Unsupervised Multiway Data Analysis: A Literature Survey

Evrim Acar
Department of Computer Science
Rensselaer Polytechnic Institute
acare@cs.rpi.edu

Bulent Yener
Department of Computer Science
Rensselaer Polytechnic Institute
yener@cs.rpi.edu

ABSTRACT

Multiway data analysis captures multilinear structures in higher-order datasets, where data have more than two modes. Standard two-way methods commonly applied on matrices often fail to find the underlying structures in multiway arrays. With increasing number of application areas, multiway data analysis has become popular as an exploratory analysis tool. We provide a review of significant contributions in literature on multiway models, algorithms as well as their applications in diverse disciplines including chemometrics, neuroscience, computer vision, and social network analysis.

1. INTRODUCTION

Multiway data analysis, originating in psychometrics back in the sixties [Tucker 1964], is the extension of two-way data analysis to higher-order datasets. Multiway analysis is often used for extracting hidden structures and capturing underlying correlations between variables in a multiway array. For example, multi-channel electroencephalogram (EEG) data are commonly represented as an $m \times n$ matrix containing signals recorded for m time samples at n electrodes. In order to discover hidden brain dynamics, often frequency content of the signals, for instance signal power at p particular frequencies, also needs to be considered. In that case, EEG data can be arranged as an $m \times p \times n$ three-way dataset [Miwakeichi et al. 2004]. Multiway analysis of a three-way EEG array then enables us to extract the signatures of brain dynamics in time, frequency and electrode domains. As opposed to two-way analysis showing brain activities at certain time periods at certain electrodes, multiway analysis can differentiate between the brain activities with different spectral signatures.

It has been shown in numerous research areas including social networks [Acar et al. 2005], neuroscience [Estienne et al. 2001], process analysis [Gourvnec et al. 2005] that underlying information content of the data may not be captured accurately or identified uniquely by two-way analysis methods. Two-way analysis methods, e.g. factor models, suffer from rotational freedom unless specific constraints such as statistical independence, orthogonality, etc. are enforced. On the other hand, these constraints requiring prior knowledge or unrealistic assumptions are not often necessary for multiway models. For example, in fluorescence spectroscopy, a Parallel Factor Analysis (PARAFAC) model can uniquely identify the pure spectra of chemicals from measurements of mixtures of chemicals. Consequently, multiway analysis with advantages over two-way analysis in terms of uniqueness, robustness to noise, ease of interpretation, etc. has been a popular exploratory analysis tool in a variety of application areas, which we discuss throughout this survey.

1.1 Multiway Arrays

First difference between two-way and multiway data analysis is the format of the data being analyzed. Multiway arrays, often referred to as tensors, are higher-order generalizations of vectors and matrices. Higher-order arrays are represented as $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \ldots \times I_N}$, where the order of $\underline{\mathbf{X}}$ is N ($N > 2$) while a vector and a matrix is an array of order 1 and 2, respectively.

Higher-order arrays have a different terminology compared to that of two-way datasets. Each dimension of a multiway array is called *a mode (or a way)* and the number of variables in each mode is used to indicate the dimensionality of a mode. For instance, $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \ldots \times I_N}$ is a multiway array with N modes (called N-way array or N^{th} order tensor) with I_1, I_2, ... dimensions in the first, second, etc. mode, respectively. Each entry of $\underline{\mathbf{X}}$ is denoted by $x_{i_1 i_2 \ldots i_N}$. For a special case, where $N = 3$, let $\underline{\mathbf{X}} \in R^{I_1 \times I_2 \times I_3}$ be a three-way array. Then $x_{i_1 i_2 i_3}$ denotes the entry in the i_1^{th} row, i_2^{th} column and i_3^{th} tube of $\underline{\mathbf{X}}$ (Figure 1 [Bader and Kolda 2006a]). When an index is fixed in one mode and the indices vary in the two other modes, this data partition is called *a slice (or a slab)* in higher-order terminology. For example, when the i^{th} row of $\underline{\mathbf{X}}$ is fixed, then it is a horizontal slice of size $I_2 \times I_3$ or similarly, if the j^{th} column of $\underline{\mathbf{X}}$ is fixed, it is a vertical slice of size $I_1 \times I_3$, etc. (Figure 2). A common trend in multiway terminology and notation is to follow the guidelines outlined in [Kiers 2000].

1.2 Models

So far, we have briefly introduced the type of data being analyzed by multiway analysis techniques. These types of data require extensions to analysis methods already available for two-way data analysis. In general, multiway data analysis methods are generalizations of two-way analysis techniques based on the idea of factor models.

A model, which is an approximation of data, consists of two parts: a structural part describing the structure in data and a residual part expressing the part of the data, which cannot be captured by the structural part. Analysis of residuals quantifies how well a model fits the data. More formally, sum of squares of residuals often accounts for the unexplained variation in a least squares sense. Model fit is then

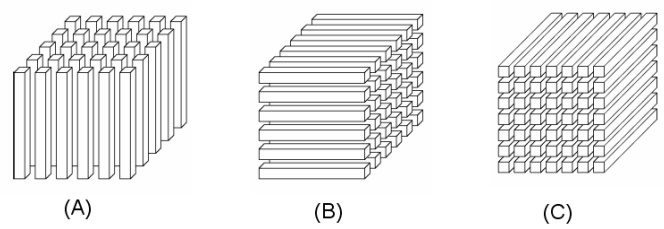

Figure 1: (A) Columns, (B) Rows, (C) Tubes.

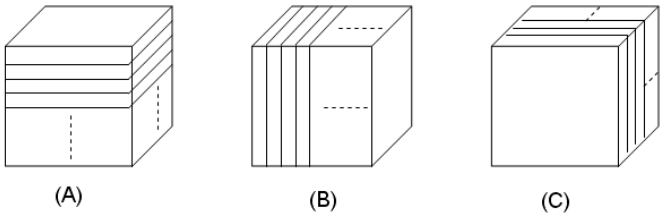

Figure 2: (A) Horizontal Slices, (B) Vertical Slices, (C) Frontal Slices.

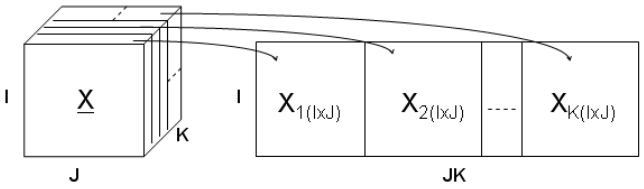

Figure 3: Matricization of a three-way array in the first mode. A three-way array $\underline{\mathbf{X}} \in R^{I \times J \times K}$ is unfolded in the first mode and a matrix of size $I \times JK$, denoted by $\mathbf{X}_{(1)}$ is formed. Subscript in $\mathbf{X}_{(i)}$ indicates the mode of matricization.

defined as the ratio of explained variation in the structural part to the total variation in data. Using bilinear or multilinear models, *factors (or components, loadings)*, which are linear combinations of variables, are extracted. These factors are later used to interpret the underlying information content of the data.

While most multiway analysis techniques preserve the multiway nature of the data, some techniques such as Tucker1 [Tucker 1964; Tucker 1966] are based on matricization of a multiway array, which means transforming a third or higher-order array into a two-way dataset (Figure 3). *Matricization (or unfolding, flattening)* has multiple definitions in literature [Kiers 2000; Lathauwer et al. 2000] but the definition in [Kiers 2000] is commonly followed. Once a three-way array is flattened and arranged as a two-way dataset, two-way analysis methods, e.g. Singular Value Decomposition (SVD) [Golub and Loan 1996] and other factor models [Kim and Mueller 1978], can be employed in understanding the structure in data.

Rearranging multiway arrays as two-way datasets and analyzing them with two-way methods, though, may result in information loss and misinterpretation especially if the data are noisy. An intuitive example is often given on a sensory dataset, where eight judges evaluates ten breads based on eleven attributes [Bro 1998]. When this dataset is modeled using a PARAFAC model, the model assumes that there is a common sense of evaluation among judges and each judge pertains to this sense of evaluation at different amounts. On the other hand, when sensory data are unfolded in bread mode and modeled using a two-way factor model, there is no assumption being made about a common sense of evaluation. Every judge may behave completely independently. In such a scenario, a two-way factor model may extract as many factors as possible to explain the variation in data. However, a PARAFAC model can only explain the variation that follows the basic assumption. Extra variation captured by a two-way factor model might actually explain noise rather than a certain structure. Thus, multiway models are more advantageous in terms of interpretation and accuracy compared to two-way models. Multilinear models (i.e. PARAFAC [Harshman 1970], Tucker [Tucker 1964; Tucker 1966] and their derivatives) capture the multilinear structure in data. Multilinearity of the model denotes that the model is linear in each mode and factors extracted from each mode are linear combinations of the variables in that mode. A component matrix, whose columns are the factors determined by the model, is then constructed to summarize the structure in each mode. These models have been applied on various datasets shown to contain multilinear structure, e.g. three-way fluorescence spectroscopic datasets with modes: *samples* $\times$ *emission* $\times$ *excitation* [Andersen and Bro 2003] or wavelet-transformed multi-channel EEG arranged as a three-way array with modes: *frequency* $\times$ *time samples* $\times$ *electrodes* [Miwakeichi et al. 2004; Acar et al. 2007].

The organization of this paper is as follows: We categorize multiway models (both original models and recent derivatives of original models) and study their similarities and differences in Section 2. In Section 3, we give a brief overview of the algorithms used in fitting these models to multiway datasets. Multiway data analysis applications from several fields and available software tools are summarized in Section 4 and Section 5. After a brief summary of the survey in Section 6, future research directions, particularly in computer science, are discussed in Section 7.

2. MULTIWAY MODELS

The most well-known and commonly applied multiway models in literature are Tucker models and the PARAFAC model, which is also called CANDECOMP (Canonical Decomposition) (CANDECOMP [Carroll and Chang 1970] was proposed independently but considered equivalent to PARAFAC). We will briefly describe these models as well as recent models built on the principles of PARAFAC and Tucker. This survey studies multiway models under three categories given in Figure 4. First category describes PARAFAC and other models, which have relaxed the restrictions enforced by a PARAFAC model to capture data-specific structures. Second category contains the models that belong to Tucker family as well as the extensions of Tucker models. Last category includes the models, which fall under neither the first nor the second category but still address the problem of analyzing multiway arrays. In spite of the categorization, models in different families are closely related to each other, e.g. PARALIND can be considered as a constrained version of a Tucker3 model. This categorization is primarily for the

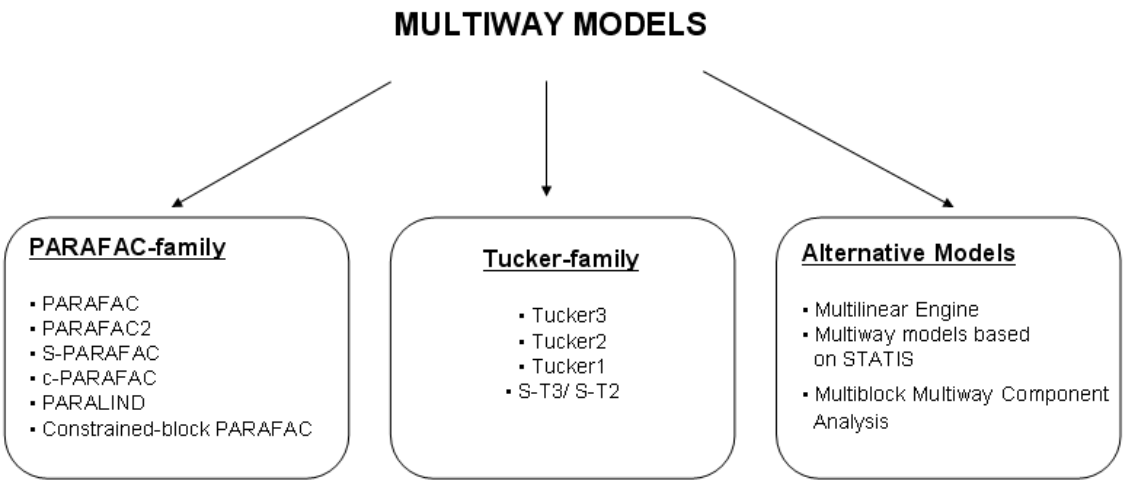

Figure 4: The categorization of multiway models studied in this survey.

ease of presentation and understanding of the models. In the rest of the paper, we discuss these models in the context of three-way arrays but most of these models (Table 1, Table 2) have already been extended to N-way arrays .

2.1 PARAFAC-family

2.1.1 PARAFAC

PARAFAC [Harshman 1970] is an extension of bilinear factor models to multilinear data. It is based on Cattell's principle of Parallel Proportional Profiles [Cattell 1944]. The idea behind Parallel Proportional Profiles is that if the same factors are present in two samples under different conditions, then each factor in the first sample is expected to have the same pattern in the second sample but these patterns will be scaled depending on the conditions. Mathematically, a PARAFAC model can be represented as the decomposition of a tensor as a linear combination of rank-1 tensors (An N^{th} order rank-1 tensor is a tensor that can be written as the outer product of N vectors). Let $\underline{\mathbf{X}} \in R^{I \times J \times K}$ be a three-way array. Then an R-component PARAFAC model can be expressed as in Equation 1 (or Equation 2), where a_i, b_i and c_i indicate the i^{th} column of component matrices $\mathbf{A} \in R^{I \times R}$, $\mathbf{B} \in R^{J \times R}$ and $\mathbf{C} \in R^{K \times R}$, respectively. $\underline{\mathbf{E}} \in R^{I \times J \times K}$ is a three-way array containing the residuals. x_{ijk} represents an entry of a three-way array $\underline{\mathbf{X}}$ in the i^{th} row, j^{th} column and k^{th} tube while a_{ij} is a matrix entry located in the i^{th} row and j^{th} column.

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r + \underline{\mathbf{E}} \tag{1}$$

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \tag{2}$$

The symbol $\circ$ denotes the outer product of vectors. Vector outer product is defined as follows. Let a, b and c be column vectors of size $I \times 1$ and $J \times 1$ and $K \times 1$ and $\underline{\mathbf{Y}}$ is a tensor of size $I \times J \times K$, then $\underline{\mathbf{Y}} = a \circ b \circ c$ if and only if $y_{ijk} = a_i b_j c_k$.

Illustration of a 3-component PARAFAC model on a three-way dataset is also given in Figure 5.

The motivation behind PARAFAC is to obtain a unique solution such that component matrices are determined uniquely up to a permutation and scaling of columns. It is this uniqueness property that makes PARAFAC a popular technique in various fields. For example in fluorescence spectroscopic data analysis [Andersen and Bro 2003], a unique PARAFAC model allows us to find physically and chemically meaningful factors directly from measurements of mixtures of chemicals. Uniqueness is achieved by the restrictions imposed by the model. The most significant restriction is that factors in different modes can only interact factorwise. The interaction between factors in different modes are represented by a core array in multiway models. For a three-way model, the core array is a third-order tensor, $\underline{\mathbf{G}}$ as given in Figure 6, where g_{pqr} represents the interaction of the p^{th} factor in the first, q^{th} factor in the second and r^{th} factor in the third mode. In a PARAFAC model, the core array is restricted to be a super-diagonal core array. For instance for a three-way PARAFAC mode, a super-diagonal core indicates that i^{th} factor in the first mode (a_i) can only interact with i^{th} factor in the second (b_i) and the third mode (c_i). As a consequence of this restriction, the same number of factors should be extracted in each mode. There are several techniques for determining the number of factors in a PARAFAC model, i.e. residual analysis, visual appearance of loadings, number of iterations of the algorithm, core consistency [Bro and Kiers 2003], etc. Among these techniques, core consistency diagnostic quantifies the resemblance between a Tucker3 core and a super-diagonal PARAFAC core and suggests whether a PARAFAC model is a valid model for the data. Core consistency diagnostic has been commonly applied in literature [Estienne et al. 2001; Andersen and Bro 2003; Miwakeichi et al. 2004]. However, there is no bulletproof way to determine the optimal number of factors (optimal in terms of interpretation) for real data. Therefore, it is often suggested that several diagnostic tools are used together rather than a single method [Andersen and Bro 2003; Bro and Kiers 2003].

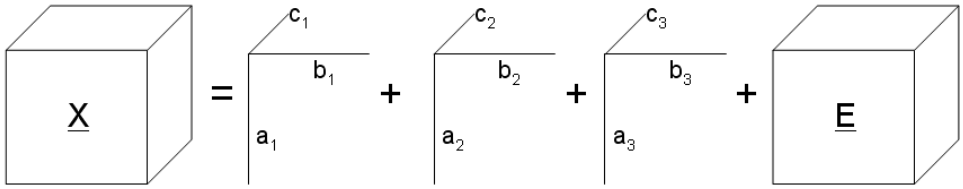

Figure 5: Illustration of a PARAFAC model. 3-component PARAFAC model, where a three-way array $\underline{\mathbf{X}}$ is expressed as the sum of three rank-1 tensors. a_i, b_i and c_i are the i^{th} components in the first, second and third mode, respectively. $\underline{\mathbf{E}}$ is a three-way array containing the residual terms.

It is possible to fit a PARAFAC model to raw data such as a three-way array with modes: $objects \times variables_1 \times variables_2$ (direct fitting). Equation 3 demonstrates the direct fitting approach using an alternative formulation of a PARAFAC model.

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}^T + \mathbf{E}_k \tag{3}$$

where $\mathbf{X}_k$ represents the k^{th} frontal slice of a three-way array and $\mathbf{A}$ and $\mathbf{B}$ are the component matrices in the first and second mode, respectively. $\mathbf{D}_k$ is a diagonal matrix, whose diagonal elements are the k^{th} row of the third component matrix C. Finally, $\mathbf{E}_k$ contains the error terms corresponding to the entries in the k^{th} frontal slice.

While direct fitting is applied on the raw data, it is also possible to apply indirect fitting on covariance matrices of data slices. For indirect fitting, raw data are rearranged as a three-way dataset of covariance matrices, for instance in the form of $objects \times objects \times variables_1$ or $objects \times objects \times variables_2$ assuming one is particularly interested in the object mode. The relationship between direct and indirect fitting approach is similar to the one between SVD on a data matrix and eigenvalue decomposition on a covariance matrix.

2.1.2 Extensions of PARAFAC

Some of the extensions of PARAFAC model are PARAFAC2, Shifted PARAFAC (S-PARAFAC), Parallel Factors with Linear Dependency (PARALIND) and Convolutive PARAFAC (cPARAFAC). We discuss these models and their similarities briefly. Their mathematical formulations are also given in Table 1.

1. PARAFAC2 [Harshman 1972]: is introduced as a less restrictive model than PARAFAC. For instance, indirectly fitting PARAFAC has a restriction such that PARAFAC can model data indirectly if one of the component matrices in the model is columnwise orthogonal. Consequently, indirect fitting approach using a PARAFAC model cannot be applied to extract oblique factors in every mode of a dataset. On the other hand, PARAFAC2 can model data indirectly to extract either orthogonal or oblique factors. PARAFAC2 relaxes a PARAFAC model by requiring the invariance of the matrix multiplication of a component matrix with its transpose in one mode rather than the invariance of the components themselves.

$$\mathbf{X}_k = \mathbf{A}_k\mathbf{D}_k\mathbf{B}^T + \mathbf{E}_k \tag{4}$$

$$\text{s.t. } \mathbf{A}_k^T\mathbf{A}_k = \Phi \qquad \text{k} = 1,...\text{K}$$

where $\mathbf{A}_k$ is the component matrix in the first mode corresponding to the k^{th} frontal slice. Φ, which is the matrix product of $\mathbf{A}_k$ with its transpose, is required to be invariant for all slices $k = 1,..K$. In Equation 4, we observe that unlike in a PARAFAC model, component matrix in the first mode (or one of the modes) can vary across slices in a PARAFAC2 model. This relaxation enables the use of multiway models in the cases, where a PARAFAC model cannot fully recover the underlying structure, e.g. modeling chromatographic data with retention time shifts [Bro et al. 1999]. Furthermore, PARAFAC2 solves the problem of modeling three-way arrays with slices of different dimensionality (dimensionality differs only in one mode). An example of such a multiway array is an environmental dataset that contains the concentrations of some chemical compounds measured at certain time periods across several sampling sites ($sampling\ sites \times parameters \times time$) [Stanimirova et al. 2004]. It is quite common to have measurements from sampling sites for varying time periods, which would result in a three-way array with different dimensionality in one of the modes (e.g. time mode in this case). A PARAFAC2 model using an indirect fitting approach can also handle different dimensionality across slices. Nevertheless, directly fitting PARAFAC2 on raw data has more advantages than indirect fitting in terms of imposing constraints, handling missing data and generalization of the model to N-way arrays [Kiers et al. 1999].

2. S-PARAFAC [Harshman et al. 2003]: has been introduced in order to deal with shifting factors in sequential data such as time series or spectral data. While PARAFAC restricts the data to have the same factor in various proportions in all samples based on Cattell's idea, S- PARAFAC relaxes this restriction by incorporating shifting information into the model and capturing the factors even if they are available in shifted positions in different samples. One limitation of S-PARAFAC is, though, it only considers one-dimensional shifts such as time shifts but does not handle multidimensional shifts that might be encountered in image sequences like brain scans. When compared to PARAFAC2, S-PARAFAC is quite similar. First of all, both models are less-constrained versions of PA-

Table 1: Selection of models from PARAFAC family. PARAFAC2 and PARALIND are expressed in matrix notation to make it easier to understand and compare with PARAFAC representation in Equation 3. In PARALIND, $\mathbf{H}$ represents the dependency matrix. For Shifted PARAFAC, s_{jr} represents the shift at column j for the r^{th} factor. In Convolutive PARAFAC, θ is used to capture the shifts in the log-frequency spectrogram.

Model Name	Mathematical Formulation	Handles Rank-deficieny	Extended to Nway data
PARAFAC	$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}$	×	✓
PARAFAC2	$\mathbf{X}_k = \mathbf{A}_k \mathbf{D}_k \mathbf{B}^T + \mathbf{E}_k$	×	✓
S-PARAFAC	$x_{ijk} = \sum_{r=1}^{R} a_{(i+s_{jr})r} b_{jr} c_{kr} + e_{ijk}$	×	✓
PARALIND	$\mathbf{X}_k = \mathbf{AHD}_k \mathbf{B}^T + \mathbf{E}_k$	✓	✓
cPARAFAC	$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{(j-\theta)r} c_{kr}^{\theta} + e_{ijk}$	×	✓

RAFAC and can model multiway datasets, which do not follow Cattell's Parallel Proportional Profiles idea. Secondly, PARAFAC2, similar to S-PARAFAC, can model data with shifting factors. In fact, both models have been used in the analysis of chromatographic data with retention time shifts [Bro et al. 1999; Hong and Harshman 2003]. However, PARAFAC2 can only capture shifts that maintain the inner product of the factors (i.e. the constraint in Equation 4) while S-PARAFAC can model independent shifts at each factor. Still though S-PARAFAC can be considered as a restricted version of PARAFAC2 since S-PARAFAC can handle only shifting factors whereas PARAFAC2 can capture the structure in data as long as the inner product of the factors across slices are the same.

3. cPARAFAC [Mørup and Schmidt 2006]: Another extension on PARAFAC is cPARAFAC, which is a generalization of Non-negative Matrix Factor Deconvolution (NMFD) to multiway spectral data. cPARAFAC, closely related to S-PARAFAC, has been proposed for multi-channel spectral data analysis in order to model convolutive mixtures. Convolution basically means generating a mixture by sending the sources through a filter. When convolution filter is sparse, cPARAFAC becomes equivalent to S-PARAFAC.

4. PARALIND [Bro et al. 2005]: A common problem that arises in real data analysis is that ranks of the component matrices may not be the same (called *rank deficiency*). That would require extracting different number of factors in different modes. In that case, fitting a PARAFAC model would perfectly give rank deficient solutions and would not guarantee meaningful uniqueness. PARALIND is proposed as an approach for modeling such cases. This model introduces dependency (or interaction) matrices among component matrices to enable the modeling of the data with component matrices with different ranks and capture the dependency between components. Besides, via dependency matrices, prior knowledge about the data and constraints can be incorporated into the model.

In addition to the extensions of PARAFAC model discussed so far, there is also another recent model called Constrained-block PARAFAC [Almeida et al. 2006]. Constrained-block PARAFAC proposes to model a multiway array as a sum of PARAFAC blocks. Each block is modeled using a PARAFAC model but has certain constraints enforced by the constraint matrices introduced in the model, similar to dependency matrices of a PARALIND model.

2.2 Tucker-family

Structural models in the PARAFAC family can be considered as constrained versions of less restricted multiway models, i.e. Tucker models, which are also called N-way principal component analysis techniques.

2.2.1 Tucker3

Similar to PARAFAC, Tucker3 is an extension of bilinear factor analysis to higher-order datasets. Equation 5 shows a commonly-used formulation for a Tucker3 model applied on a three-way array $\underline{\mathbf{X}} \in R^{I \times J \times K}$.

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk} \tag{5}$$

where $\mathbf{A} \in R^{I \times P}$, $\mathbf{B} \in R^{J \times Q}$ and $\mathbf{C} \in R^{K \times R}$ are the component matrices corresponding to the first, second and third modes, respectively. $\underline{\mathbf{G}} \in R^{P \times Q \times R}$ is the core array and $\underline{\mathbf{E}} \in R^{I \times J \times K}$ contains the residuals. Illustration of a Tucker3 model on a three-way array is given in Figure 6. Compared to PARAFAC, a Tucker3 model is a more flexible model. This flexibility is due to the core array, $\underline{\mathbf{G}}$, which allows an interaction between a factor with any factor in the other modes. While the core array enables us to explore the underlying structure of a multiway dataset much better than a restricted PARAFAC model, full-core array structure in Tucker3 has some drawbacks. First, this property is the reason for rotational indeterminacy in Tucker3 models. Unlike PARAFAC, a Tucker3 model cannot determine component matrices uniquely. When a component matrix is rotated by a rotation matrix, it is possible to apply the inverse of the rotation matrix to the core and still obtain the same model fit. Therefore, a Tucker3 model can determine component matrices only up to a rotation. Second, interpretation of Tucker3 models is much more difficult compared to PARAFAC models.

Originally, Tucker family contains Tucker1, Tucker2 and Tucker3 models (Table 2). Tucker1 is based on the simple idea of rearranging data as a matrix and decomposing the unfolded data using SVD. Tucker2 and Tucker3 models allow rank reduction in more than one mode and named after the number of modes rank reduction is allowed. Desired rank reduction in each mode are user-specified model parameters and determining these parameters in Tucker models is a tedious task. While using ranks indicated by SVD on unfolded data in each mode is a practical option, systematic methods, e.g. cross validation, DIFFIT [Timmerman and Kiers 2000], have also been developed. DIFFIT (Difference in Fit)

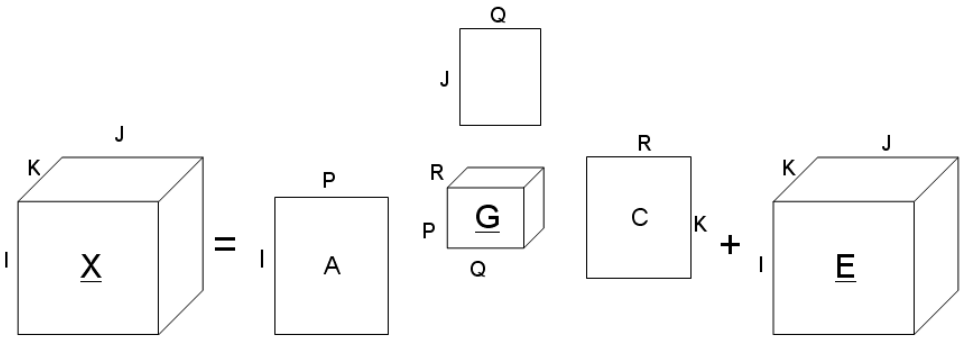

Figure 6: Illustration of a Tucker3 model. (P, Q, R)-component Tucker3 model, where a three-way array $\underline{\mathbf{X}} \in R^{I \times J \times K}$ is modeled with component matrices $\mathbf{A} \in R^{I \times P}$, $\mathbf{B} \in R^{J \times Q}$ and $\mathbf{C} \in R^{K \times R}$ in the first, second and third mode, respectively. $\underline{\mathbf{G}} \in R^{P \times Q \times R}$ is the core array and $\underline{\mathbf{E}} \in R^{I \times J \times K}$ contains the error terms.

enumerates all possible models and uses the differences between model fits to determine the number of components. However, high computational complexity of DIFFIT makes it inefficient. Therefore, it has later been improved by comparing approximate model fit values rather than exact model fits [Kiers and Kinderen 2003]. The most recent work in finding the number of components is based on searching for the convex hull on the plot of model fit values vs. number of free parameters [Ceulemans and Kiers 2006]. This approach is more general than previously-proposed methods and helps in determining the model parameters in not only Tucker3 but also Tucker1, Tucker2 and PARAFAC models. Even though empirical comparison of DIFFIT and the convex hull approach on simulation data suggests that the convex hull approach gives promising results and outperforms previous methods, there is no straightforward way to find the optimal number of components [Ceulemans and Kiers 2006]. Similar to the case of determining component numbers in a PARAFAC model, several diagnostics should be used to have a true understanding of the structure of a multiway dataset.

2.2.2 *Extensions of Tucker3*

Similar extensions as in PARAFAC models have also been studied for Tucker models in order to capture shifting factors. Shifted Tucker3 (S-T3) and Shifted Tucker2 (S-T2), combination of Shifted Factor Analysis with Tucker3 and Tucker2 models, have been introduced by [Harshman et al. 2003]. Although it is not proven formally, it has been discussed that incorporating shifting information in S-T3 suggests the uniqueness of a S-T3 model [Harshman et al. 2003]. Structural models in Tucker-family are given in Table 2.

2.3 Tucker3 vs. SVD vs. PARAFAC

Tucker3 can be considered as the generalization of SVD to higher-order tensors. The link between Tucker3 and SVD and how singular values and singular vectors generalize to those of higher-order datasets have been extensively studied in [Lathauwer et al. 2000]. This is a significant milestone in multiway literature since it links multilinear algebra with models originated in psychometrics and chemometrics. Later, computation of singular values and singular vectors of tensors using Lagrangian approach have been discussed in depth in [Lim 2005], which complements the theoretical background of generalization of singular value decomposition to higher-order datasets.

In [Lathauwer et al. 2000], Tucker3 with orthogonality constraints on the components, has been named as Higher-Order Singular Value Decomposition (HOSVD). HOSVD can simply be computed by flattening the tensor in each mode and calculating the singular vectors corresponding to that mode, which are also called $n-mode$ singular vectors. Given the singular vectors, a core tensor can be computed as shown in Step 5 of Tucker3-ALS algorithm in Figure 7. However, unlike SVD, HOSVD does not provide the best rank-$(R_1, R_2, ..R_N)$ approximation of a tensor [Lathauwer et al. 2000], where R_i is the rank of tensor in i^{th} mode. The rank of a tensor in n^{th} mode is called $n-rank$ and it is the dimension of the vector space spanned by the columns of the matrix obtained by flattening the tensor in n^{th} mode. Nevertheless, it does give a good approximation of the data as shown in many applications, e.g. face recognition on an image dataset, where images are affected by several factors such as viewpoints, facial expressions, lighting conditions, etc. [Vasilescu and Terzopoulos 2002].

Compared to SVD, which is a decomposition that represents a matrix as a sum of rank-1 matrices, HOSVD does not decompose a tensor as a sum of rank-1 tensors. In that sense, PARAFAC is considered to be another generalization of SVD to higher-order arrays because PARAFAC decomposes a tensor as the sum of rank-1 tensors. However, orthogonality constraints on the component matrices of a PARAFAC model, in general, cannot be satisfied. In order to be able to decompose a tensor with a PARAFAC model, which will give component matrices with orthogonal columns, a tensor should be diagonalizable and in general they are not [Kolda 2001].

SVD has been quite popular in every field of data analysis from signal processing to social network analysis, from chemometrics to image compression because it enables noise filtering through dimensionality reduction. The first R significant singular vectors may represent the data very well and SVD provides the best rank-R approximation for a matrix. Besides, if the best rank $R+1$ approximation is sought, then first R singular vectors are kept the same and only one more singular vector is computed. This property has played an important role in the development of online SVD algorithms, which compute SVD of a data stream by updating singular vectors rather than computing SVD of the whole

Table 2: Selection of models from Tucker family. In Shifted Tucker3, s_{jp} indicates the shift at j^{th} column for p^{th} factor. Shifted Tucker2 is formulated similarly.

Model Name	Mathematical Formulation	Handles Rank-deficieny	Extended to Nway data
Tucker1	$x_{ijk} = \sum_{p=1}^{P} g_{pjk} a_{ip} + e_{ijk}$	✓	✓
Tucker2	$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} g_{pqk} a_{ip} b_{jq} + e_{ijk}$	✓	✓
Tucker3	$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}$	✓	✓
S-Tucker3	$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{(i+s_{jp})p} b_{jq} c_{kr} + e_{ijk}$	✓	✓

dataset every time data are updated [Levy and Lindenbaum 2002]. However, the best rank-R approximation through SVD in matrices cannot be generalized to tensors [Kolda 2001; Kolda 2003]. Through a counterexample, [Kolda 2003] demonstrates that best rank-$(R+1)$ approximation of a tensor does not necessarily contain the components present in best rank-R approximation of the tensor. By best rank-R approximation of a tensor, we refer to orthogonal rank decomposition of a tensor, where tensor $\underline{\mathbf{X}}$ is expressed as the weighted sum of rank-one tensors as in Equation 6.

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \sigma_r \underline{\mathbf{U}}_r \tag{6}$$

where $\underline{\mathbf{U}}_i \perp \underline{\mathbf{U}}_j$ for all $i \neq j$ and the minimal number of rank-one tensors (minimal R) needed to express $\underline{\mathbf{X}}$ in this form is called the *orthogonal rank* of $\underline{\mathbf{X}}$. For a detailed discussion on rank decompositions and algorithms computing the best rank approximations, the reader is referred to [Lathauwer et al. 2000; Kolda 2001; Zhang and Golub 2001].

2.4 Alternative Models

There exist several other models based on approaches other than PARAFAC and Tucker models for unsupervised multiway data analysis. We briefly introduce Multilinear Engine (ME) [Paatero 1999], multiway models based on STATIS [Stanimirova et al. 2004] and multiblock multiway models [Smilde et al. 2000].

ME is a program that is capable of fitting different structural models including PARAFAC and PARAFAC2 on multiway arrays using a general-purpose optimization/curve fitting approach. Although models mentioned so far are only capable of modeling multilinearity in data, structure tables created by specified variables and functions enable ME to fit multilinear as well as quasi-multilinear models. Multilinear models are based on mathematical expressions, which are linear with respect to each set of variables corresponding to different modes whereas quasi-multilinear models contain nonlinearity in the sense of polynomials. Therefore, the multilinear engine can explore a wider range of structures in data compared to PARAFAC, Tucker3, etc.

Another model focusing on three-way data analysis is STATIS [Stanimirova et al. 2004] originally studied in [Carlier et al. 1989; Lavit et al. 1994]. When compared to N-way analysis methods, which explore each mode simultaneously, this STATIS-based model explores each mode separately. It considers each observation/sample as a slice of a three-way array and computes the covariance matrix corresponding to that slice. The basic principle in the model is to apply Principal Component Analysis (PCA) on a global covariance matrix formed as a linear combination of covariance matrices corresponding to individual slices. Similar to indirect fitting approach, it is possible to analyze three-way arrays with slices of different sizes using STATIS. One disadvantage of STATIS is that it cannot be generalized to N-way arrays.

Methods, referenced so far, focus on the analysis of a single multiway array. On the other hand, multiblock multiway arrays are also encountered in various studies such as control of batch processes, where more than two blocks of multiway arrays need to be analyzed simultaneously. One approach to deal with multiblock multiway component problems is to analyze each multiway array using a certain structural model such as a Tucker3 or a PARAFAC model and then combine summaries of information from different multiway arrays in a single matrix [Smilde et al. 2000]. The matrix containing summaries from different arrays can then be analyzed using bilinear factor models. This approach can be considered as a generalized version of Collective PCA [Kargupta et al. 2001] to higher-order datasets.

3. ALGORITHMS

Algorithms for fitting multiway models are, in general, iterative algorithms and based on Alternating Least Squares (ALS). We briefly discuss the algorithms used for fitting the structural models studied in the previous section. Furthermore, we include robust algorithms developed to handle outliers and noise problems in multiway data analysis.

3.1 Alternating Least Squares

In ALS, component matrices are usually estimated one at a time keeping the estimates for other component matrices fixed. Estimation of component matrices is repeated until a convergence criterion, e.g. no change in model fit, is satisfied. Apart from its simplicity, ALS is a frequently used algorithm since it can be combined with constraints, generalized to higher-order arrays and modified to handle objective functions in the form of weighted least squares. Here we briefly describe a Tucker3-ALS algorithm with orthogonality constraints on the component matrices and a PARAFAC-ALS algorithm on a three-way array $\underline{\mathbf{X}} \in R^{I \times J \times K}$.

1. Tucker3-ALS: A Tucker3 model on $\underline{\mathbf{X}}$ can be expressed as $\mathbf{X}_{(1)} = \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)}$, where $\mathbf{X}_{(1)}$ and $\mathbf{E}_{(1)}$ are unfolded tensors in the first mode of size $I \times JK$, $\mathbf{G}_{(1)}$ is the core array matricized in the first mode, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are the component matrices as given in Figure 6. The symbol $\otimes$ denotes Kronecker product. Kronecker product of matrices is defined as follows. Let $\mathbf{M} \in R^{I \times J}$ and $\mathbf{N} \in R^{K \times L}$. Kronecker product denoted by $\mathbf{M} \otimes \mathbf{N}$ is then given by:

Tucker3-ALS

Input: $\underline{\mathbf{X}}$ and the number of components to be extracted in each mode (P, Q, R)

Output: $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\underline{\mathbf{G}}$

(Step 1) Initialize $\mathbf{B}$ and $\mathbf{C}$, where $\mathbf{B}$ is of size J x Q and $\mathbf{C}$ is of size K x R

Do while convergence criterion not satisfied

(Step 2) Decompose $\mathbf{X}_{(1)}(\mathbf{C}\otimes\mathbf{B})$ using SVD

$$\mathbf{X}_{(1)}(\mathbf{C}\otimes\mathbf{B})=\mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^{\mathbf{T}}$$

$$\mathbf{A}=\mathbf{U}_1(:,\ 1{:}\mathrm{P})$$

where columns of $\mathbf{A}$ are the first P left singular vectors of $\mathbf{X}_{(1)}(\mathbf{C}\otimes\mathbf{B})$.

(Step 3) Decompose $\mathbf{X}_{(2)}(\mathbf{C}\otimes\mathbf{A})$ using SVD

$$\mathbf{X}_{(2)}(\mathbf{C}\otimes\mathbf{A})=\mathbf{U}_2\mathbf{S}_2\mathbf{V}_2^{\mathbf{T}}$$

$$\mathbf{B}=\mathbf{U}_2(:,\ 1{:}\mathrm{Q})$$

where columns of $\mathbf{B}$ are the first Q left singular vectors of $\mathbf{X}_{(2)}(\mathbf{C}\otimes\mathbf{A})$.

(Step 4) Decompose $\mathbf{X}_{(3)}(\mathbf{B}\otimes\mathbf{A})$ using SVD

$$\mathbf{X}_{(3)}(\mathbf{B}\otimes\mathbf{A})=\mathbf{U}_3\mathbf{S}_3\mathbf{V}_3^{\mathbf{T}}$$

$$\mathbf{C}=\mathbf{U}_3(:,\ 1{:}\mathrm{R})$$

where columns of $\mathbf{C}$ are the first R left singular vectors of $\mathbf{X}_{(3)}(\mathbf{B}\otimes\mathbf{A})$.

Loop

(Step 5) $\mathbf{G}_{(1)}=\mathbf{A}^{\mathbf{T}}\mathbf{X}_{(1)}(\mathbf{C}\otimes\mathbf{B})$

Figure 7: Tucker3-ALS algorithm with orthogonality constraints on the component matrices. ALS algorithm for computing a (P, Q, R)-component Tucker3 model on a three-way array $\underline{\mathbf{X}} \in R^{I\times J\times K}$. Commonly used convergence criterion is epsilon difference between the values of relative change in fit computed in consecutive iterations.

$$M \otimes N = \begin{bmatrix} m_{11}N & m_{12}N & ... & m_{1J}N \\ m_{21}N & m_{22}N & ... & m_{2J}N \\ ... & ... & ... & ... \\ m_{I1}N & m_{I2}N & ... & m_{IJ}N \end{bmatrix} \quad (7)$$

The objective function in a Tucker3 model is to minimize the error given as follows:

$$min_{A,B,C,G} \left|\left|\mathbf{X}_{(1)} - \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C}\otimes\mathbf{B})^T\right|\right|^2$$

Tucker3-ALS algorithm with orthogonality constraints on the component matrices is given in Figure 7.

2. PARAFAC-ALS: Similarly, a PARAFAC model given in Equations 1, 2 and 3 can alternatively be expressed in matrix notation as $\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{C}\odot\mathbf{B})^T + \mathbf{E}_{(1)}$, where $\odot$ denotes Khatri-Rao product. Khatri-Rao product is defined as follows. Let $\mathbf{M} \in R^{I\times K}$ and $\mathbf{N} \in R^{J\times K}$ and m_i and n_i represent the i^{th} column of $\mathbf{M}$ and $\mathbf{N}$, respectively for $i = 1, 2, ...K$. Then Khatri-Rao product denoted by $\mathbf{M} \odot \mathbf{N}$ is given as $\mathbf{M} \odot \mathbf{N} = [m_1 \otimes n_1\ m_2 \otimes n_2\ ...\ m_K \otimes n_K]$, which can be defined, in other words, as columnwise kronecker product. The objective function that minimizes the residuals is then given by:

$$min_{A,B,C} \left|\left|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C}\odot\mathbf{B})^T\right|\right|^2$$

Figure 8 shows an algorithm satisfying the objective function based on ALS.

In addition to ALS, there exist alternative algorithms for PARAFAC. These algorithms can be studied under three categories as specified in [Tomasi and Bro 2006]: alternating algorithms, closed form solutions and gradient-based methods. Representatives of alternating algorithms, i.e. Alternating Slice-wise Diagonalization (ASD) [Jiang et al. 2000], Self Weighted Alternating Trilinear Diagonalization (SWATLD) [Chen et al. 2000], have recently been studied and compared in depth in [Faber et al. 2003; Tomasi and Bro 2006]. These algorithms particularly aim to improve PARAFAC -ALS in terms of convergence rate and robustness to overfactoring, i.e. more factors than the model rank of an array are extracted. However, their objective functions are not based on least squares and are, in fact, not well-defined. The second category contains Generalized Rank Annihilation Method (GRAM) and Direct Trilinear Decomposition (DTLD), which find the solution of a PARAFAC model by solving a generalized eigenvalue problem. GRAM handles the cases where a three-way dataset has only two slices in one mode and DTLD is the generalization of GRAM to cases where more than two matrices/slices are explored.

PARAFAC-ALS

Input: $\underline{\mathbf{X}}$ and the number of components to be extracted in each mode R

Output: $\mathbf{A}, \mathbf{B}, \mathbf{C}$

(Step 1) Initialize $\mathbf{B}$ and $\mathbf{C}$, where $\mathbf{B}$ is of size J x R and $\mathbf{C}$ is of size K x R

Do while convergence criterion not satisfied

(Step 2) $Z=C \odot B$

$A=X_{(1)}Z(Z^TZ)^{-1}$

(Step 3) $Z=C \odot A$

$B=X_{(2)}Z(Z^TZ)^{-1}$

(Step 4) $Z=B \odot A$

$C=X_{(3)}Z(Z^TZ)^{-1}$

Loop

Figure 8: PARAFAC-ALS algorithm. ALS algorithm for computing an R-component PARAFAC model on a three-way array $\underline{\mathbf{X}} \in R^{I \times J \times K}$. Similar to Tucker3-ALS, the algorithm converges when the relative change in fit does not change much from one iteration to another.

Finally, the last category contains the algorithms based on Gauss-Newton approach. PMF3, the algorithmic method also employed in ME, is based on a modification of a Gauss-Newton method. For further information and comparison of these algorithms, the reader is referred to [Faber et al. 2003; Tomasi and Bro 2006].

Algorithms employed in multiway models other than PARAFAC also rely on the algorithms discussed in the previous paragraph. Both indirect [Harshman 1972] and direct fitting [Kiers et al. 1999] approaches proposed for PARAFAC2 are based on ALS. Similarly, an ALS-based approach improved by weighting schemes is used in Shifted Factor models. Besides, most popular algorithms used in fitting a Tucker model are Tucker3-ALS given above and closed form solutions based on SVD.

3.2 Robust Algorithms

In addition to the algorithms mentioned so far, there are also algorithms that particularly focus on robustness to outliers and different types of noise/measurement errors. In the area of outlier detection, robust algorithms for both Tucker3 (Robust Tucker3) [Pravdova et al. 2001] and PARAFAC (Robust PARAFAC) [Engelen and Hubert 2005] have been developed. The motivation in these robust algorithms is to improve ALS method, which can dramatically suffer from outliers. In these algorithms, PARAFAC and Tucker3 are combined with methods that can automatically handle outlier detection. Robust PARAFAC combines Robust PCA with a PARAFAC model while Robust Tucker3 applies multivariate trimming or minimum covariance determinant methods to remove outliers before applying a Tucker model. When we consider the types of noise or measurement errors, in general noise is either omitted in the studies or simply assumed to be additive noise. Besides, noise is often assumed to be identically and independently distributed Gaussian. However, if this assumption does not hold and measurement errors follow a Laplacian or Cauchy distribution instead, ALS does not perform as well as some recently proposed algorithms [Vorobyov et al. 2005], which rely on an objective function, that is least absolute error based on L1-norm rather than least squares. It has been demonstrated in the context of multiuser separation detection for a direct-sequence code-division multiple access that when measurement errors do not come from a Gaussian distribution, ALS fails to capture the underlying structure as well as the recently developed algorithms [Vorobyov et al. 2005]. Even if algorithms are improved for different distributions, the assumption of noise being identically and independently distributed remains as a challenge. With a goal of handling the cases where noise is not identically distributed, a weighted least squares approach has been proposed by [Kiers 1997]. This approach has later been generalized to model data, which might have correlated measurement errors, in [Bro et al. 2002] by an algorithm called Maximum likelihood via Iterative Least Squares (MILES). Similarly, in another study, a generalization of maximum likelihood PCA, called maximum likelihood PARAFAC (MLPARAFAC), has been proposed in order to fit a PARAFAC model to multiway arrays containing correlated noise structure in general or in a specific mode [Vega-Montoto and Wentzell 2003]. While MILES has been developed as a general algorithm to fit any model with a least squares solution, MLPARAFAC has been particularly developed to fit a PARAFAC model.

Even though it is often not mentioned separately as a part of many of the algorithms, preprocessing is also a crucial step in data analysis. Similar to preprocessing in two-way data analysis, centering and scaling both generalize to multiway arrays [Bro and Smilde 2003]. Centering across one mode of a three-way dataset is performed by simply matricizing the data in the desired mode and applying two-way centering. Scaling, on the other hand, requires scaling each slice of the array corresponding to each variable rather than scaling

columns as in two-way analysis.

4. APPLICATIONS

As already mentioned, multiway models are employed in numerous disciplines addressing the problem of finding hidden multilinear structure in multiway datasets. There are many applications in various fields and this survey offers some representative examples from different research areas.

4.1 Chemistry

We start with one of the most popular application of a PARAFAC model: modeling fluorescence excitation-emission data, which is a commonly used data type in chemistry, medicine and food science. Such data typically consist of samples containing different concentrations of several chemical compounds. The goal of PARAFAC analysis on this data type is to determine the compounds found in each sample as well as the relative concentrations of compounds. Fluorescence spectroscopy enables the generation of three-way datasets with modes: *samples* $\times$ *emission* $\times$ *excitation*. Among many other applications of PARAFAC, modeling of fluorescence spectroscopy is the one, which demonstrates the modeling power and interpretation of factors of a PARAFAC model best. An example of a PARAFAC model on a fluorescence spectroscopic dataset is given as an in-depth study on a Fish dataset and data with known fluorophores in [Andersen and Bro 2003]. This study is an important resource demonstrating the underlying idea of the structural model of PARAFAC, its benefits and limitations. Limitations of a PARAFAC model have also been addressed in another study modeling chromatographic data [Bro et al. 1999]. Even though initial challenge of uniquely identifying the components in chromatographic data can be solved by a PARAFAC model, PARAFAC falls short in extracting the components when elution profiles of the components follow a shifting pattern throughout the experiments. On the other hand, PARAFAC2 succeeds in modeling the shifting factors and recovering the underlying components in chromatographic data with retention time shifts. In addition to these studies, a recent review on multiway analysis in chemistry lists a broader range of applications on chemical datasets [Bro 2006].

4.2 Neuroscience

In neuroscience, multiway models have been initially used in studying the effect of a new drug on brain activity [Estienne et al. 2001]. In this study, EEG data and data collected through experiments with different doses of a drug over several patients under certain conditions are arranged as a six-way array with modes: EEG, patients, doses, conditions, etc. Results demonstrate that significant information is successfully extracted from a complex drug dataset by a Tucker3 model rather than two-way models such as PCA. Multiway models have become even more popular in neuroscience with the idea of decomposing EEG data into space-time-frequency components [Miwakeichi et al. 2004]. PCA and Independent Component Analysis (ICA) have been frequently used in analyzing EEG datasets that are in the form of *time samples* $\times$ *electrodes* or *frequency* $\times$ *electrodes*. However, these techniques have not taken into account the frequency content of the signals in specific time periods across different electrodes. In [Miwakeichi et al. 2004], wavelet transformation is applied on the signals recorded at each electrode. Then wavelet-transformed data are arranged as a three-way array with modes *time samples* $\times$ *frequency* $\times$ *electrodes* and analyzed using a PARAFAC model. Factors in the first, second and third component matrices represent the temporal, spectral and spatial signatures of the EEG data, respectively. PARAFAC models with nonnegativity constraints have later been used in another study on event-related potentials (ERP) to find the underlying structure of brain dynamics [Mørup et al. 2006]. These studies have also motivated the application of multiway models in understanding the structure of epileptic seizures, localizing focus origins [Acar et al. 2006] and removing artifacts using multilinear subspace analysis [Acar et al. 2007]. Most recently, a toolbox called ERPWAVELAB [Mørup et al. 2006] running under MATLAB [MathWorks] has been released for multichannel time-frequency analysis of brain activity using EEG and MEG data. This toolbox enables the use of multiway models in the analysis of brain dynamics.

4.3 Social Network Analysis/Web-mining

Multiway data analysis has also often been employed in extracting relationships in social networks. The aim of social networks analysis is to study and discover hidden structures in social networks, for instance, extracting communication patterns among people or within organizations. In [Acar et al. 2005], chatroom communications data have been arranged as a three-way array with modes: *users* $\times$ *keywords* $\times$ *time samples* and the performance of multiway models in capturing the underlying user group structure has been compared with that of two-way models. Another recent study [Acar et al. 2006] assesses the performance of collective and centralized tensor analysis approaches again in the context of chatroom communications. Not only chatroom but also email communications have been analyzed using multiway models [Bader et al. 2006]. In the context of web link analysis, [Kolda et al. 2005] and [Kolda and Bader 2006] combine hyperlink and anchor text information and rearrange web graph data as a sparse three-way tensor with modes: *webpages* $\times$ *webpages* $\times$ *anchor text*. Web graph is then analyzed using an algorithm improved to fit a PARAFAC model to large and sparse datasets efficiently in order to capture the groupings of webpages and identify the main topics. Finally, with a goal of improving personalized web searches, click-through data have been analyzed using an algorithm called CubeSVD [Sun et al. 2005], which is indeed the same as HOSVD. Click-through data are arranged as a three-way array with modes: *users* $\times$ *queries* $\times$ *webpages* and CubeSVD is compared with two-way methods such as Latent Semantic Indexing (LSI) and shown to outperform the two-way approaches.

4.4 Computer Vision

Approximations of tensors have proved to be useful in computer vision applications such as image compression and face recognition. In computer vision and graphics, data often have multiway nature, e.g a color image is a three-way array with x and y-coordinates being two of the modes and color being the third mode. Previously-developed image coding techniques consider images as vectors or matrices by reshaping the data as a vector or a matrix. For instance, x and y-coordinates are vectorized and a color image is often represented as a two-way array with modes: color and spatial information. On the other hand, recently it has been

shown that when images are represented as tensors and iteratively obtained rank-one approximations of tensors are used to compress these images [Wang and Ahuja 2004], the error between the original and reconstructed images is less than the error obtained when PCA is used for compression. Another application of multiway models in computer vision is face recognition, where a set of face images are arranged as a tensor, that represents images using not only pixel information but also illuminations, expressions, viewpoints and person identities. HOSVD is then used to decompose this tensor and basis vectors called TensorFaces [Vasilescu and Terzopoulos 2002] are determined. Component matrices extracted from each mode is used to construct person-specific, viewpoint-specific, etc. TensorFaces, which improve the understanding of the underlying factors in an image dataset. Apart from these applications, tensors have also been employed in several other fields, e.g. textured surface rendering, identifying human-motion signatures, in computer vision community.

4.5 Process Analysis

Last research area we will mention in this survey is process monitoring. Real-time batch process monitoring is a challenging task since the whole data are needed for the analysis in general and that would require waiting till the completion of a batch. However, in-filling of missing future data and modeling using a PARAFAC model have shown to overcome this challenge [Meng et al. 2003] and PARAFAC has been demonstrated to be an applicable approach in controlling batch processes arranged as a three-way array with modes: *batches* $\times$ *variables* $\times$ *time samples*. Similarly, STATIS has been applied in monitoring batch processes on datasets that come from various areas such as pharmacology, spectroscopy and yeast bakery production. These datasets are arranged as three-way arrays but they have different number of dimensions in the time mode [Gourvnec et al. 2005] so STATIS can handle this problem easily. On these datasets, STATIS approach has been compared with unfolded PCA on variables mode, which forms a matrix of *variables* $\times$ *batches - time samples*. It has been observed that detecting bad batches is much easier with STATIS than it is with unfolded PCA. Therefore, this study is a good example for the cases when unfolded PCA will suffer from loss of information while three-way analysis techniques will perform better.

5. SOFTWARE

As multiway analysis is spreading from chemometrics and psychometrics to other fields, software tools have also been developed and improved. Currently available softwares for multiway data analysis are the Nway Toolbox [Andersson and Bro 2000], Tensor Toolbox [Bader and Kolda 2006a; Bader and Kolda 2007], PLS Toolbox [Eigenvector 2006] and CuBatch [Gourvnec et al. 2005], which all run under MATLAB. The Nway toolbox is the original toolbox, which has combined multiway analysis techniques such as PARAFAC, Tucker models in a software package and enabled the application of these models in different fields. The Tensor Toolbox has been initially introduced as a TensorClass, which handles mathematical operations on multiway arrays such as tensor-matrix multiplications, matricization of tensors and many other algebraic operations. It has later been improved to manipulate efficiently not only multiway arrays but also sparse tensors, where only small fraction of the elements are nonzero. CuBatch is another software package recently introduced as a multiway analysis toolbox with a user-friendly interface. It has been originally built for analyzing batch process data but it is also applicable on multiway datasets in general. Available models for unsupervised learning in this toolbox are PCA, PARAFAC, PARAFAC2, and Tucker models. Preprocessing techniques such as centering and scaling and different techniques for identifying outliers are also included in this toolbox. CuBatch contains Nway Toolbox functions and it is a more developed version of the initial toolbox. Apart from these freely available toolboxes, there also exists a commercial toolbox called PLS Toolbox, which enables the analysis of multiway arrays with numerous multiway models providing visual analysis tools. An efficient approach for analyzing multiway arrays would be to combine the Tensor Toolbox with one of the other toolboxes to have modeling, algorithmic and visualization power as a readily-available package.

In addition to software running under MATLAB, there is another software package called Multilinear Engine [Paatero 1999] implemented using FORTRAN. Structural models and algorithms used in Multilinear Engine have already been discussed in Section 2 and 3. There are also other software packages for manipulating multiway arrays but they do not particularly focus on multiway data analysis or multiway models. For more information on these software packages, interested users are referred to [Bader and Kolda 2006b] and references therein.

6. SUMMARY

Multiway data analysis has recently attracted attention in many fields. That is mostly due to the nature of the datasets, which cannot be truly captured by ordinary two-way analysis techniques. As datasets started to be rearranged as multiway arrays rather than matrices, multilinear models originated in late sixties and seventies have become popular. These models have been followed by extended versions of original models and techniques, e.g. Shifted factor models (S-PARAFAC, S-T3), PARALIND, in an effort to capture data-specific structures in multiway datasets. Theoretical aspects of these models such as model uniqueness and rank properties of multiway arrays have been studied more in depth. New algorithms, e.g. ASD, SWATLD, as alternatives to ALS have been developed. Multilinear algebra, a less-known field compared to linear algebra, has been explored to perform operations between multiway arrays and develop computationally efficient algorithms. Enhanced software tools, e.g. Nway Toolbox, CuBatch, enabling multiway data analysis and mathematical operations on multiway arrays, e.g. Tensor Toolbox, have been implemented.

This study mainly focuses on structural multiway models and briefly describes the algorithms employed in fitting these models to multiway datasets. We also give representative applications of multiway analysis from a variety of disciplines to illustrate the diversity of the fields making use of multiway data analysis. However, we have not mentioned some important aspects of multiway analysis in order to have a compact and rather self-contained survey. There are concepts such as uniqueness properties of models and uniqueness conditions for models, e.g. well-known Kruskal's

condition for PARAFAC [Kruskal 1989], understanding rotations in factor models and using those rotations to simplify the interpretation of a model and handling missing data. Furthermore, most of the structural models given in Section 2 can be employed for both unsupervised and supervised learning. In addition to those, there are also multiway models especially developed for supervised learning, e.g. Multilinear Partial Least Squares [Bro 1996]. We refer interested users to [Bro 1998] and to the most recent book in multiway data analysis [Smilde et al. 2004].

7. DISCUSSIONS

Recent studies show that multiway models have many application areas in computer science such as social network analysis, web link analysis and a variety of other problems in data mining as well as computer vision. Besides, recent theoretical studies focusing on multiway models improve the understanding of the models originally developed in chemistry and psychometrics and make them more popular in handling computer science problems.

Even though current algorithms and models are applicable on numerous datasets, there are still further progress needed in several fields. First area we want to emphasize is the summarization and analysis of data streams. Techniques discussed so far are based on offline and centralized dimensionality reduction models. On the other hand, there is also a demand for online methods to analyze data streams efficiently, especially in communication networks. For instance, a recent study introduces online algorithms for tensor analysis using an approach called dynamic tensor analysis [Sun et al. 2006]. Similarly, developing distributed versions of these methods would enable efficient analysis of massive datasets.

In addition to these, concepts of multilinearity and nonlinearity should be studied further so that limitations of multilinear models on capturing the structures in multiway datasets are better understood. Factors extracted by common multilinear models, i.e. PARAFAC-family and Tucker-family, are linear combinations of variables in each mode. On the other hand, these models will fail to discover nonlinear relationships between variables. Nonlinearity has been a recent topic of interest in data mining community especially since kernel methods became popular [Shawe-Taylor and Cristianini 2004]. Most two-way analysis techniques, e.g. PCA, Canonical Correlation Analysis (CCA), are combined with kernels in order to capture the underlying nonlinear structure among variables. Similar to a recent study in computer vision community, which has combined HOSVD with kernel methods for face recognition to capture the nonlinear relationship between image features [Y. Li and Lin 2005], embedding kernels into multiway models should be explored more, especially from a theoretical perspective.

Finally, performance of multiway data analysis in terms of space and computational complexity should be analyzed further. Most studies applying multiway data analysis demonstrate how multiway methods improve the interpretation and accuracy of the underlying structure captured when multiway models are used instead of two-way methods. However, computational and space complexity of multiway models and algorithms have not often been discussed in literature except for few studies including [Andersson and Bro 1998a; Andersson and Bro 1998b], where the speed of algorithms for a Tucker3 model is compared and Tucker3 is suggested as a compression method for speeding up multiway analysis, [Zhang and Golub 2001], which compares different algorithms for computing rank-1 approximation of a tensor, and [Bader and Kolda 2006b], which discusses efficient tensor storage schemes. Recently, for efficient analysis of massive multiway datasets, e.g. recommendation systems, hyperspectral image datasets, Tensor-CUR decomposition has been proposed [Mahoney et al. 2006]. Unlike multiway models discussed in this survey, Tensor-CUR does not employ an approach based on factor models. Instead, this algorithm relies on sampling subtensors, which consist of the original data elements, based on a given probability distribution and approximating the data using the sampled subtensors. While sampling reduces the complexity of the problem, how well it captures the multilinear structure in datasets in general is another open problem.

8. ACKNOWLEDGEMENTS

We would like to acknowledge the members of multiway community especially Henk Kiers, Lieven De Lathauwer, Morten Mørup and Alex Vasilescu for early discussions and references on topics included in this survey. In particular, we would like to thank Rasmus Bro and Tamara G. Kolda for their contributions to this survey through their comments and invaluable guidance in multiway data analysis.

9. REFERENCES

Acar, E., Bingol, C. A., Bingol, H., Bro, R., and Yener, B. 2007. Multiway analysis of epilepsy tensors. *Technical Report 07-04, Rensselaer Polytechnic Institute*.

Acar, E., Bingol, C. A., Bingol, H., and Yener, B. 2006. Computational analysis of epileptic focus localization. In *Proc. of The Fourth IASTED International Conference on Biomedical Engineering*. 317–322.

Acar, E., Camtepe, S. A., Krishnamoorthy, M., and Yener, B. 2005. Modeling and multiway analysis of chatroom tensors. In *Proc. of IEEE International Conference on Intelligence and Security Informatics*. Springer, Germany, 256–268.

Acar, E., Camtepe, S. A., and Yener, B. 2006. Collective sampling and analysis of high order tensors for chatroom communications. In *Proc. of IEEE International Conference on Intelligence and Security Informatics*. Springer, Germany, 213–224.

Almeida, A. L. F. D., Favier, G., and Mota, J. C. M. 2006. Block-parafac decompositions and applications in wireless communications. *Submitted to SIAM J. Matrix Anal. Appl.*.

Andersen, C. M. and Bro, R. 2003. Practical aspects of parafac modelling of fluorescence excitation-emission data. *J. of Chemometrics 17,* 4, 200–215.

Andersson, C. A. and Bro, R. 1998a. Improving the speed of multi-way algorithms: Part i. tucker3. *Chemometrics Intell. Lab. Systems 42,* 1-2, 93–103.

Andersson, C. A. and Bro, R. 1998b. Improving the speed of multi-way algorithms: Part ii. compression. *Chemometrics Intell. Lab. Systems 42,* 1-2, 105–113.

Andersson, C. A. and Bro, R. 2000. The n-way toolbox for matlab. *Chemometrics Intell. Lab. Systems 52,* 1, 1–4.

Bader, B. W., Harshman, R. A., and Kolda, T. G. 2006. Temporal analysis of social networks using three-way dedicom. *Technical Report SAND2006-2161, Sandia National Laboratories*.

Bader, B. W. and Kolda, T. G. 2006a. Algorithm 862: Matlab tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software 32,* 4, 635–653.

Bader, B. W. and Kolda, T. G. 2006b. Efficient matlab computations with sparse and factored tensors. *Technical Report SAND2006-7592, Sandia National Laboratories*.

Bader, B. W. and Kolda, T. G. 2007. Matlab tensor toolbox version 2.2. *http://csmr.ca.sandia.gov/ tgkolda/TensorToolbox*.

Bro, R. 1996. Multiway calibration. multilinear pls. *J. of Chemometrics 10,* 1, 47–61.

Bro, R. 1998. Multi-way analysis in the food industry: models, algorithms, and applications. Ph.D. thesis, University of Amsterdam, Amsterdam, Holland.

Bro, R. 2006. Review on multiway analysis in chemistry–2000-2005. *Critical Reviews in Analytical Chemistry 36*, 279–293.

Bro, R., Andersson, C. A., and Kiers, H. A. L. 1999. Parafac2 - part ii: Modeling chromatographic data with retention time shifts. *J. of Chemometrics 13,* 3-4, 295–309.

Bro, R., Harshman, R. A., and Sidiropoulos, N. D. 2005. Modeling multi-way data with linearly dependent loadings. *Technical Report 2005-176, KVL*.

Bro, R. and Kiers, H. A. L. 2003. A new efficient method for determining the number of components in parafac models. *J. of Chemometrics 17,* 5, 274–286.

Bro, R., Sidiropoulos, N. D., and Smilde, A. K. 2002. Maximum likelihood fitting using ordinary least squares algorithms. *J. of Chemometrics 16,* 8-10, 387–400.

Bro, R. and Smilde, A. K. 2003. Centering and scaling in component analysis. *J. of Chemometrics 17,* 1, 16–33.

Carlier, A., Lavit, C., Pages, M., Pernin, M., and Turlot, J. 1989. A comparative review of methods, which handle a set of indexed data tables. In *Multiway Data Analysis*. Elsevier, North-Holland, 85–101.

Carroll, J. D. and Chang, J. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of 'eckart-young' decomposition. *Psychometrika* 35, 218–319.

Cattell, R. B. 1944. Parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika* 9, 267–283.

Ceulemans, E. and Kiers, H. A. L. 2006. Selecting among three-mode principal component models of different types and complexities: A numerical convex-hull based method. *British Journal of Mathematical and Statistical Psychology 59,* 1, 133–150.

Chen, Z. P., Wu, H. L., Jiang, J. H., Li, Y., and Yu, R. Q. 2000. A novel trilinear decomposition algorithm for second-order linear calibration. *Chemometrics Intell. Lab. Systems 52,* 1, 75–86.

Eigenvector, R. 2006. Plstoolbox. In *http://www.eigenvector.com/*.

Engelen, S. and Hubert, M. 2005. A robust parafac method. In *International Conference on Robust Statistics*.

Estienne, F., Matthijs, N., Massart, D. L., Ricoux, P., and Leibovici, D. 2001. Multi-way modelling of high-dimensionality electroencephalographic data. *Chemometrics Intell. Lab. Systems 58,* 1, 59–72.

Faber, N. M., Bro, R., and Hopke, P. K. 2003. Recent developments in candecomp/ parafac algorithms: A critical review. *Chemometrics Intell. Lab. Systems 65,* 1, 119–137.

Golub, G. H. and Loan, C. F. V. 1996. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD.

Gourvnec, S., Stanimirova, I., Saby, C. A., Airiau, C. Y., and Massart, D. L. 2005. Monitoring batch processes with the statis approach. *J. of Chemometrics 19,* 5-7, 288–300.

Gourvnec, S., Tomasi, G., Durvillec, C., Crescenzo, E. D., Saby, C. A., Massart, D. L., Bro, R., and Oppenheim, G. 2005. Cubatch, a matlab interface for n-mode data analysis. *Chemometrics Intell. Lab. Systems 77,* 1-2, 122–130.

Harshman, R. A. 1970. Foundations of the parafac procedure:models and conditions for an 'explanatory' multimodal factor analysis. *UCLA working papers in phonetics* 16, 1–84.

Harshman, R. A. 1972. Parafac2: Mathematical and technical notes. *UCLA working papers in phonetics* 22, 30–44.

Harshman, R. A., Hong, S., and Lundy, M. E. 2003. Shifted factor analysis - part i: Models and properties. *J. of Chemometrics 17,* 7, 363–378.

Hong, S. and Harshman, R. A. 2003. Shifted factor analysis - part iii: N-way generalization and application. *J. of Chemometrics 17,* 7, 389–399.

Jiang, J. H., Wu, H. L., Li, Y., and Yu, R. Q. 2000. Three-way data resolution by alternating slice-wise diagonalization (asd) method. *J. of Chemometrics 14,* 1, 15–36.

Kargupta, H., Huang, W., Sivakumar, K., and Johnson, E. 2001. Distributed clustering using collective principal component analysis. *Knowledge and Information Systems Journal 3,* 4, 422–448.

KIERS, H. A. L. 1997. Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika 62*, 251–266.

KIERS, H. A. L. 2000. Towards a standardized notation and terminology in multiway analysis. *J. of Chemometrics 14,* 3, 105–122.

KIERS, H. A. L., BERGE, J. M. F. T., AND BRO, R. 1999. Parafac2 - part i: a direct fitting algorithm for the parafac2 model. *J. of Chemometrics 13,* 3-4, 275–294.

KIERS, H. A. L. AND KINDEREN, A. D. 2003. A fast method for choosing the numbers of components in tucker3 analysis. *British Journal of Mathematical and Statistical Psychology 56,* 1, 119–125.

KIM, J. AND MUELLER, C. W. 1978. *Introduction to factor analysis: What it is and how to do it.* Sage Publications, Newbury Park, CA.

KOLDA, T. G. 2001. Orthogonal tensor decompositions. *SIAM J. Matrix Anal. Appl. 23,* 1, 243–255.

KOLDA, T. G. 2003. Counterexample to the possibility of an extension of the eckart-young low-rank approximation theorem for orthogonal rank tensor decomposition. *SIAM J. Matrix Anal. Appl. 24,* 3, 762–767.

KOLDA, T. G. AND BADER, B. W. 2006. The tophits model for higher-order web link analysis. In *Workshop on Link Analysis, Counterterrorism and Security.*

KOLDA, T. G., BADER, B. W., AND KENNY, J. P. 2005. Higher-order web link analysis using multilinear algebra. In *Proc. of The 5th IEEE International Conference on Data Mining.* 242–249.

KRUSKAL, J. B. 1989. Rank decomposition, and uniqueness for 3-way and n-way arrays. In *Multiway Data Analysis.* Elsevier, North-Holland, 8–18.

LATHAUWER, L. D., MOOR, B. D., AND J.VANDEWALLE. 2000. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl. 21,* 4, 1253–1278.

LATHAUWER, L. D., MOOR, B. D., AND VANDEWALLE, J. 2000. On the best rank-1 and rank-$(r_1, r_2, ..., r_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl. 21,* 4, 1324–1342.

LAVIT, C., ESCOUFIER, Y., SABATIER, R., AND TRAISSAC, P. 1994. The act (statis method). *Computational Statistics and Data Analysis 18,* 1, 97–119.

LEVY, A. AND LINDENBAUM, M. 2002. Sequential karhunen-loeve basis extraction and its applications to images. *IEEE Trans. Image Processing 9,* 8, 1371–1374.

LIM, L. H. 2005. Singular values and eigenvalues of tensors: a variational approach. In *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP).* 129–132.

MAHONEY, M. W., MAGGIONI, M., AND DRINEAS, P. 2006. Tensor-cur decompositions for tensor-based data. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 327–336.

MATHWORKS. Matlab. In *http:www.mathworks.com.*

MENG, X., MORRIS, A. J., AND MARTIN, E. B. 2003. On-line monitoring of batch processes using a parafac representation. *J. of Chemometrics 17,* 1, 65–81.

MIWAKEICHI, F., MARTNEZ-MONTES, E., VALDS-SOSA, P., NISHIYAMA, N., MIZUHARA, H., AND YAMAGUCHI, Y. 2004. Decomposing eeg data into space-time-frequency components using parallel factor analysis. *NeuroImage 22,* 3, 1035–1045.

MØRUP, M., HANSEN, L. K., AND ARNFRED, S. M. 2006. ERPWAVELAB a toolbox for multi-channel analysis of time-frequency transformed event related potentials. *(Preprint) accepted for publication in Journal of Neuroscience Methods.*

MØRUP, M., HANSEN, L. K., HERMANN, C. S., PARNAS, J., AND ARNFRED, S. M. 2006. Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage 29,* 3, 938–947.

MØRUP, M. AND SCHMIDT, M. N. 2006. Sparse nonnegative tensor 2d deconvolution (sntf2d) for multichannel time-frequency analysis. *Technical Report, Technical University of Denmark, DTU.*

PAATERO, P. 1999. The multilinear engine - a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. of Computational and Graphical Statistics 8,* 4, 854–888.

PRAVDOVA, V., ESTIENNE, F., WALCZAK, B., AND MASSART, D. L. 2001. A robust version of the tucker3 model. *Chemometrics Intell. Lab. Systems 59,* 1-2, 75–88.

SHAWE-TAYLOR, J. AND CRISTIANINI, N. 2004. *Kernel Methods for Pattern Analysis.* Cambridge University Press, NewYork, NY.

SMILDE, A. K., BRO, R., AND GELADI, P. 2004. *Multiway Analysis. Applications in the chemical sciences.* Wiley, England.

SMILDE, A. K., WESTERHUIS, J. A., AND BOQU, R. 2000. Multiway multiblock component and covariates regression models. *J. of Chemometrics 14,* 3, 301–331.

STANIMIROVA, I., WALCZAK, B., MASSART, D. L., SIMEONOV, V., SABY, C. A., AND CRESCENZO, E. D. 2004. Statis, a three-way method for data analysis. application to environmental data. *Chemometrics Intell. Lab. Systems 73,* 2, 219–233.

SUN, J., TAO, D., AND FALOUTSOS, C. 2006. Beyond streams and graphs: dynamic tensor analysis. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 374–383.

SUN, J. T., ZENG, H. J., LIU, H., LU, Y., AND CHEN, Z. 2005. Cubesvd: A novel approach to personalized web search. In *Proc. of the 14th International World Wide Web conference.* 382–390.

Timmerman, M. E. and Kiers, H. A. L. 2000. Three mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology 53,* 1, 1–16.

Tomasi, G. and Bro, R. 2006. A comparison of algorithms for fitting the parafac model. *Computational Statistics and Data Analysis 50,* 7, 1700–1734.

Tucker, L. R. 1964. The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology.* Holt, Rinehart and Winston, New York, 110–182.

Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika 31*, 279–311.

Vasilescu, M. A. O. and Terzopoulos, D. 2002. Multilinear image analysis for facial recognition. In *Proc. of International Conference on Pattern Recognition.* Vol. 2. 511–514.

Vega-Montoto, L. and Wentzell, P. D. 2003. Maximum likelihood parallel factor analysis. *J. of Chemometrics 17,* 4, 237–253.

Vorobyov, S. A., Rong, Y., Sidiropoulos, N. D., and Gershman, A. B. 2005. Robust iterative fitting of multilinear models. *IEEE Trans. Signal Processing 53,* 8, 2678–2689.

Wang, H. and Ahuja, N. 2004. Compact representation of multidimensional data using tensor rank-one decomposition. In *Proc. of International Conference on Pattern Recognition.* Vol. 1. 44–47.

Y. Li, Y. D. and Lin, X. 2005. Kernel-based multifactor analysis for image synthesis and recognition. In *Proc. of International Conference on Computer Vision.* Vol. 1. 114–119.

Zhang, T. and Golub, G. H. 2001. Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl. 23,* 2, 534–550.