J Math Chem (2014) 52:2546–2558
DOI 10.1007/s10910-014-0396-z

ORIGINAL PAPER

# Reductive enhanced multivariance product representation for multi-way arrays

**Evrim Korkmaz Özay · Metin Demiralp**

Received: 3 July 2014 / Accepted: 19 July 2014 / Published online: 17 August 2014


**Abstract** The main purpose of this work is to develop a new multi-way decomposition technique by considering statistical structure of target multi-way array. To this end enhanced multivariance products representation (EMPR), which is an expansion extended from high dimensional model representation (HDMR), is used. EMPR provides quite successful results on representation and approximation of multivariate functions when they have high level multiplicativity. Hence this urges us to reconstruct EMPR as a multi-way array decomposer. This paper presents this decomposition technique with all reconstruction formulations and numerical experiments on synthetic and real-life data sets to denote EMPR's efficiency as a decomposer and also presents a combined method Reductive-EMPR (R-EMPR) as a multi-way array decomposition technique.

**Keywords** High dimensional data · Multi-way array decomposition · Enhanced multivariance product representation · Reductive array decomposition

## 1 Introduction

High dimensional data modelling became an indispensable subject in applied science with the development of technology used in the collection of data. Thus, there are a lot of algorithms and models which have been developed to tackle with high dimensional data sets. Sometimes high dimensional data sets are processed as multi-way arrays

E. K. Özay (✉) · M. Demiralp
Computational Science and Engineering Program, Group for Science and Methods of Computation, Informatics Institute, İstanbul Technical University, Maslak 34469, Istanbul, Turkey
e-mail: korkmazevr@itu.edu.tr

M. Demiralp
e-mail: demiralp@itu.edu.tr

to capture relationships between the features more accurately. A multi-way array is a type of high-dimensional data with multi-indices. Multi-way arrays (N-way arrays, tensors) are the generalization of vectors (1-way arrays) and matrices (2-way arrays) in a simple definition. Here we prefer to say ‘multi-way array’instead of ‘tensor’ to avoid confusion with the entities used in the areas like continuum mechanics. Those tensors are required to satisfy certain transformational properties which do not need to exist in multi-way arrays. To give a clear explanation it can be stated that a multi-way array is somehow product of vector spaces of each way [1]. Multi-way array applications can be found in all areas which include multi-way data. Chemometrics, neuroscience, psychometry, data mining are some of these areas [2]. Even though multi-way arrays are not subject of linear algebra, there are certain similarities between the matrix-vector operations and multi-way arrays, for instance, the norm of a multi-way array can be evaluated just like Frobenius norm in ordinary linear algebra [3]. Let $\mathcal{X}$ be an $N$-way array which has dimensions $I_1 \times I_2 \times \cdots \times I_N$, then its norm can be defined as follows

$$\| \mathcal{X} \| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2}, \quad \mathcal{X} \in \mathcal{R}^{I_1 \times I_2 \times \cdots \times I_N} \tag{1}$$

Inner product definition has same analogy with the well-known inner product of ordinary linear algebra. The inner product of two different multi-way arrays which have same number of ways and same number of dimensions on related ways can be evaluated just like the one below

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N} y_{i_1 i_2 \ldots i_N}, \quad \mathcal{X}, \mathcal{Y} \in \mathcal{R}^{I_1 \times I_2 \times \cdots \times I_N} \tag{2}$$

Although there are different kind of multiplications and features, such as, Kronecker product, Hadamard product, Khatri-Rao product, $n$-mode product and rank, they are not explained here since our method does not require these entities explicitly and they can be found in reference [4] by curious readers. Also there are different kinds of decomposition techniques for multi-way arrays to work with less data or to understand the relations between the ways. Since multi-way array is a data type with high dimensionality, dimensionality reduction is naturally the most important application on this subject just like matrix decomposition techniques. However, due to the special form of multi-way array’s dimensionality, matrix decomposition techniques may be insufficient for dimensionality reduction. Therefore matrix decomposition techniques were extended and approach is specialized for multi-way arrays. However this study aims to develop a new technique by using enhanced multivariance products representation (EMPR) which is a recently developed algorithm used for multivariate function de composition exactly and mostly approximately. Next section contains preliminaries of mathematical background and literature for the method we propose.

## 2 Mathematical background

### 2.1 Decomposition techniques

Decomposition is a basic process for high dimensional data analysis, hence it, almost always, remains a popular topic for researchers. In many applications researchers had formerly used matrix decomposition methods for all kinds of data set regardless of the way concept. This approach may cause information loss for multi-way data even though it may cause wrong inference for noisy multi-way data [5]. Therefore the latter decomposition techniques are extended for multi-way arrays with inspiration from matrix decomposition techniques such as singular value decomposition (SVD), principal component analysis (PCA), factor analysis (FA) [6–8]. Multi-way array decomposition has place in many application areas, from computational chemistry to neuroscience, however root s of multi-way array decomposition can be found at Hitchcock's studies [9]. Different two basic decomposition techniques were developed afterwards by Cattel [10,11] and Tucker [12,13]. These methods motivate researchers for developing new decomposition techniques and for using multi-way ar rays more frequently in many applications. Nowadays there are many decomposition techniques which arise from these two basics and alternative methods are used in applications [14, 15]. These techniques usually handle multi-way data with respect to spectral pattern of data set. For example Tucker decomposition for a target multi-way array gives factor matrices as the number of ways and a core multi-way array, but still it is possible to gain a new algorithm by considering some constraints and condition s. As an example let $\mathcal{X}$ be a 3-way array and its dimensions be $I \times J \times K$, $\mathbf{A} \in \mathbb{R}^{I\times P}$, $\mathbf{B} \in \mathbb{R}^{J\times Q}$ and $\mathbf{C} \in \mathbb{R}^{K\times R}$ are factor matrices and $\mathcal{G} \in \mathbb{R}^{P\times Q\times R}$ is core array. Then Tucker decomposition for $\mathcal{X}$ can be written as follows

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{i=1}^{P} \sum_{j=1}^{Q} \sum_{k=1}^{R} g_{ijk} \mathbf{a}_i \mathbf{b}_j \mathbf{c}_k, \tag{3}$$

where $\times_i$, $i = 1, 2, 3$ show mode product of a multi-way array with a matrix according to related way. It is possible to find components of Tucker decomposition with different condition s on factor matrices such as orthogonality [16], best rank approximation [17] or Bayesian perspective [18]. Another commonly used method, canonical polyadic decomposition (CP), de compose a multi-way array as finite sum of rank-one tensors [19–21] and it can be considered as the generalization of SVD. As an example CP decomposition for a 3-way array is as follows

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \tag{4}$$

where $r$ stands for the rank of a multi-way array and $\circ$ shows outer product of vectors, $\mathbf{a}_r, \mathbf{b}_r, \mathbf{c}_r$. CP method decomposes a multi-way array with respect to rank of array which is shown as $R$ above. Although there are different rank definitions for multi-way arrays, here, it is defined as the number of terms that formed as outer product of each way's vector. All the methods described so far are obtained through procedures based on the features of multi-linear algebra, but it is possible to find a different way for decomposition. So we propose a new approach to decompose multi-way data considering statistical relationships between ways of the data by using EMPR. The following section explains general features and analytical structure of EMPR.

### 2.2 Enhanced multivariance products representation

High Dimensional Model Representation (HDMR) based algorithms are used recently at different scientific areas which have high dimensional data. Actually HDMR was designed to approximate a multivariate function and was based on divide-and conquer philosophy by Sobol [22], then it was extended by Rabitz [24] and Demiralp[23]. Mathematical expression of HDMR is as follows

$$f\left(x_1,\ldots,x_N\right)=f_0+\sum_{i=1}^{N} f_i\left(x_i\right)+\sum_{\substack{i_1,i_2=1\\ i_1<i_2}}^{N} f_{i_1,i_2}\left(x_{i_1},x_{i_2}\right)\cdots+f_{12\cdots N}\left(x_1,\ldots,x_N\right) \tag{5}$$

where $f$ is an analytical function with $N$ independent variables, $f_0$ is constant component of HDMR, $f_i\left(x_i\right)$'s are univariate components for $i=1,\ldots,N$ and the other components have increasing number of variable dependencies. To uniquely determine components at the right hand side of the Eq. (5) certain conditions should be imposed. These constraints are defined also for EMPR under the existence of support functions which are univariate functions multiplied by components of HDMR and they are explained for EMPR of multi-way arrays in the next section. Enhanced multivariance products representation was developed by Demiralp to get better approximation and to overcome some weaknesses of HDMR [25,26]. Although different methods based on HDMR have been developed for different kind of functions [27–29], EMPR can be considered as a generalization of HDMR. The main reason for developing EMPR is additive nature of HDMR. We know that HDMR works well in approximating additive natures. The structure of EMPR having similar philosophy with HDMR increases the performance of that method and makes this divide-and-conquer algorithm applicable to all kinds of analytical structures. Another main property of HDMR that distinguishes HDMR from the other expansions (Taylor, Maclaurin, etc…) is to be finite, in other words, HDMR decomposes a multivariate function starting from a constant component end ing with an $N$-dimensional component but EMPR has the same multivariance at each term with the help of support functions ($s_j(x_j)$, $j=1,\ldots,N$) as shown in the following equation for an $N$-variate function, $f$.

$$f\left(x_1,\ldots,x_N\right)=f_0\prod_{j=1}^{N}s_j\left(x_j\right)+\sum_{i=1}^{N}f_i(x_i)\prod_{\substack{j=1\\ j\neq i}}^{N}s_j\left(x_j\right)$$

$$+\sum_{\substack{i_1,i_2=1\\ i_1<i_2}}^{N} f_{i_1,i_2}\left(x_{i_1},x_{i_2}\right)\prod_{\substack{j=1\\ j\neq i_1,i_2}}^{N} s_j\left(x_j\right)+\cdots \tag{6}$$
$$+f_{12\cdots N}\left(x_1,x_2,\cdots,x_N\right).$$

The existence of support functions empowers the quality of approximation to multivariate functions. Here support functions are univariate functions and derived from $f\left(x_1,\ldots,x_N\right)$ itself, but they can be adapted to different kind of structures or they can be chosen manually so to speak if there is some knowledge about the structure of target multivariate function. These powerful and flexible aspects of EMPR urge us to apply this method to discrete structures like multi-way arrays. Next section describes EMPR's general structure for multi-way arrays and also introduces a new method for choosing support terms of EMPR.

## 3 Enhanced multivariate products representation as a multi- way array decomposer

### 3.1 Discrete enhanced multivariate products representation

Enhanced multivariance products representation, which is an ex tended version of high dimensional model representation (HDMR), has been used for discrete structures in recent years [30,31]. Based on this development, EMPR is applied to multi-way arrays and the preliminary results are quite remarkable, so that it is imperative to explain EMPR algorithm and the progress on multi-way arrays. Mathematical structure of EMPR on multi-way arrays is similar to application on continuous functions. However it is needed to bring it into line with mathematical formula of EMPR on continuous functions, so EMPR expansion on multi-way array is explicitly shown be low

$$\mathcal{X}_{i_1\ldots i_N}=\prod_{j=1}^{N} s_{i_j}^{(j)}\mathcal{X}^{(0)}+\sum_{j_1=1}^{N}\mathcal{X}_{i_{j_1}}^{(j_1)}\prod_{\substack{j=1\\ j\neq j_1}}^{N} s_{i_j}^{(j)}+\sum_{\substack{j_1,j_2=1\\ j_1<j_2}}^{N}\mathcal{X}_{i_{j_1},i_{j_2}}^{(j_1j_2)}\prod_{\substack{j=1\\ j\neq j_1,j_2}}^{N} s_{i_j}^{(j)}$$
$$+\cdots+\mathcal{X}_{i_{j_1}i_{j_2}\ldots i_{j_N}}^{(j_1j_2\ldots j_N)},\qquad i_j=1,2,\ldots,n_j,\quad j=1,2,\ldots,N \tag{7}$$

where $\mathcal{X}$ is a multi-way array of $N$ way (or indexes) and $\mathcal{X}_{i_1\ldots i_N}$ is the element with $i_1\ldots i_N$ indices, $\mathcal{X}_0$ shows constant term of EMPR, $\mathcal{X}^{(j_1)}$, $j_1=1,\ldots N$'s are one-way arrays, and the other terms on the right hand side are the arrays which have increasing ways. $s^{(j)}$'s are the support terms of EMPR expansion and they ensure the number of way's equality for each term. Thus, when applying truncation to EMPR expansion it is possible to get knowledge about all ways of multi-way array. At this point, two important terms of EMPR must be calculated, and to do that, some conditions and constraints are defined. In fact these conditions show similarity with the EMPR on continuous functions. First of them is weight array which is composed of outer products of each way's one-way weight arrays. Outer product of vectors refers to tensor product that increase the dimensionality of the result. Besides, normalization condition is still valid for multi-way arrays.

$$\sum_{i_j=1}^{n_j} W_{i_j}^{(j)} = 1, \quad W_{i_1 \ldots i_N} \equiv \prod_{j=1}^{N} W_{i_j}^{(j)}, \qquad j = 1, 2, \ldots, N \tag{8}$$

where $W_{i_j}^{(j)}$'s are positive. In addition to the above conditions, EMPR's each support term must have unit weighted norm. This property provides determination of EMPR components uniquely.

$$\sum_{i_j=1}^{n_j} W_{i_j}^{(j)} \left(s_{i_j}^{(j)}\right)^2 = 1, \qquad j = 1, 2, \ldots, N \tag{9}$$

To determine the EMPR components uniquely certain other conditions are needed. To this end, we introduce the vanishing conditions as follows

$$\sum_{i_{j_l}=1}^{n_{j_l}} W_{i_{j_l}}^{(j_l)} s_{i_{j_l}}^{(j_l)} \mathcal{X}_{i_{j_1} \ldots i_{j_k}}^{(i_{j_1} \ldots i_{j_k})} = 0$$
$$l = 1, 2, \ldots, k, \qquad k = 1, 2, \ldots, N \tag{10}$$

As a matter of fact, EMPR's constant term can be considered as projection which projects all the multi-way array to the weighted and supported mean of this array under certain weight arrays and support terms. Constant term of EMPR for a multi-way array can be determined as follows.

$$\mathcal{X}^{(0)} = \sum_{j_1=1}^{n_1} \cdots \sum_{j_N=1}^{n_N} W_{j_1}^{(1)} \cdots W_{j_N}^{(N)} \prod_{k=1}^{N} s_{j_k}^{(k)} \mathcal{X}_{j_1 \ldots j_N} \tag{11}$$

One-way components of EMPR for a multi-way array also get the information from all ways except the focused one. Each one-way component excludes the way relevant to itself and takes the mean of the multi-way array under related weight arrays and support arrays over the other ways.

$$\mathcal{X}_{i_k}^{(k)} = \sum_{j_1=1}^{n1} \cdots \sum_{j_N=1}^{n_N} W_{j_1}^{(1)} \cdots W_{j_N}^{(N)} \frac{\delta_{i_k j_k}}{W_{j_k}^{(k)}} \prod_{\substack{l=1 \\ l \neq k}}^{N} s_{i_l}^{(l)} \mathcal{X}_{j_1 \ldots j_N} - \mathcal{X}_{j_1 \ldots j_N}^{(0)}$$
$$i_k = 1, 2, \ldots, n_k, \qquad k = 1, 2, \ldots, N, \qquad l = 1, 2, \ldots, n_l \tag{12}$$

EMPR components with two or more ways can be found within the same philosophy. However in this study it is desired to get best approximant for target multi-way array by calculating EMPR terms which have at most two-way components, so determination of higher-order terms is not explicitly given here. The main problem of building an EMPR algorithm for multi-way arrays is determination of the support terms. The outer product of all support terms should reflect the pat tern of target array as much as possible. As a matter of fact the functional structure of the elements of the given multi-way array in the indices is not known so the best and easiest way is to use the multi-way array itself. For this purpose, under a certain weight array, EMPR's support term of a way is formulated as follows.

$$s_{i_j}^{(j)} = \frac{\sum_{i_1=1}^{n_1} \cdots \sum_{i_{j-1}=1}^{n_{j-1}} \sum_{i_{j+1}=1}^{n_{j+1}} \cdots \sum_{i_N=1}^{n_N} w_{i_1}^{(1)} \cdots w_{i_{j-1}}^{(j-1)} w_{i_{j+1}}^{(j+1)} \cdots w_{i_N}^{(N)} \mathcal{X}_{i_1,\ldots,i_N}}{\left( w_{i_j}^{(j)} \left[ \sum_{i_1=1}^{n_1} \cdots \sum_{i_{j-1}=1}^{n_{j-1}} \sum_{i_{j+1}=1}^{n_{j+1}} \cdots \sum_{i_N=1}^{n_N} w_{i_1}^{(1)} \cdots w_{i_{j-1}}^{(j-1)} w_{i_{j+1}}^{(j+1)} \cdots w_{i_N}^{(N)} \mathcal{X}_{i_1,\ldots,i_N} \right]^2 \right)^{\frac{1}{2}}} \tag{13}$$

This kind of expression is easy to calculate but it reminds us whether a better way exists. To answer this question optimization works on support terms are in progress to get a powerful decomposition technique. A brand new idea that this research brings is using reductive decomposition method for multilinear arrays (RDMMA) to determine support terms which is described in the next subsection.

### 3.2 Reductive enhanced multivariance product representation

RDMMA reduces the dimensionality of an array one by one and it has been recently applied to animation data sets [32,33]. According to RDMMA an approximated array is constructed with product of two multi-way arrays as follows. Let $\mathcal{M}$ be an approximate array for original $N$ way array, $\mathcal{X}$, then $\mathcal{M}$ consists of two arrays' outer product ($\circ$), one way array $x$ and $(N-1)$ way array $B$.

$$\mathcal{M} \equiv B \circ x, \qquad B \in \mathcal{H}_{N-1}, \qquad x \in \mathcal{H}_1 \tag{14}$$

where the relevant Hilbert spaces are denoted by the symbol $\mathcal{H}$. The element representation of this product can be shown as

$$\mathcal{M}_{i_1,\ldots,i_N} = B_{i_1,\ldots,i_{N-1}} x_{i_N}. \tag{15}$$

It is possible to find $B$ and $x$ by constructing an optimization problem under certain conditions such as taking the norms of the arrays as one, $\|\mathcal{M}\| = 1$, $\|B\| = 1$, $\|x\| = 1$. Under these conditions we can use the following cost functional to proceed

$$\mathcal{J} = \Delta(\sigma, B, x) + \lambda_1(\|B\|^2 - 1) + \lambda_2(\|x\|^2 - 1) \tag{16}$$

where

$$\begin{aligned} \Delta(\sigma, B, x) &= \|D\|^2 \\ D &\equiv \mathcal{X} - \sigma B \circ x \end{aligned} \tag{17}$$

and $\lambda_1$, $\lambda_2$ are Lagrange Multipliers to be determined. This expresses the distance square between the target function and the approximant plus Lagrangian constraints terms. This optimization problem can be solved in two steps. As the first step, the first derivatives of the above cost functional with respect to $\sigma$, $\lambda_1$, $\lambda_2$ are calculated, then the gradients with respect to $B$ and $x$ are calculated in the second step. After algebraic calculations and substitutions a covariance-like matrix, $\mathbf{C}$ is defined through $\mathcal{X}$ as follows

$$C_{ij} \equiv \sum_{i_1=1}^{n_1} \cdots \sum_{i_{N-1}=1}^{n_{N-1}} \mathcal{X}_{i_1,\ldots,i} \mathcal{X}_{i_1,\ldots,j}, \quad i, j = 1, 2, \ldots, n_N \tag{18}$$

The above matrix is in fact a coefficient matrix for an eigenvalue problem which gives the desired one-way array, $x$

$$\mathbf{Cx} = \sigma^2 \mathbf{x} \tag{19}$$

All the calculations bring the other member which is an $(N-1)$-way array to us and its elements can be evaluated as below.

$$B_{i_1,\ldots,i_{N-1}} = \frac{1}{\sigma} \sum_{i_N=1}^{n_N} \mathcal{X}_{i_1,\ldots,i_N} x_{i_N}, \tag{20}$$

The proofs of the above method can be found in related reference [34]. A brand new idea of this work is to combine EMPR with the method explained above for multi-way array decomposition. The combination of these two methods yields a decomposition technique which can detect statistical structure of the data and also gets spectral perspective for the same data. Combination process begins with getting multi-way array, for example let $\mathcal{X}$ be a 3-way array and $B^{(3)}$, $x^{(3)}$ be component s of RDMMA for the way chosen as the first way of the array. If approximated array is named as $\mathcal{X}_{app}$ to refer an approximation then it can be shown as follows.

$$\mathcal{X}_{app} = \sigma_1 B^{(1)} \circ x^{(1)} \tag{21}$$

If the difference between original array and $\mathcal{X}_{app}$ and the orthogonality between $\mathcal{X}_{app}$ and difference array is taken into consideration as a new constraint then RDMAA is applied to the difference array again but this time for the second way, and then the expression below becomes appropriate for the new approximation.

$$\mathcal{X}_{app} = \sigma_1 B^{(1)} \circ x^{(1)} + \sigma_2 B^{(2)} \circ x^{(2)} \tag{22}$$

Difference between the second approximation and the original array is again processed with RDMMA and a decomposition which has two-way and one-way components for all ways is obtained.

$$\mathcal{X} \approx \sigma_1 B^{(1)} \circ x^{(1)} + \sigma_2 B^{(2)} \circ x^{(2)} + \sigma_3 B^{(3)} \circ x^{(3)} \tag{23}$$

Here $\sigma_1, \sigma_2, \sigma_3$ stand for scaling parameters of relevant components which have unit array norms. Obtained one-way arrays from Eq. (23), $(x^{(1)}, x^{(2)}, x^{(3)})$, can now be used as support terms of EMPR.

Many data types obtained from nature is referred to as 3-way so our combined method, reductive enhanced multivariance product representation (R-EMPR), is explained for three-way arrays for the ease of explanation, even so it can be applied for arrays that have more than three ways. Other important issues to be mentioned about R-EMPR are selecting directions and deciding turn of reduction. On the example above reduction starts from a determined way by user and goes on in the same way, directions can be selected not only arbitrarily. They can be selected on how the obtained approximation quality is. At this stage of this study it is also important to see the behaviour of EMPR and R-EMPR on real life multi-way data sets. To this end next section examines some 3-way arrays from synthetic data sets and data sets from experiments.

## 4 Numerical implementations

With these numerical experiments in this section it is aimed to see behaviours of EMPR on different multi-way data sets and to compare performance of our method against two main and commonly used algorithms, Tucker decomposition and CP decomposition. First of all some synthetic 3-way data sets with additive and multiplicative structures are produced by changing additivity and the dimensions of multi-way arrays. Using synthetic data set with additive nature is import ant to see whether EMPR represent the data set when non-additive behaviour increases. The algorithm coded on MATLAB and EMPR's zeroth, fir st and second order approximants' performance were measured by relative error. Table 1 shows the relative errors of EMPR approximants which produced from additive data sets and Table 2 shows results for more multiplicative 3-way data sets.

According to the results, when the additivity decreases in data set then the approximation quality decreases too. There is also same kind of relation between the dimensions of ways and the relative error, but this time increasing dimension make less efficient approximation.

In this study we also aim to see how our algorithm represents data against the other two multi-way array decomposition algorithms. For this purpose real-life data sets are taken from different chemical experiments and Tables 3, 4, and 5 show the results. Tucker and CP decomposition is applied by using Matlab Tensor Toolbox

**Table 1** Approximation results of discrete EMPR for additive synthetic data sets

| Datasets | Relative $\text{Err}_0$ | Relative $\text{Err}_1$ | Relative $\text{Err}_2$ |
|---|---|---|---|
| $\mathcal{X}^{(3,4,5)}, \mathcal{X}(i,j,k) = i + j + k$ | 0.0372 | 0.0357 | 0.0057 |
| $\mathcal{X}^{(6,8,10)}, \mathcal{X}(i,j,k) = i + j + k$ | 0.0479 | 0.0454 | 0.0084 |
| $\mathcal{X}^{(12,16,20)}, \mathcal{X}(i,j,k) = i + j + k$ | 0.0541 | 0.0508 | 0.0100 |

**Table 2** Approximation results of discrete EMPR for multiplicative synthetic data sets

| Datasets | Relative $\text{Err}_0$ | Relative $\text{Err}_1$ | Relative $\text{Err}_2$ |
|---|---|---|---|
| $\mathcal{X}^{(3,4,5)}, \mathcal{X}(i,j,k) = i^2 + j^2 + k^2$ | 0.1011 | 0.0898 | 0.0213 |
| $\mathcal{X}^{(6,8,10)}, \mathcal{X}(i,j,k) = i^2 + j^2 + k^2$ | 0.1183 | 0.1029 | 0.0265 |
| $\mathcal{X}^{(12,16,20)}, \mathcal{X}(i,j,k) = i^2 + j^2 + k^2$ | 0.1266 | 0.1091 | 0.0290 |

**Table 3** Relative errors for amino dataset

| Tucker Alg. | CP Alg. | $\text{EMPR}_0$ | $\text{EMPR}_1$ | $\text{EMPR}_2$ | R-$\text{EMPR}_2$ |
|---|---|---|---|---|---|
| 0.5996 | 0.6465 | 0.6179 | 0.5988 | 0.2147 | 0.0955 |

**Table 4** Relative errors for nose dataset

| Tucker Alg. | CP Alg. | $\text{EMPR}_0$ | $\text{EMPR}_1$ | $\text{EMPR}_2$ | R-$\text{EMPR}_2$ |
|---|---|---|---|---|---|
| 0.0323 | 0.1398 | 0.0411 | 0.0328 | 0.0032 | 0.0059 |

**Table 5** Relative errors for sugar-process dataset

| Tucker A. | CP A. | $\text{EMPR}_0$ | $\text{EMPR}_1$ | $\text{EMPR}_2$ | R-$\text{EMPR}_2$ |
|---|---|---|---|---|---|
| 0.6345 | 0.7061 | 0.6413 | 0.6352 | 0.1790 | 0.1381 |

[35]. After getting two decompositions with their best rank for each data set then decomposed array reconstruct ed and the difference between the original multi-way array and the reconstructed array is evaluated. The first of real-data samples is ami no acid data set which was used by Bro [36]. Amino acid data set includes five samples and each sample contains different amount s of tyrosine, tryptophan and phenylalanine dissolved in phosphate buffered water. These samples were measured by fluorescence with excitation 250–300 nm, emission 250–450 nm, 1 nm intervals and the dimensions of data set $5 \times 51 \times 201$.

Second chosen data set is taken from sensor based investigation, Nose data set. Data set is composed of three way as sample, time and sensor. The main purpose of the collecting this data set is to differentiate good licorices and bad licorices with the help of an electronic nose combined with multivariate chemometrics tools [37]. Also the data set ordered as sample$\times$ time $\times$ sensor and dimensions are $18 \times 241 \times 12$.

Third data set was obtained from sugar dissolved in un-buffered water and measured on spectrofluorometer [38]. Dimensions of the data are $265 \times 571 \times 7$. The first way shows samples, the second way shows emission wavelengths and the last way is for excitation wavelengths. All of the datasets can be obtained from relevant reference [39].

According to numerical results second order approximation with EMPR provides very powerful approximation for different kind of data sets. Even the first order approximation of EMPR gives results as good as Tucker Decomposition and CP Decomposition. Another observation onto numerical experiments R-EMPR may optimize the results of plain EMPR. Here given R-EMPR results show second order approximations. Because we want to see whether R-EMPR is better than the best EMPR approximation. The approximation degree has been taken as second order for R-EMPR.

## 5 Concluding remarks

This study shows the performance on multi-way arrays of two novel decomposition algorithm, EMPR and R-EMPR. They can represent multi -way data sets as good as other methods which are frequently used in the literature for decomposing multi-way arrays. Numerical results show us R-EMPR has the power to optimize the EMPR's results. In this work we have used both synthetic data sets and data sets from some chemical experiments to show the approximation quality and compare with two other methods. Second order approximants of EMPR and R-EMPR provide more powerful approximations than zeroth degree and first degree approximations. This means that the second order approximation represents the data set quite good and there fore second order approximations of these methods can be alternative decomposition techniques for at least three-way arrays or arrays have more than three ways.

Another subject to be emphasized is the selection of the ways when R-EMPR is applied iteratively. They can be chosen arbitrarily just as we mentioned above but it would give more powerful results, if the decision of R-EMPR's way can be made according to a criterion for that the success of the approximation. Another finding of this work is the selection of support terms can be modified by using RDMMA's one-way components. And this finding provides diversity for both algorithm's structure and performance value. Although the preliminary results have been given by using truncation, reconstruction and approximation, it is obvious that EMPR and R-EMPR will find their place in certain applications on multi-way arrays.

**Acknowledgments** First author is grateful to N. Abdülbaki Baykara and Mehmet Alper Tunga for their careful reading of the manuscript and their invaluable comments. First author also thanks Scientific Research Projects Unit of Istanbul Technical University for technical supports.

## References

1. T.G. Kolda, B.W. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
2. M. Mrup, Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **1**(1), 24–40 (2011)
3. G.H. Golub, C.F. Van Loan, *Matrix Computations, 55* (Johns Hopkins University Press, Baltimore, 1983)
4. A. Smilde, R. Bro, P. Geladi, *Multi-Way Analysis: Applications in the Chemical Sciences* (Wiley, West Sussex, 2004)
5. A. Cichocki, R. Zdunek, A.H. Phan, S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation* (Wiley, UK, 2009), pp. 1–8
6. G.H. Golub, C. Reinsch, Singular value decomposition and least squares solutions. Numer. Math. **14**(5), 403–420 (1970)

7. S. Wold, K. Esbensen, P. Geladi, Principal component analysis. Chemom. Intell. Lab. Syst. **2**(1), 37–52 (1987)
8. B. Thompson, *Exploratory and Confirmatory Cactor Analysis: Understanding Concepts and Applications* (American Psychological Association, Washington, DC, 2004)
9. F.L. Hitchcock, Multiple invariants and generalized rank of a p-way matrix or tensor. J. Math. Phys. **7**, 39–79 (1927)
10. R.B. Cattel, Parallel proportional profiles and other principles for determining the choice of factors by rotation. Psychometrika **9**, 267–283 (1944)
11. R.B. Cattel, The three basic factor-analytic research designs their interrelations and derivatives. Psychol. Bull. **49**, 499–452 (1952)
12. L.R. Tucker, Implications of factor analysis of three-way matrices for measurement of change, in *Problems in Measuring Change* (University of Wisconsin Press, 1963), pp. 122–137
13. L.R. Tucker, The extension of factor analysis to three-dimensional matrices, in *Contributions to Mathematical Psychology* (1964), pp. 109–127
14. P. Paatero, The multilinear enginea table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. J. Comput. Graph. Stat. **8**(4), 854–888 (1999)
15. I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, C.A. Saby, E.D. Crescenzo, Statis, a three-way method for data analysis. Application to environmental data. Chemom. Intell. Lab. Syst. **73**(2), 219–233 (2004)
16. B. De Moor, L. De Lathauwer, J. Vandewalle, A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000)
17. L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(R 1, R 2,., Rn) approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000)
18. Z. Xu, F. Yan, A. Qi, InfTucker: t-process based infinite tensor decomposition, arXiv:1108.6296 (2011)
19. J.D. Carroll, J.J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart–Young decomposition. Psychometrika **35**, 283–319 (1970)
20. R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis. UCLA Work. Pap. Phon. **16**, 1–84 (1970)
21. H.A.L. Kiers, Towards a standardized notation and terminology in multiway analysis. J. Chemom. **14**, 105–122 (2000)
22. I.M. Sobol, Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exp. (MMCE) **1**(4), 407–414 (1993)
23. M. Demiralp, High dimensional model representation and its application varieties. Math. Res. **9**, 146–159 (2003)
24. H. Rabitz, Ö. Alış, General foundations of high dimensional model representations. J. Math. Chem. **25**, 19–233 (1999)
25. B. Tunga, M. Demiralp, The influence of the support functions on the quality of enhanced multivariance product representation. J. Math. Chem. **48**(3), 827–840 (2010)
26. B. Tunga, M. Demiralp, Support function influences on the univariance of the enhanced multivariance product representation, in *AIP Conference Proceedings*, vol. 1504 (2012)
27. M.A. Tunga, M. Demiralp, A factorized high dimensional model representation on the partitioned random discrete data. Appl. Numer. Anal. Comput. Math. **1**, 231–241 (2004)
28. M.A. Tunga, M. Demiralp, A factorized high dimensional model representation on the nodes of a finite hyperprismatic regular grid. Appl. Math. Comput. **164**, 865–883 (2005)
29. M. Demiralp, Logarithmic high dimensional model representation, in *Proceedings of 6th WSEAS International Conference on Mathematics (MATH'06)* (2006), pp. 157–161
30. M. Demiralp, E. Demiralp, Reductive multilinear array decomposition based support functions in enhanced multivariance product representation (EMPR), in *Proceedings of the International Conference on Applied Computer Science (WSEAS)* (2010), pp. 448–454
31. E.K. Özay, A new multiway array decomposition via enhanced multivariance product representation, in *AIP Conference Proceedings*, vol. 1479 (2012), pp. 2015–2018
32. E. Demiralp, Application of reductive decomposition method for multilinear arrays (RDMMA) to animations, in *Proceedings of the 11th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering*, vol. 1, no. 2 (2009), pp. 648–656
33. L. Divanyan, M. Demiralp, Weighted reductive multilinear array decomposition, in *AIP Conference proceedings*, vol. 1389 (2011)

34. M. Demiralp, E. Demiralp, An orthonormal decomposition method for multidimensional matrices, numerical analysis and applied mathematics, in *International Conference on Numerical Analysis and Applied Mathematics*, vol. 1168, no. 1 (AIP Publishing, 2009), pp. 424–427
35. B.W. Bader, T.G. Kolda, MATLAB Tensor Toolbox Version 2.2. http://www.sandia.gov/~tgkolda/TensorToolbox/ (2007)
36. R. Bro, PARAFAC: tutorial and applications. Chemom. Intell. Lab. Syst. **38**, 149–171 (1997)
37. T. Skov, R. Bro, A new approach for modelling sensor based data. Sens. Actuators B Chem. **106**(2), 719–729 (2005)
38. R. Bro, Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. Chemom. Intell. Lab. Syst. **46**, 133–147 (1999)
39. http://www.models.kvl.dk/datasets