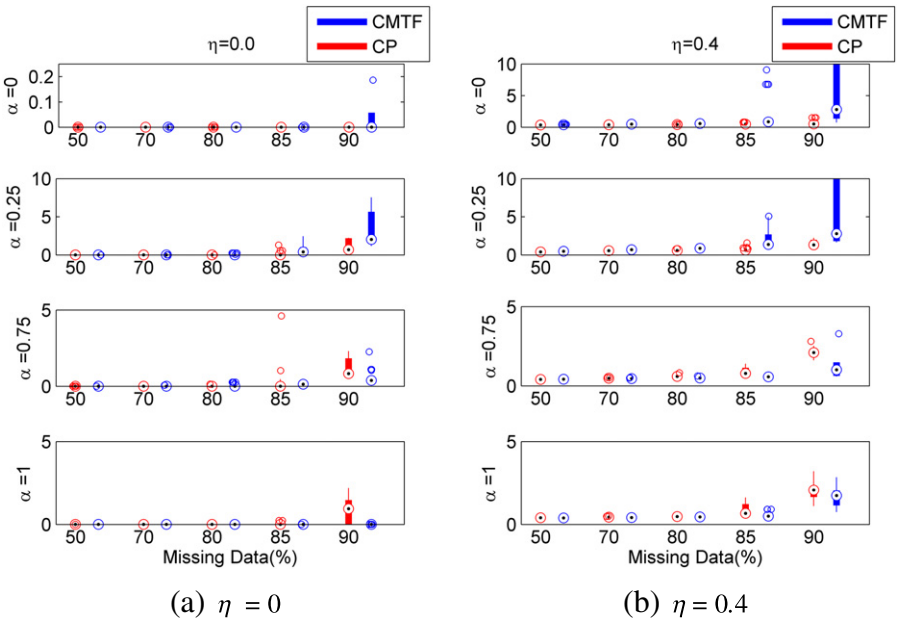

Fig. 3. Tensor completion score of CP and CMTF for coupled data sets of size $(I,J,K,M) = (10,10,10,100)$ at different noise levels.

performing R changes depending on the α value and the model. For instance, for $\alpha = 0$, the tensor has only one non-overlapping component, i.e., $R_{no} = 1$, and no common component; therefore, there is only one underlying component. Fig. 6 shows that extracting only one component achieves the best missing data recovery error for CP while we often need $R = 2$ components to recover the missing entries accurately using CMTF. For $\alpha = 1$, the tensor has two overlapping components, i.e., $R_o = 2$, and no non-overlapping component, which means that, in total, there are two true underlying components. We observe, in Fig. 6, for $\alpha = 1$, $R = 2$ performs the best for both CP and CMTF. For higher amounts of missing data, CP fails while CMTF can still accurately recover missing entries by achieving low missing data recovery error for $R = 2$. Note that some of the bar plots are not shown, e.g., $R = 4$ for all α values for CMTF at 90% missing data, indicating that the algorithm does not converge in those cases. The algorithms fail to converge for difficult problems, i.e., large component number and high amounts of missing data. By increasing the maximum number of iterations and function evaluations by an order of 10, we can decrease the number of non-converging cases. However, this forces the algorithms to run longer for difficult cases and eventually when/if they converge, they return a high tensor completion score. We should point out that the relative performance of the models does not change since these cases are encountered for wrong models, e.g., too many components. See Appendix A for results obtained by increasing the maximum number of iterations and function evaluations.

3.2. Tensor completion performance on real coupled data sets with simulated missingness

In this section we demonstrate the performance of CMTF compared to CP in terms of missing data recovery on real data sets from sensory data analysis and metabolomics.

3.2.1. Sensory data analysis

In sensory data analysis, we often deal with data from multiple sources. For instance, one source of data is the sensory data, i.e., a set of products assessed by several judges based on various attributes. In addition to sensory data, other measurements, such as spectroscopic measurements of products, may also be available. We use the sensory data and the NIR (near-infrared) measurements described in [35]. The sensory data is constructed based on the assessment of 60 pea samples over 6 attributes (i.e., pea flavor, sweetness, fruity flavor, off-flavor, mealiness and hardness) by 10 judges and forms a third-order tensor of the form *products* by *judges* by *attributes*. The NIR data contains 116 spectral values for each of the 60 pea samples and forms a matrix of *products* by *variables*. The data sets share the *products* mode and are coupled in that dimension. For the details of sample preparation and measurements, see [35].

In order to see the performance of CMTF compared to a CP model in terms of missing data recovery, we set sensory data entries to missing uniformly at random (see Section 3.1.1), construct $\overline{\mathcal{X}}$ (as illustrated in Fig. 1a), and estimate missing entries using a CP model of $\overline{\mathcal{X}}$ and a CMTF model of $\overline{\mathcal{X}}$ and $\mathbf{Y}$. We randomly generate 25 such incomplete data sets for each amount of missing data. Fig. 7 shows the average tensor completion score at different amounts of missing data for different number of components, i.e., $R = 1,2,3,4$. We observe that 2-component models almost always give the lowest missing data recovery for both models, especially at high amounts of missing data.

Based on Fig. 7, we pick $R = 2$ for both models and take a closer look at the performance in Fig. 8. We observe that both CP and CMTF perform equally well for low amounts of missing data and can recover the missing values with a TCS of around 0.20. The error increases for CP as the amount of missing entries increases to 85–90%; however, CMTF is still able to keep the missing data recovery error low at such high amounts of missing data.

3.2.2. Metabolomics

Another field, where we commonly encounter data sets from multiple sources, is metabolomics [1]. In metabolomics, biological fluids such as urine and plasma samples are often measured using several analytical methods. In this paper, we consider joint factorization of fluorescence excitation-emission (represented as a third-order tensor) and NMR measurements (in the form of a matrix) of 286 human blood samples. The details of the specific subset used in this paper are given in [36] and the

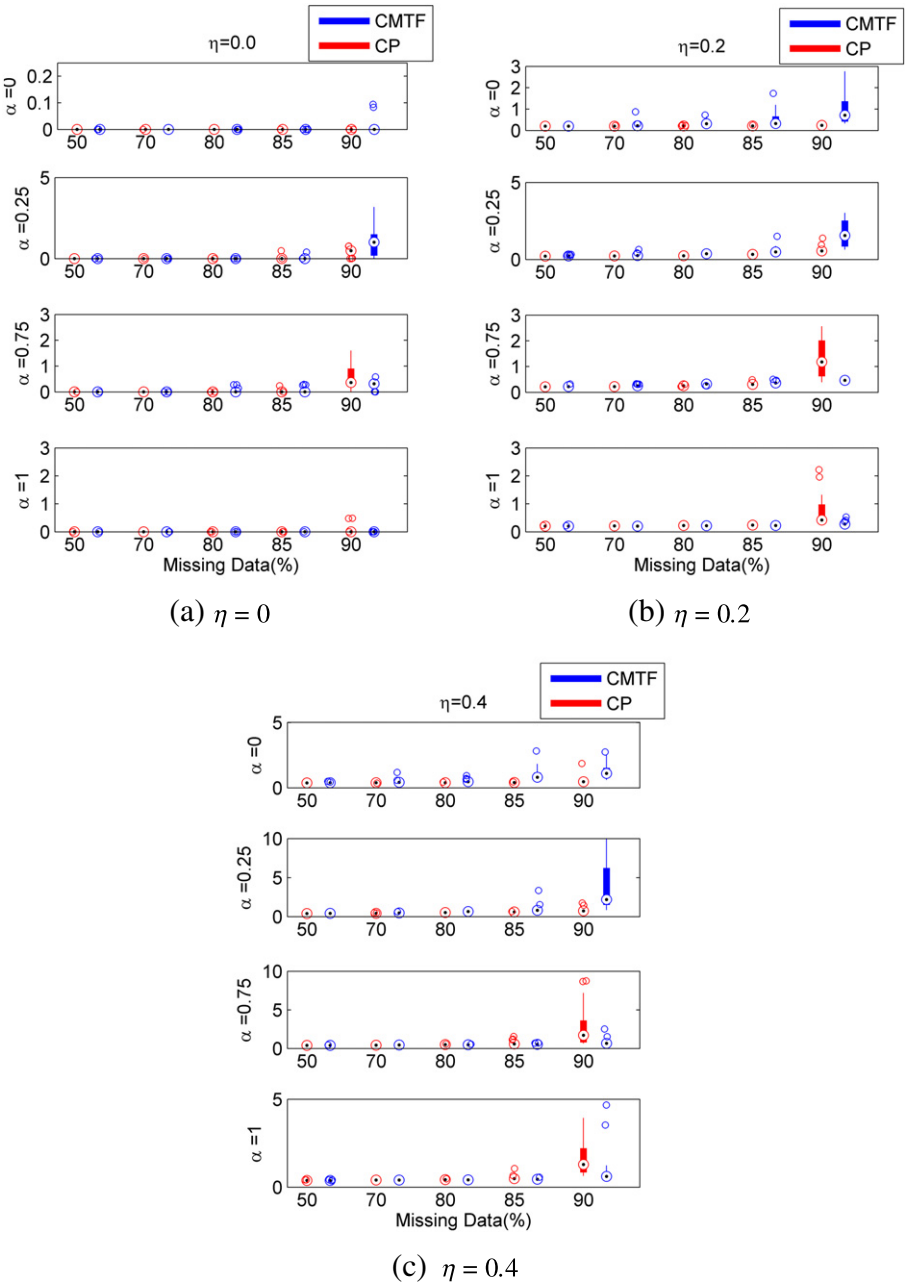

Fig. 4. Tensor completion score of CP and CMTF for different values of α and different amounts of missing data, S, for coupled data sets of size $(I,J,K,M) = (10,20,10,100)$ at different noise levels.

fluorescence measurements are available from http://www.models.life.ku.dk. We have used the undiluted samples measured in the spectral region of emission wavelengths from 300 nm to 600 nm with 20 nm increment and excitation wavelengths from 250 nm to 450 nm with 20 nm increment. This forms a third-order tensor, $\mathcal{X}$, of type 286 *samples* × 16 *emission wavelengths* × 11 *excitation wavelengths*.[4] For the same samples, we have also acquired CPMG ^{1}H-NMR spectra using a 600 MHz spectrometer and formed a *samples* by *chemical shifts* (*ppm*) matrix, $\mathbf{Y}$, of size 286 × 4132. The NMR spectra were preprocessed by aligning [37] and referencing them towards the doublet signal of the anomeric alpha-glucose proton (5.25 ppm) and removing the spectral region containing the suppressed water signal. The matrix was formed by downsampling the raw NMR datascaled using Pareto scaling.

Here, we set tensor entries from fluorescence measurements in $\mathcal{X}$ to missing uniformly at random and construct $\overline{\mathcal{X}}$. We randomly generate 25 such incomplete data sets for varying amounts of missing data. Fig. 9 shows the average tensor completion of CP vs. CMTF for different number of components. We have previously seen in simulation experiments and sensory analysis application that modeling additional source of information makes a difference in terms of missing data estimation for higher amounts of missing data. Therefore, we have only focused

[4] We have used larger increments for excitation and emission spectra compared to the original data set in order to work with a smaller data set and see the advantage of coupled approach in smaller amounts of missing data. Downsampling the spectra emulates a filter-based instrument.

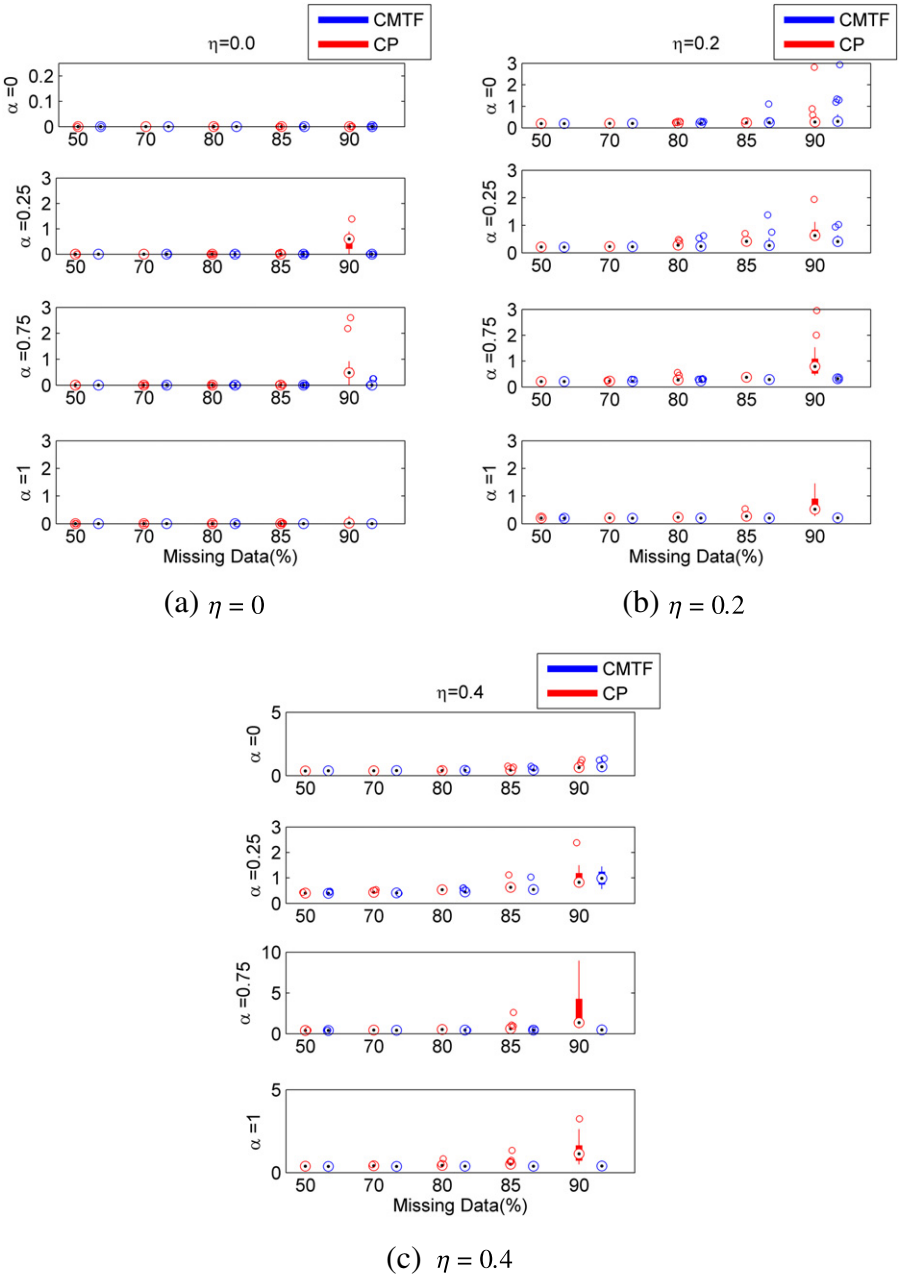

Fig. 5. Tensor completion score of CP and CMTF for different values of α and different amounts of missing data, S, for coupled data sets of size $(I,J,K,M) = (60,10,6,116)$ at different noise levels.

on the 70%–90% interval. The best performing number of components in terms of TCS is either $R = 1$ or $R = 2$. Since $R = 2$ gives slightly lower TCS for both CP and CMTF at 70% and 80%, we show tensor completion scores for different amounts of missing data for CP vs. CMTF using $R = 2$ components[5] in Fig. 10. Fig. 10 illustrates a pattern similar to the one observed for coupled data sets from sensory data analysis, i.e., compared to a CP model, CMTF can handle higher amounts of missing data more accurately.

4. Discussions

In complex problems, we often need to jointly analyze data from multiple sources through data fusion. In this paper, we have studied data fusion using the coupled matrix and tensor factorization formulation, which can jointly analyze data sets of different orders, i.e., both matrices and higher-order tensors.

Data fusion studies typically focus on the gain obtained through fusing data sets, e.g., joint analysis of multiple data sets outperforms the analysis of a single data set in terms of missing data estimation [7,8,16,20,38]. In this paper, instead, we were interested in the

[5] The best performing number of components is not clear in Fig. 9; therefore, we have also inspected the relative behavior of the models for the following cases: (i) $R = 1$ is used for both models, (ii) $R = 1$ for CP and $R = 2$ for CMTF, and (iii) best performing R is picked for each model. In all cases, the relative performance is still similar to the behavior in Fig. 10.

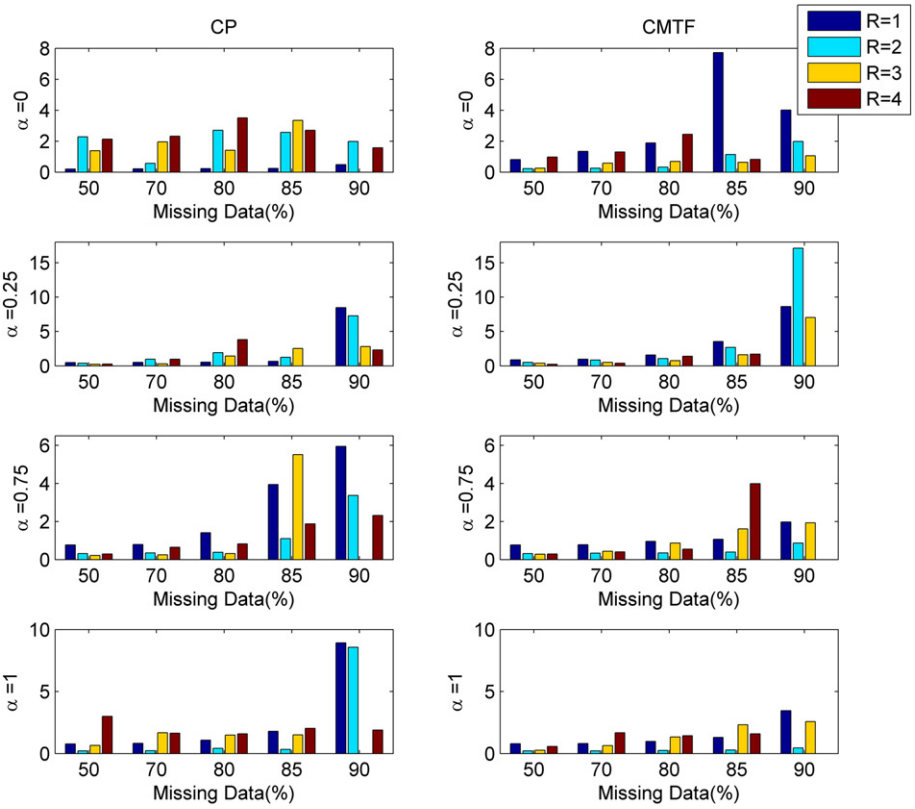

Fig. 6. Average tensor completion score of CP and CMTF for different number of extracted components, R, for coupled data sets of size $(I,J,K,M) = (10,10,10,100)$ and noise level $\eta = 0.2$. Bars are missing (i) for CMTF, at 90% missing data for $R = 4$ and (ii) for CP, at 90% missing data for $R = 3$ (except for $\alpha = 0.25$) and at 85% missing data for $R = 4$ for $\alpha = 0.25$ since the algorithm does not converge in those cases.

limitations as well as the advantages of data fusion and tried to answer the following question: Is it possible that coupled data sets have shared factors and still their joint analysis performs worse than the analysis of a single data set? In order to address that question, we have conducted extensive simulations generating coupled data sets with shared and/or individual components and demonstrated cases where coupled analysis can enhance knowledge discovery in terms of missing data estimation as well as cases where it fails to do so due to the underlying structures of coupled data sets. For instance,

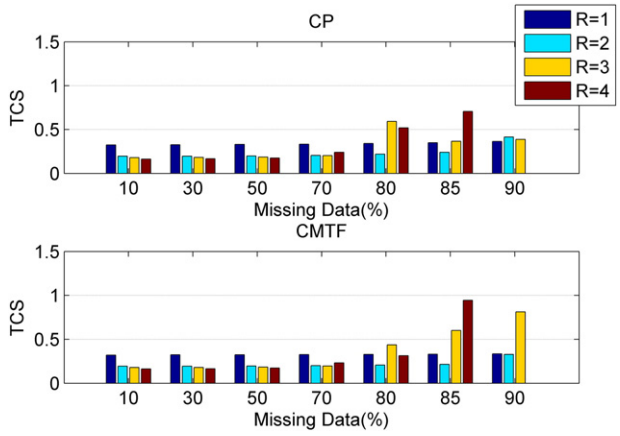

Fig. 7. Average tensor completion score for different number of components for CP (of sensory data) and CMTF (of sensory data coupled with NIR). Bars are missing for both CMTF and CP at 90% missing data for $R = 4$ since the algorithm does not converge in those cases.

coupled analysis performs worse than the analysis of a single data set when the underlying common factors constitute only a small part of the data under inspection. We have further studied the relative performance of models under different noise levels and data set sizes. While the relative performance of models is similar across different noise levels, data set size of the additional data set used to help with missing data recovery has a clear impact. For instance, coupled matrix and tensor factorization performs worse than a CP model in terms of tensor completion if the coupled data sets have the same number of entries and shared factors only contribute little to each data set. On the other hand, if the matrix has twice the number of entries than the third-order tensor, CMTF outperforms CP even if the shared factors are still only a small part of the data. In summary, numerical experiments demonstrate that (i) data fusion often (but not necessarily) enhances knowledge discovery and (ii) it is indeed possible to have coupled data sets with shared factors and their joint analysis does not improve missing data estimation.

While this paper demonstrates potential limitations of coupled analysis that have not been discussed before, we should acknowledge that we have only explored data fusion in a limited setting. This study can be extended in many ways, e.g., estimation of missing entries in both the matrix and the third-order tensor, different weighting schemes in the multi-objective of the data fusion problem and different data generation schemes using different α values in each data set. For instance, a more general situation would be to investigate the missing data estimation performance using data sets generated based on the following set-up:

$$\mathcal{X} = \alpha_t [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]\!] + (1-\alpha_t)[\![\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \mathbf{B}^{(3)}]\!], \tag{7}$$

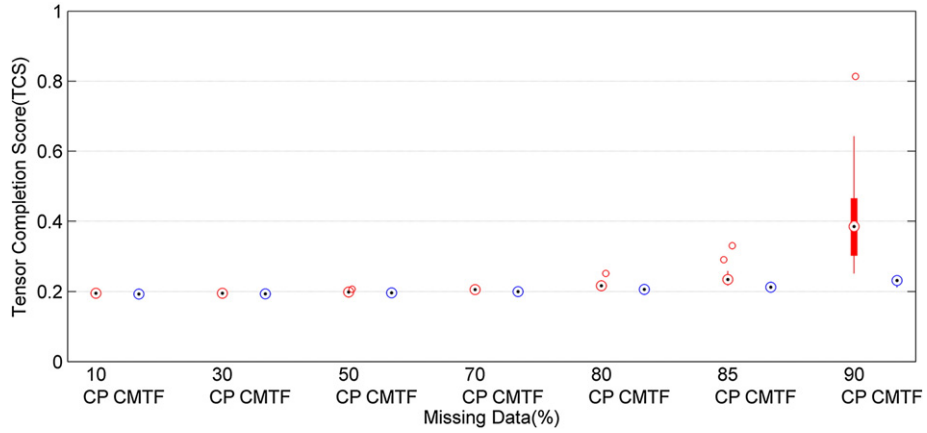

Fig. 8. Tensor completion score for randomly missing entries using CP of sensory data vs. CMTF of sensory data coupled with NIR.

$$\mathbf{Y} = \alpha_m \mathbf{A}^{(1)}\mathbf{V}^{\mathsf{T}} + (1-\alpha_m)\mathbf{C}^{(1)}\mathbf{C}^{(2)\mathsf{T}},$$

where $\alpha_t \neq \alpha_m$. In our experiments, we have used $\alpha_t = \alpha_m$ and model the generated coupled data sets using CMTF as in (3), where both parts of the objective function, i.e., $f_{\mathcal{W}_1}$ and $f_{\mathcal{W}_2}$, are given equal weights. In Figs. 11 and 12, we illustrate tensor completion scores obtained using a CP model of a third-order tensor of size $10 \times 10 \times 10$ vs. the TCS obtained by a CMTF model of the third-order tensor coupled with a matrix of size 10×100 at different amounts of missing data, where noise level is $\eta = 0.2$. Unlike in Fig. 2, here, we report the results where data sets are generated with unequal α_t and α_m values, i.e., $\alpha_t = 0.75$, $\alpha_m = 0.25$ in Fig. 11 and $\alpha_t = 0.25$, $\alpha_m = 0.75$ in Fig. 12. The top plot in each figure shows that CMTF still performs worse than CP, similar to our results in Fig. 2. However, we should note that we can get completely different results by weighting $f_{\mathcal{W}_1}$ and $f_{\mathcal{W}_2}$ differently. For instance, the bottom plot in Figs. 11 and 12 shows the performance when $f_{\mathcal{W}_2}$ and $f_{\mathcal{W}_1}$, respectively, is given three times more importance by weighting that specific part of the objective more heavily. Besides, while we have studied data fusion formulated as CMTF, alternative formulations using different loss functions and/or tensor factorization models can also be of interest [16].

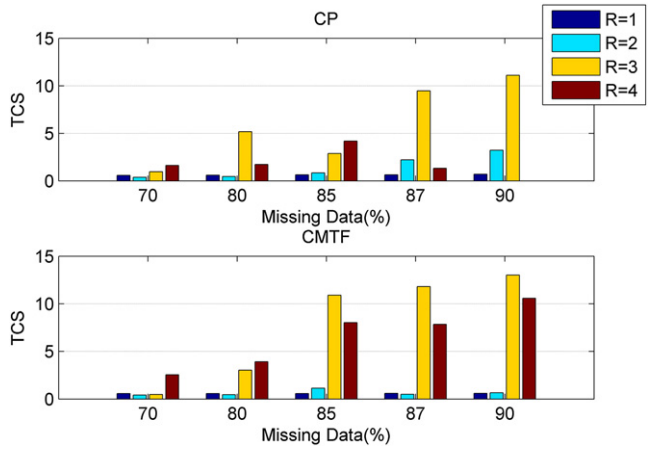

Fig. 9. Average tensor completion score for different number of components for CP (of fluorescence data) and CMTF (of fluorescence data coupled with NMR). The bar corresponding to $R = 4$ at 90% missing data for CP is missing since the algorithm fails to converge.

Coupling data sets with structured missing data is definitely another area of interest [16,38]. For instance, we may have a tensor with missing slices. It is not possible to reconstruct a missing slice of a tensor using its low-rank approximation. Coupled analysis, on the other hand, may prove useful in such cases since it makes use of an additional source of information [16,38].

Since our focus so far has been missing data recovery (for missing data estimation, one does not need a unique model), we have not discussed the uniqueness properties of the CMTF model. When CP model is used in the CMTF formulation, it preserves its uniqueness properties in coupled analysis [19]; however, that is limited to the case where all factors are shared by the matrix and the tensor. We plan to study uniqueness properties of CMTF with overlapping and non-overlapping components in future studies.

Finally, while we have studied data fusion in terms of its missing data estimation performance in this paper, it is also important to understand data fusion models in terms of their performance in providing robustness in the case of noisy data sets as well as identification of the underlying factors in coupled data sets.

Acknowledgments

This work is funded by the Danish Council for Independent Research (DFF) — Technology and Production Sciences (FTP) Program under the projects 11-116328 and 11-120947. The contribution of Tormod Næs is partly financed by the project "Data Integration" supported by the Research Council of Norway.

Appendix A

In Section 3, we report the best performance of each model across different number of components in order not to influence the performance comparison with the potential random effect of the model selection approach. Here, we include additional results using the number of components which one may be tempted to select based on the experimental set-up used in the simulations. The true number of components for the CP model is 1, 3, 3 and 2 for $\alpha = 0, 0.25, 0.75$ and 1, respectively. The true number of components for CMTF, on the other hand, is not clear except for $\alpha = 1$. We may set the number of components based on the total number of available components, i.e., the sum of non-overlapping components in each data set and the overlapping components. In that case, the number of components for CMTF can be

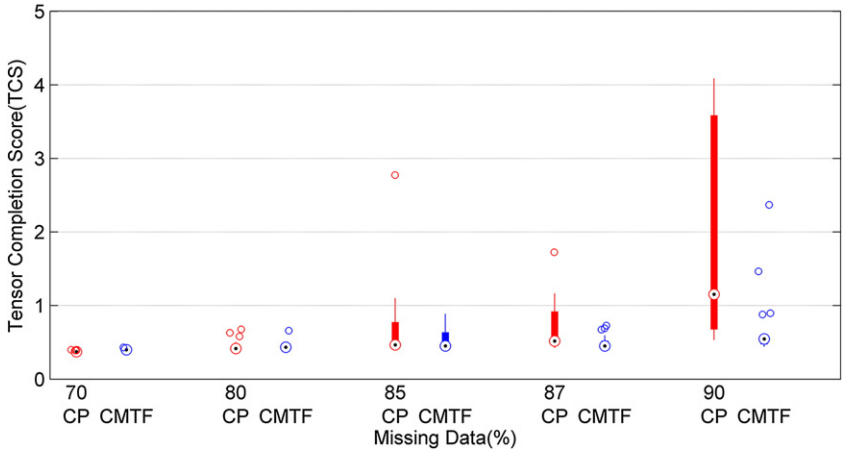

Fig. 10. Tensor completion score for randomly missing entries using CP of fluorescence data vs. CMTF of fluorescence data coupled with NMR.

picked as 2, 4, 4 and 2 for $\alpha = 0, 0.25, 0.75$ and 1, respectively. Note that the CMTF model is based on the idea of modeling only overlapping components; therefore, the total number of available components is not necessarily the true number of components for CMTF.

When these specific number of components are used, as we observe in Fig. 6, the algorithm fails to converge at high amounts of missing data, e.g., for $\alpha = 0.75$, $R = 3$ fails for CP while $R = 4$ fails for CMTF at 90% missing data. We run the algorithm longer by increasing the maximum number of iterations and function evaluations to 10^5 and 10^6, respectively, and Fig. A.13 shows the missing data recovery error for CP and CMTF for a third-order tensor of size $10 \times 10 \times 10$ coupled with a matrix of size 10×100 at noise level $\eta = 0.2$. In Fig. A.13(a), we compare the models by using the best performing R for each model across $R = 1,2,3,4$ as we have done in Section 3. The relative performances of the models are similar to the results in Fig. 2. Fig. A.13(b), on the other hand, demonstrates the tensor completion scores when the specific number of components given in the previous paragraph are used. In this case, even though it looks like CMTF is improving the performance for $\alpha = 0.25$ and $\alpha = 0.75$, when we pay attention to the y-axis, we observe that both CP and CMTF are performing extremely poorly and CMTF is useful only for $\alpha = 1$. This plot still enables us to observe that coupled data sets may have common components but we may not see the effect of coupling unless α is high. Nevertheless, we get a more complete picture of the performance of the models by choosing the best performing R. In practice, as it is not possible to know the best performing R, we need to rely on model selection schemes.

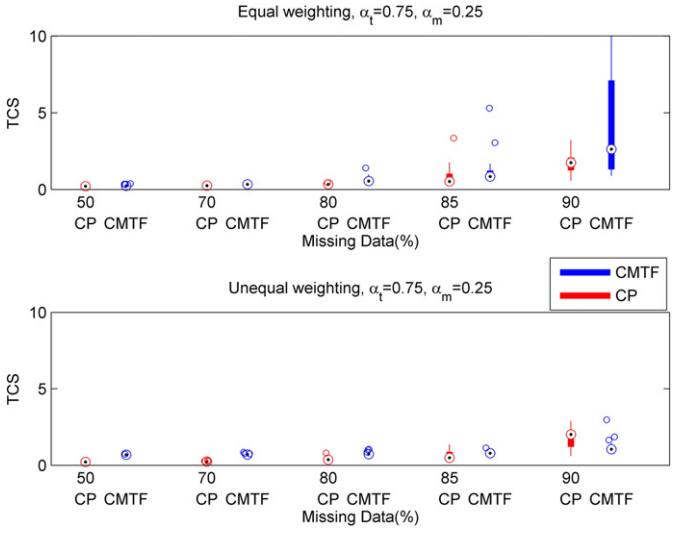

Fig. 11. Tensor completion score for CP vs. CMTF for coupled data sets of size $(I,J,K, M) = (10,10,10,100)$ at noise level $\eta = 0.2$ with $\alpha_t = 0.75$ and $\alpha_m = 0.25$ using equal and unequal weighting schemes.

References

[1] S.E. Richards, M.-E. Dumas, J.M. Fonville, T.M. Ebbels, E. Holmes, J.K. Nicholson, Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework, Chemometrics and Intelligent Laboratory Systems 104 (2010) 121–131.
[2] I. Måge, E. Menichelli, T. Næs, Preference mapping by po-pls: separating common and unique information in several data blocks, Food Quality and Preference 24 (2012) 8–16.
[3] H. Hotelling, Relations between two sets of variates, Biometrika 28 (1936) 321–377.
[4] J. Levin, Simultaneous factor analysis of several Gramian matrices, Psychometrika 31 (1966) 413–419.
[5] J.A. Westerhuis, T. Kourti, J.F. Macgregor, Analysis of multiblock and hierarchical PCA and PLS models, Journal of Chemometrics 12 (1998) 301–321.
[6] B. Long, Z.M. Zhang, X. Wu, P.S. Yu, Spectral clustering for multi-type relational data, ICML'06: Proc. 23rd Int. Conf. on Machine Learning, 2006, pp. 585–592.
[7] H. Ma, H. Yang, M.R. Lyu, I. King, SoRec: social recommendation using probabilistic matrix factorization, CIKM'08: Proc. 17th ACM Conf. on Information and Knowledge Management, 2008, pp. 931–940.
[8] A.P. Singh, G.J. Gordon, Relational learning via collective matrix factorization, KDD'08: Proc. 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2008, pp. 650–658.
[9] N.M. Correa, T. Adali, Y.-O. Li, V.D. Calhoun, Canonical correlation analysis for data fusion and group inferences, IEEE Signal Processing Magazine 27 (2010) 39–50.
[10] J. Yoo, M. Kim, K. Kang, S. Choi, Nonnegative matrix partial co-factorization for drum source separation, ICASSP'10: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal, 2010, pp. 1942–1945.
[11] A. Banerjee, S. Basu, S. Merugu, Multi-way clustering on relation graphs, SDM'07: Proc. SIAM Int. Conf. on Data Mining, 2007, pp. 145–156.
[12] I.V. Mechelen, A.K. Smilde, A generic linked-mode decomposition model for data fusion, Chemometrics and Intelligent Laboratory Systems 104 (1) (2010) 83–94.
[13] A. Smilde, J.A. Westerhuis, R. Boque, Multiway multiblock component and covariates regression models, Journal of Chemometrics 14 (2000) 301–331.
[14] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, A. Kelliher, MetaFac: community discovery via relational hypergraph factorization, KDD'09: Proc. 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2009, pp. 527–536.
[15] V.W. Zheng, B. Cao, Y. Zheng, X. Xie, Q. Yang, Collaborative filtering meets mobile recommendation: A user-centered approach, AAAI'10: Proc. 24th Conf. on Artificial Intelligence, 2010, pp. 236–241.
[16] B. Ermis, E. Acar, A.T. Cemgil, Link prediction via generalized coupled tensor factorisation, ECML/PKDD Workshop on Collective Learning and Inference on Structured Data, 2012, (arXiv:1208.6231).
[17] E. Acar, G.E. Plopper, B. Yener, Coupled analysis of in vitro and histology tissue samples to quantify structure–function relationship, PLoS One 7 (3) (2012) e32227.
[18] T.F. Wilderjans, E. Ceulemans, I.V. Mechelen, Simultaneous analysis of coupled data blocks differing in size: a comparison of two weighting schemes, Computational Statistics and Data Analysis 53 (2009) 1086–1098.

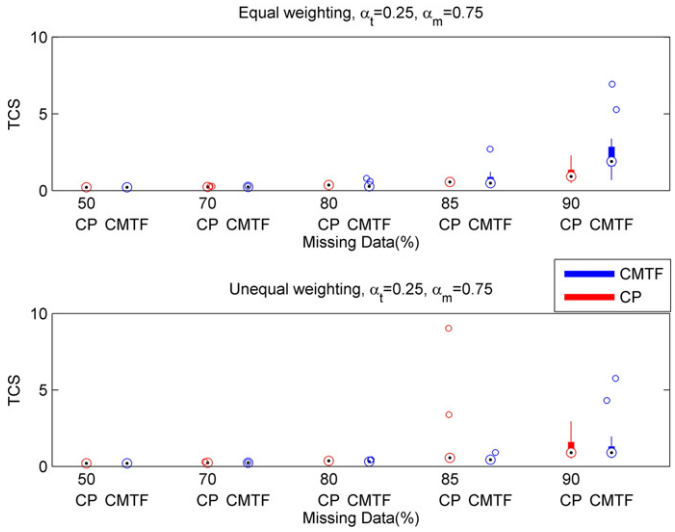

Fig. 12. Tensor completion score for CP vs. CMTF for coupled data sets of size $(I,J,K,M) = (10,10,10,100)$ at noise level $\eta = 0.2$ with $\alpha_t = 0.25$ and $\alpha_m = 0.75$ using equal and unequal weighting schemes.

[19] T.F. Wilderjans, E. Ceulemans, H.A.L. Kiers, 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 (2009) 1073–1082.

[20] E. Acar, T.G. Kolda, D.M. Dunlavy, All-at-once optimization for coupled matrix and tensor factorizations, KDD Workshop on Mining and Learning With Graphs, 2011, (arXiv:1105.3422).

[21] F.L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, Journal of Mathematics and Physics 6 (1) (1927) 164–189.

[22] J.D. Carroll, J.J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition, Psychometrika 35 (1970) 283–319.

[23] R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis, UCLA Working Papers in Phonetics, 16, 1970, pp. 1–84.

[24] A. Smilde, R. Bro, P. Geladi, Multi-way Analysis: Applications in the Chemical Sciences, Wiley, West Sussex, England, 2004.

[25] E. Acar, B. Yener, Unsupervised multiway data analysis: a literature survey, IEEE Transactions on Knowledge and Data Engineering 21 (1) (2009) 6–20.

[26] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.

[27] R.A. Harshman, M.E. Lundy, PARAFAC: parallel factor analysis, Computational Statistics and Data Analysis 18 (1994) 39–72.

[28] H. Shan, Probabilistic Models for Multi-relational Data Analysis. Ph.D. thesis The University of Minnesota, 2012.

[29] G. Tomasi, R. Bro, A comparison of algorithms for fitting the PARAFAC model, Computational Statistics and Data Analysis 50 (7) (2006) 1700–1734.

[30] E. Acar, D.M. Dunlavy, T.G. Kolda, A scalable optimization approach for fitting canonical tensor decompositions, Journal of Chemometrics 25 (2011) 67–86.

[31] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, 2006.

[32] D.M. Dunlavy, T.G. Kolda, E. Acar, Poblano v1.0: a Matlab toolbox for gradient-based optimization, Tech. Rep. SAND2010-1422, Sandia National Laboratories, 2010.

[33] U. Simsekli, B. Ermis, A.T. Cemgil, E. Acar, Optimal weight learning for coupled tensor factorization with mixed divergences, EUSIPCO'13, 2013.

[34] E. Acar, D. Dunlavy, T. Kolda, M. Morup, Scalable tensor factorizations for incomplete data, Chemometrics and Intelligent Laboratory Systems 106 (2011) 41–56.

[35] T. Næs, B. Kowalski, Predicting sensory profiles from external instrumental measurements, Food Quality and Preference 1 (1989) 135–147.

[36] A.J. Lawaetz, R. Bro, M. Kamstrup-Nielsen, I.J. Christensen, L.N. Jørgensen, H.J. Nielsen, Fluorescence spectroscopy as a potential metabonomic tool for early detection of colorectal cancer, Metabolomics 8 (2012) S111–S121.

[37] F. Savorani, G. Tomasi, S.B. Engelsen, icoshift: a versatile tool for the rapid alignment of 1D NMR spectra, Journal of Magnetic Resonance 202 (2010) 190–202.

[38] A. Narita, K. Hayashi, R. Tomioka, H. Kashima, Tensor factorization using auxiliary information, ECML PKDD'11: Proc. European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2011, pp. 501–516.

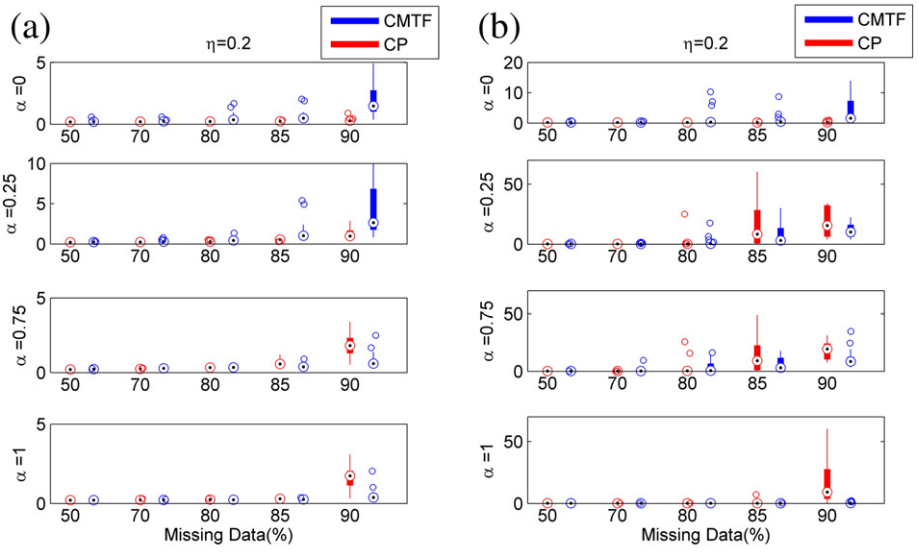

Fig. A.13. Tensor completion score of CP and CMTF for coupled data sets of size $(I,J,K,M) = (10,10,10,100)$ at $\eta = 0.2$ using (a) the best performing R across $R = 1,2,3,4$ and (b) specific R values (given in text for each model).