Overview of Recent Advances in Numerical Tensor Algebra

Göran Bergqvist[1] and Erik G. Larsson[2]

[1]Department of Mathematics, [2]Department of Electrical Engineering

Linköping University, SE-58183 Linköping, Sweden

[1]gober@mai.liu.se, [2]erik.larsson@isy.liu.se

Abstract—We present a survey of some recent developments for decompositions of multi-way arrays or tensors, with special emphasis on results relevant for applications and modeling in signal processing. A central problem is how to find low-rank approximations of tensors, and we describe some new results, including numerical methods, algorithms and theory, for the higher order singular value decomposition (HOSVD) and the parallel factors expansion or canonical decomposition (CP expansion).

I. INTRODUCTION

Tensors or multi-arrays are used to describe data structures with more than two dimensions. Direct studies of tensors can reveal patterns and properties that may be difficult to find using two-dimensional matrix methods. It can also produce algorithms which are more efficient than matrix algorithms. Important issues are compression and approximation of tensors. While the theory for matrices is simple and clear, the best low-rank approximation of a matrix is obtained by truncating its singular value decomposition (SVD), the theory for tensors is more complicated. In spite of this, tensor methods have shown to give improved models and algorithms for a number of applications. The higher order singular value decomposition (HOSVD) and the CP expansion (Canonical decomposition / Parallel factors) are two generalizations of the matrix SVD. We review the basic properties of these and discuss various numerical algorithms.

Tensors methods are being used in many scientific areas, such as chemometrics, psychometrics, image processing, bioinformatics, visualization, pattern recognition, data mining, and brain modeling [1], [18].

In communications and signal processing, tensor modeling and tensor decompositions have become increasingly important tools. This is the case for amplify and forward MIMO two-way relays [26], [27], [28]. The possibility to introduce simpler and cheaper relay stations into mobile communication systems to support the communication between mobile terminals is important. In amplify and forward two-way relaying, data signals are transmitted by both terminals over unknown channels to the relay, then the combined signal received at the

relay is amplified and transmitted again back to the terminals. The received signal at each terminal involves products of the data signals with both channel matrices and the amplification matrix. Note that all transmissions occur in the same spectrum and both communication ways use the same relay, hence we have spectrum and infrastructure sharing. The two-way communication is enabled by the use of multiple antennas at the relay, which gives the necessary degrees of freedom to design the amplification matrix so that each terminal can subtract its signal of interest out the combined one that the relay sends. In one-way communication with single-antenna relay, the amplification is just a scalar.

More precisely, suppose two terminals have n_i ($i = 1, 2$) antennas and that a relay station has n_R antennas. The transmitted vectors $\mathbf{x}^{(i)} \in \mathbb{C}^{n_i}$ are multiplied by channel $n_R \times n_i$ matrices $\mathbf{H}^{(i)}$ to produce the total received signal $\mathbf{r} = \mathbf{H}^{(1)}\mathbf{x}^{(1)} + \mathbf{H}^{(2)}\mathbf{x}^{(2)} + \mathbf{n} \in \mathbb{C}^{n_R}$ at the relay station, where $\mathbf{n}$ represents the noise. This is amplified by an $n_R \times n_R$ matrix $\mathbf{A}$ and transmitted to the terminals which receive $\mathbf{y}^{(i)} = \mathbf{H}^{(i)T}\mathbf{A}\mathbf{r} + \mathbf{n}^{(i)} = \mathbf{H}^{(i)T}\mathbf{A}(\mathbf{H}^{(1)}\mathbf{x}^{(1)} + \mathbf{H}^{(2)}\mathbf{x}^{(2)}) + \tilde{\mathbf{n}}^{(i)} \in \mathbb{C}^{n_i}$ (so each $\mathbf{y}^{(i)}$ depends on both $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$, and the vectors $\mathbf{n}^{(i)}$ and $\tilde{\mathbf{n}}^{(i)}$ represent noise). With $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$ known, it is a matrix problem to decode $\mathbf{x}^{(2)}$ at terminal 1 and vice versa, but in order to first determine the channel matrices $\mathbf{H}^{(i)}$, a training phase where known signals $\mathbf{x}^{(i,j)}$ ($i = 1, 2$, $j = 1, \ldots, N$) are transmitted for different fixed relay amplification matrices $\mathbf{A}^{(k)}$ ($k = 1, \ldots, n_R$), is needed. With the extra dimensions, represented by j and k, the received signals at the terminals form a 3-tensor $\mathcal{Y}$ (with components $\mathcal{Y}_{ijk}$), and it is a tensor decomposition problem to determine the channel matrices [26], [28].

Some other examples from signal processing are blind detection of CDMA spread spectrum signals [31], [32], blind identification [9], [10], [22], MIMO signal processing [7], [23], [36], multidimensional harmonic retrieval [16], and beamforming. For example, [31] considers a sensor array composed of several subarrays that receive a linear superposition of signals emitted by several sources. The received tensor has three dimensions: time, antenna index, and subarray index. Another example is blind multi-antenna receivers for code-division multiple-access (CDMA) systems [22]. Here, the tensor represents the received data along the dimensions antenna, chip, and symbol index.

Tensors models have also been used with success in image processing and face recognition. In [35], face image data was

modeled with 3-tensors with texels, illuminations and views as the three modes. Algorithms based on the HOSVD were shown to improve on matrix methods for recognition of test faces presented from new viewpoints and new illuminations. Another interesting example using the HOSVD is the automatic recognition of hand-written digits [29], [3].

We use the following notation, some of which was already introduced above for the amplify and forward two-way relay. Tensors or multi-arrays will be denoted by $\mathcal{T}$, $\mathcal{S}$ etc. and their elements $\mathcal{T}_{ijk...}$, matrices by $\mathbf{U}$, $\mathbf{V}$ etc. and their elements U_{ij}, vectors $\mathbf{u}$, $\mathbf{v}$ etc. and their components u_j. $\mathbf{U}^T$ denotes the transpose and $\mathbf{U}^H$ the Hermitian conjugate of $\mathbf{U}$, $\otimes$ is outer or tensor product, i.e., $\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T = \mathbf{A}$, where $A_{ij} = u_i v_j$, and $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} = \mathcal{T}$ where the 3-tensor $\mathcal{T}$ has components $\mathcal{T}_{ijk} = u_i v_j w_k$.

We consider only 3-tensors since all the important differences between matrices (2-tensors) and higher-order tensors appear already for 3-tensors.

II. RANK AND THE CP EXPANSION

An $m \times n \times p$ tensor $\mathcal{T}$ has elements $\mathcal{T}_{ijk}$ (which are real or complex numbers), with $1 \leq i \leq m$, $1 \leq j \leq n$ and $1 \leq k \leq p$. A tensor $\mathcal{S}$ that can be written $\mathcal{S} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$ for some vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ is said to be of *rank 1*. The *rank* r of a general tensor $\mathcal{T}$ is the minimum number of rank-1 tensors that need to be added to get $\mathcal{T}$. Then we have

$$\mathcal{T} = \sum_{l=1}^{r} \mathbf{x}^{(l)} \otimes \mathbf{y}^{(l)} \otimes \mathbf{z}^{(l)} \quad \text{or} \quad \mathcal{T}_{ijk} = \sum_{l=1}^{r} X_{il} Y_{jl} Z_{kl} \quad (1)$$

Here the vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)} \in \mathbb{C}^m$ are the columns of the $m \times r$ matrix $\mathbf{X}$, i.e., $x_i^{(l)} = X_{il}$, and analogously for $\mathbf{Y}$ and $\mathbf{Z}$.

The rank of a tensor is always well defined. Note that the definition of tensor rank is consistent with the definition of matrix rank as the number of terms in the singular value expansion. There are however some important differences. A rank-r matrix can be written

$$\mathbf{A} = \sum_{l=1}^{r} \mathbf{x}^{(l)} \otimes \mathbf{y}^{(l)} , \quad (2)$$

where the vectors $\mathbf{x}^{(l)}$ and $\mathbf{y}^{(l)}$ are highly non-unique. If they are required to be orthogonal, as in the SVD,

$$\mathbf{A} = \sum_{l=1}^{r} \sigma_l \mathbf{u}^{(l)} \otimes \mathbf{v}^{(l)} , \quad (3)$$

essential uniqueness is obtained. For tensors it is not possible in general to have orthogonality between the different $\mathbf{x}^{(l)}$. In fact, uniqueness in (1) up to trivial rescalings and permutations is guaranteed under mild conditions. A sufficient condition for uniqueness is $k_X + k_Y + k_Z \geq 2(r+1)$, where k_X (the Kruskal rank of $\mathbf{X}$) is the largest number such that any k_X columns of $\mathbf{X}$ are linearly independent [19], [20].

For an $m \times n$ matrix $\mathbf{A}$ we always have $r \leq \min(m,n)$ and generically (with probability 1 if the elements are drawn from a continuous probability distribution), $r = \min(m,n)$. Furthermore, if $\mathbf{A}$ is real, the rank is the same independently of whether $\mathbf{x}^{(l)}$ and $\mathbf{y}^{(l)}$ are allowed real or complex.

There are several important differences between matrix rank and tensor rank [18]. The determination of the rank of a tensor is a non-trivial problem and there is no finite algorithm for it.

For an $m \times n \times p$ tensor $\mathcal{T}$ of rank r it is possible that $r > \max\{m,n,p\}$. It is known that $r \leq \min(mn, mp, np)$. If the elements of $\mathcal{T}$ are chosen randomly from a continuous probability distribution (real or complex), then with probability 1, $\mathcal{T}$ will be of some particular rank over $\mathbb{C}$, the *generic* rank of $\mathcal{T}$. Still, there may be tensors with higher rank than the typical. As an example, the generic rank of $2 \times 2 \times 2$ tensors over $\mathbb{C}$ is 2, the maximal rank is 3, and ranks 0, 1 and 3 occur with probability 0. If $\mathcal{T}$ is real, there may not be any generic rank over $\mathbb{R}$. Then ranks which appear with positive probability are called *typical* ranks. If $\mathcal{T}$ is a real $2 \times 2 \times 2$ tensor with elements from a normal distribution of zero mean and standard deviation 1, then it has rank 2 with probability $\pi/4$ and rank 3 with probability $1 - \pi/4$ over $\mathbb{R}$ [2]. Ranks 2 and 3 are typical ranks, and 3 is the maximal rank. The only other known exact values of probabilities for typical ranks are that a $3 \times 3 \times 2$ real tensor has rank 3 with probability $1/2$ and rank 4 with probability $1/2$ [2]. In [4] the generic and typical ranks of $m \times n \times p$ tensors for some small values of m, n and p were recently determined. The maximal rank for general $m \times n \times p$ tensors is not known, but bounds exist.

The decomposition (1) of $\mathcal{T}$ is called the CANDECOMP (CANonical DECOMPosition) or PARAFAC (PARAllel FACtors expansion), or just the *CP expansion*.

Generalizing the Frobenius norm

$$||\mathbf{A}||^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2 \quad (4)$$

for matrices, we define a norm of $\mathcal{T}$

$$||\mathcal{T}||^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} |\mathcal{T}_{ijk}|^2 \quad (5)$$

A fundamental problem is to find the best rank-s approximation $\hat{\mathcal{T}}$ of $\mathcal{T}$, where $s < r$, i.e., to find

$$\hat{\mathcal{T}} = \sum_{l=1}^{s} \mathbf{a}^{(l)} \otimes \mathbf{b}^{(l)} \otimes \mathbf{c}^{(l)} \quad \text{or} \quad \hat{\mathcal{T}}_{ijk} = \sum_{l=1}^{s} A_{il} B_{jl} C_{kl} \quad (6)$$

such that $||\mathcal{T} - \hat{\mathcal{T}}||$ is minimized over all rank-s tensors. The first problem is that the truncated CP expansion of $\mathcal{T}$ does not provide the solution, i.e., we can not take $\mathbf{a}^{(l)} = \mathbf{x}^{(l)}$, $\mathbf{b}^{(l)} = \mathbf{y}^{(l)}$, and $\mathbf{c}^{(l)} = \mathbf{z}^{(l)}$, for $1 \leq l \leq s$. Thus we have a more complicated situation than for matrices where the truncated SVD is the best low-rank approximation. In fact, given a tensor $\mathcal{T}$, there is in general no analytic method to find its full CP expansion (1).

Even more serious is that the low-rank approximation problem is in general ill-posed, the set of tensors of a fixed rank is not closed [14]. Thus, no rank-s $\hat{\mathcal{T}}$ that minimizes $||\mathcal{T} - \hat{\mathcal{T}}||$ need to exist. A sequence of rank-s tensors may converge to a tensor of higher rank, not only of rank $s+1$ but even higher.

A best rank-1 approximation does exist, but it might not be one of the two terms in the best rank-2 approximation (if that

exists). Furthermore, subtracting the best rank-1 approximation from $\mathcal{T}$ may result in a tensor of higher rank than $\mathcal{T}$ [33].

If the tensor is non-negative (all elements non-negative), and all vectors in (1) required to be non-negative, then the low-rank approximation problem is well-posed [21]. Another special type of tensor are the symmetric tensors which were studied in [5].

In spite of these problems, low-rank approximations of tensors are applied in many scientific areas and the reason is that such methods can outperform matrix methods for the same problems. Some algorithms for finding low-rank approximations will be described below.

III. MULTI-RANK AND THE HOSVD

Besides the above definition of rank for tensors, there is also the so-called the multi-rank of $\mathcal{T}$. The mode-1 rank or column rank r_1 of $\mathcal{T}$ is the dimension of the space spanned by all mode-1 (column) vectors of $\mathcal{T}$, i.e. the np vectors $\mathbf{t}_{\cdot jk} \in \mathbb{C}^m$, where $(\mathbf{t}_{\cdot jk})_i = \mathcal{T}_{ijk}$, $i = 1, ..., m$, are the components of $\mathbf{t}_{\cdot jk}$. Similarly one defines the mode-2 (row) and mode-3 ranks r_2 and r_3. For a matrix it is well known that the column and row ranks are equal, and that they are equal to the number r of non-zero singular values of the matrix. Thus there is simple rank concept for matrices. For tensors, however, r_1, r_2 and r_3 can all be different and one therefore defines the *multi-rank* (r_1, r_2, r_3) of $\mathcal{T}$.

One can then formulate a new approximation problem for $\mathcal{T}$. Given a triple (s_1, s_2, s_3), with $s_1 \leq r_1$, $s_2 \leq r_2$ and $s_3 \leq r_3$, find the tensor $\hat{\mathcal{T}}$ of multi-rank (s_1, s_2, s_3) that minimizes $||\mathcal{T} - \hat{\mathcal{T}}||$. This problem has been shown to well-posed, the best low multi-rank approximation always exists [14].

The *Tucker decomposition* of an $m \times n \times p$ tensor $\mathcal{T}$ is

$$\mathcal{T} = \sum_{I=1}^{M} \sum_{J=1}^{N} \sum_{K=1}^{P} \mathcal{G}_{IJK}\, \mathbf{u}^{(I)} \otimes \mathbf{v}^{(J)} \otimes \mathbf{w}^{(K)} \qquad (7)$$

or, for the components,

$$\mathcal{T}_{ijk} = \sum_{I=1}^{M} \sum_{J=1}^{N} \sum_{K=1}^{P} \mathcal{G}_{IJK} U_{iI} V_{jJ} W_{kK} \qquad (8)$$

Here the vectors $\mathbf{u}^{(I)}$ are columns of the $m \times M$ matrix $\mathbf{U}$ so $u_i^{(I)} = U_{iI}$, and in the same way $v_j^{(J)} = V_{jJ}$ and $w_k^{(K)} = W_{kK}$. The scalars $\mathcal{G}_{IJK}$ are elements of an $M \times N \times P$ core tensor $\mathcal{G}$. If $M \leq m$, $N \leq n$ and $P \leq p$, then $\mathcal{G}$ is a compression of $\mathcal{T}$.

In [11] it was shown that $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ can be taken to be unitary (orthogonal if $\mathcal{T}$ is real) matrices, so that $\mathcal{G}$ has the same size as $\mathcal{T}$. Simultaneously, the matrix slices of $\mathcal{G}$ along any mode can be chosen to be mutually orthogonal (with respect to the standard inner product $(\mathbf{A}, \mathbf{B}) = \text{tr}(\mathbf{A}^H \mathbf{B})$ for matrices), and with decreasing Frobenius norm. Slices along the first mode are matrices $\mathbf{G}^{(i)1}$ defined by the elements $G_{jk}^{(i)1} = \mathcal{G}_{1jk}$. Hence we have $\text{tr}((\mathbf{G}^{(i)1})^H \mathbf{G}^{(q)1}) = 0$ if $i \neq q$ and $||\mathbf{G}^{(i)1}|| \geq ||\mathbf{G}^{(q)1}||$ if $i > q$. The same holds for matrix slices along the other two modes. This is therefore a generalization of the matrix SVD, in which rows and columns of the singular value matrix $\mathbf{\Sigma}$ are mutually orthogonal and

with decreasing norm. The Tucker decomposition is in this case called the *Higher Order Singular Value Decomposition* (HOSVD) [11], [3], [18]. Because of the orthogonality conditions, the HOSVD is essentially unique. The HOSVD is rank revealing which means that if $\mathcal{T}$ has multirank (r_1, r_2, r_3), then the last $m - r_1$, $n - r_2$ and $p - r_3$ slices along the different modes in $\mathcal{G}$ are zero matrices. One may then use thin matrices $\mathbf{U}$ $(m \times r_1)$, $\mathbf{V}$ $(n \times r_2)$ and $\mathbf{W}$ $(p \times r_3)$ and a smaller $(r_1 \times r_2 \times r_3)$ core tensor to write the expansion.

In HOSVD, $\mathbf{U}$ can be calculated by performing a matrix SVD on the $(m \times np)$ matrix obtained by a flattening or matrization of $\mathcal{T}$. $\mathbf{V}$ and $\mathbf{W}$ are found in the same way [11]. Since $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are unitary, $\mathcal{G}$ is then easily calculated from

$$\mathcal{G}_{IJK} = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \mathcal{T}_{ijk} \bar{U}_{iI} \bar{V}_{jJ} \bar{W}_{kK} \qquad (9)$$

Hence, as opposed to the situation of the CP expansion, there is a straightforward method for calculating the HOSVD of a tensor.

The truncated HOSVD is in general not the best low multi-rank approximation of $\mathcal{T}$ but it is still used in many applications as an approximation of $\mathcal{T}$.

IV. ALGORITHMS FOR LOW-RANK APPROXIMATION

The traditional method for finding a low-rank approximation $\hat{\mathcal{T}}$ of $\mathcal{T}$ is to use alternating least squares (ALS). Fixing the rank s of the approximation, the goal is to minimize

$$||\mathcal{T} - \hat{\mathcal{T}}||^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} |\mathcal{T}_{ijk} - \sum_{l=1}^{s} A_{il} B_{jl} C_{kl}|^2 \qquad (10)$$

Thus, we are searching for the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of sizes $m \times s$, $n \times s$, and $p \times s$ respectively. If two of the matrices are known, then it is a least square problem to find the elements of the third. Based on this observation, the alternating least squares method is as follows. Fix some $\mathbf{B}^{(\kappa)}$ and $\mathbf{C}^{(\kappa)}$ and then define $\mathbf{A}^{(\kappa+1)}$ to be the least square solution to

$$\mathcal{T}_{ijk} = \sum_{l=1}^{s} A_{il} B_{jl}^{(\kappa)} C_{kl}^{(\kappa)}, \quad 1 \leq i \leq m, 1 \leq j \leq n, 1 \leq k \leq p$$

(mnp equations, ms unknowns). In the next step fix $\mathbf{A}^{(\kappa+1)}$ and $\mathbf{C}^{(\kappa)}$, and define $\mathbf{B}^{(\kappa+1)}$ to be the least square solution to the corresponding problem over $\mathbf{B}$. Then use $\mathbf{A}^{(\kappa+1)}$ and $\mathbf{B}^{(\kappa+1)}$ to find $\mathbf{C}^{(\kappa+1)}$ in the same way, and $\mathbf{A}^{(\kappa+2)}$ will be determined from $\mathbf{B}^{(\kappa+1)}$ and $\mathbf{C}^{(\kappa+1)}$ and so on. This method has monotone decrease of $||\mathcal{T} - \hat{\mathcal{T}}||$, but the convergence rate is slow and it may converge to a local minimum of $||\mathcal{T} - \hat{\mathcal{T}}||$ or not converge at all [6].

Alternating least squares are also used to calculate the rank and the full CP expansion of $\mathcal{T}$. One may start by guessing the rank, run the ALS and if the obtained $\hat{\mathcal{T}}$ is sufficiently close to $\mathcal{T}$, the rank and the CP expansion have been found. Otherwise increase the guess of the rank by 1 and repeat the procedure.

Various improvements of the method have been suggested. To avoid ending up in local minima the algorithm may be executed with many different initial values of the matrices $\mathbf{B}^{(0)}$

and $\mathbf{C}^{(0)}$. One may also use enhanced line search. Represent $\mathbf{A}^{(\kappa+1)}$, $\mathbf{B}^{(\kappa+1)}$ and $\mathbf{C}^{(\kappa+1)}$ by q_{k+1}, and $\mathbf{A}^{(\kappa)}$, $\mathbf{B}^{(\kappa)}$ and $\mathbf{C}^{(\kappa)}$ by q_k. Then $q_{k+1} - q_k = \Delta_k$, and line search means looking for the optimal value of μ in $q_k + \mu\Delta_k$ as the next point rather than just using $\mu = 1$ to get q_{k+1} as the next point. See [34], [6] and [18] for descriptions of improved ALS and other methods, including conjugate gradient, Newton, and higher order methods.

An interesting approach is to use simultaneous matrix diagonalization [8], [13], [24], [25]. Following [24], first write

$$\mathcal{T} \approx \mathcal{T}^s \tag{11}$$

where $\mathcal{T}^s$ is the multi-rank (s,s,s) truncation of the HOSVD of $\mathcal{T}$. The core tensor corresponding to $\mathcal{T}^s$ is denoted $\mathcal{G}^s$ and we still use the notation $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ for the HOSVD matrices. In [24] it is shown how there are non-singular $s \times s$ matrices $\mathbf{T_1}$, $\mathbf{T_2}$ and $\mathbf{T_3}$ such that $\mathbf{A} = \mathbf{U}\mathbf{T_1}$, $\mathbf{B} = \mathbf{V}\mathbf{T_2}$ and $\mathbf{C} = \mathbf{W}\mathbf{T_3}$. The matrices $\mathbf{T_1}$ and $\mathbf{T_2}$ simultaneously diagonalize a certain set of matrices which are slices of the tensor $\mathcal{H}_3^s$ given by

$$(\mathcal{H}_3^s)_{ijk} = \sum_l \mathcal{G}_{ijl}^s W_{kl} \tag{12}$$

multiplied by inverses of such slices, or similarly along the other modes. $\mathbf{T_1}$ will diagonalize $n + p$ such matrices giving estimates firstly for $\mathbf{C}$ and $\mathbf{T_1}$, from which an estimate for $\mathbf{A}$ is also obtained. Repeating this along all modes for all $\mathbf{T_i}$ one gets several estimates for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, and finally chooses among these the ones that minimize (10). Numerical performance compares well with improved ALS methods with enhanced line search.

V. Algorithms for low multi-rank approximation

The HOSVD can be calculated using the SVD for matrices as explained above, and in many applications truncation of the HOSVD has been applied with success. This is in spite of the fact that in general the $s_1 \times s_2 \times s_3$ truncation of the HOSVD is not the best multirank-(s_1, s_2, s_3) approximation of $\mathcal{T}$. However, since the matrix slices along all modes have decreasing norm, it is expected that truncation gives a fairly good low multi-rank approximation, and for many applications it has been considered to be sufficiently good. It can also serve as an initial value in other algorithms for finding the best approximation, and as mentioned above it is used as a first step when searching for low-rank approximations using simultaneous matrix diagonalization.

One example where truncation of HOSVD was used is that of automatically recognizing a handwritten digit [29]. A $256 \times 1194 \times 10$ 3-tensor $\mathcal{T}_{ijk}$ of data from the training set is by HOSVD truncation reduced to an $m \times n \times 10$ 3-tensor. The HOSVD of $\mathcal{T}_{ijk}$ is

$$\mathcal{T}_{ijk} = \sum_{I=1}^{256} \sum_{J=1}^{1194} \sum_{K=0}^{9} \mathcal{G}_{IJK} U_{iI} V_{jJ} W_{kK}$$

$$\approx \sum_{I=1}^{m} \sum_{J=1}^{n} \sum_{K=0}^{9} \mathcal{G}_{IJK} U_{iI} V_{jJ} W_{kK} = \sum_{I=1}^{m} \sum_{J=1}^{n} \mathcal{F}_{IJk} U_{iI} V_{jJ} ,$$

where

$$\mathcal{F}_{IJk} = \sum_{K=0}^{9} \mathcal{G}_{IJK} W_{kK} = \sum_{i=1}^{256} \sum_{j=1}^{1194} \mathcal{T}_{ijk} U_{iI} V_{jJ}$$

is an $m \times n \times 10$ tensor. By choosing m and n between 32 and 64 the data is compressed by about 99%, and only the first m and n columns of $\mathbf{U}$ and $\mathbf{V}$, respectively, need to be calculated. An unknown digit $\mathbf{z} \in \mathbb{R}^{256}$ is projected to $(\mathbf{U}^T \mathbf{z})_i = \sum_{j=1}^{256} U_{ji} z_j \in \mathbb{R}^m$. For each $d = 0, \ldots, 9$ least-squares are used to approximate $\mathbf{U}^T z$ by the columns of the $m \times n$ matrix $(F_d)_{ij} = \mathcal{F}_{ijd}$, which by a matrix SVD is further reduced to an $m \times k$ matrix with $k \approx 10$. The value of d with the smallest residual determines the classification of the digit. With the tensor model all digits are projected to a single common subspace. Only one projection of a test digit $\mathbf{z}$ is needed rather than 10, one for each d, needed if the training data is modeled as 10 matrices. Tests show that the algorithm is computationally efficient compared to other methods, and has an error rate of only 5% when the data is compressed by 99%.

To calculate a truncated HOSVD, only the desired number of columns of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ in (9) need to be calculated, which can be done using higher order orthogonal iteration (HOOI) [12].

For the problem of finding the best multirank-(s_1, s_2, s_3) approximation of $\mathbf{T}$, alternating least-squares has been the traditional method but very recently improved methods have been developed [18].

In general, to find the best multi-rank (s_1, s_2, s_3) approximation

$$\hat{\mathcal{T}} = \sum_{I=1}^{s_1} \sum_{J=1}^{s_2} \sum_{K=1}^{s_3} \hat{\mathcal{G}}_{IJK} \, \mathbf{a}^{(I)} \otimes \mathbf{b}^{(J)} \otimes \mathbf{c}^{(K)} \tag{13}$$

of $\mathcal{T}$, one has to minimize

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} |\mathcal{T}_{ijk} - \sum_{I=1}^{s_1} \sum_{J=1}^{s_2} \sum_{K=1}^{s_3} \hat{\mathcal{G}}_{IJK} A_{iI} B_{jJ} C_{kK}|^2 \tag{14}$$

over all $s_1 \times s_2 \times s_3$ tensors $\hat{\mathcal{G}}$, $m \times s_1$ matrices $\mathbf{A}$, $n \times s_2$ matrices $\mathbf{B}$, and $p \times s_3$ matrices $\mathbf{C}$. The matrices may be assumed to satisfy $\mathbf{A}^H \mathbf{A} = \mathbf{I}_{s_1}$, $\mathbf{B}^H \mathbf{B} = \mathbf{I}_{s_2}$ and $\mathbf{C}^H \mathbf{C} = \mathbf{I}_{s_3}$ [15]. The minimizing problem (15) can be shown to equivalent to maximizing

$$\sum_{I=1}^{s_1} \sum_{J=1}^{s_2} \sum_{K=1}^{s_3} |\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \mathcal{T}_{ijk} A_{iI} B_{jJ} C_{kK}|^2 \tag{15}$$

still assuming $\mathbf{A}^H \mathbf{A} = \mathbf{I}_{s_1}$, $\mathbf{B}^H \mathbf{B} = \mathbf{I}_{s_2}$ and $\mathbf{C}^H \mathbf{C} = \mathbf{I}_{s_3}$. Note that (15) is invariant under multiplication by unitary matrices on the right of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Such equivalence classes are points on Grassmann manifolds so it implies that one actually has a maximizing problem on the product of three Grassmann manifolds [15], [30], [17].

Newton's method for maximizing a function $f(x)$, $x_{k+1} = x_k + p_k$ with p_k satisfying $H_f(x_k) p_k = -\nabla f(x_k)$ (H_f the Hessian of f), or with line search $x_{k+1} = x_k + \mu p_k$, is in [15], [30] adapted to the manifold structure. This means that covariant derivatives must be used and in the iteration

one moves along geodesic curves. The algorithms show much improved performance compared to other methods.

REFERENCES

[1] E Acar and B Yener "Unsupervised multiway data analysis: a literature survey" *IEEE Trans. Knowl. Data Eng.* **21** 6–20, 2009

[2] G Bergqvist "Exact probabilities for typical ranks of 2x2x2 and 3x3x2 tensors" preprint, 2010

[3] G Bergqvist and E G Larsson "The higher order singular value decomposition: theory and an application" *IEEE Signal Process. Mag.* **27** 151–154, 2010

[4] P Comon, J M F ten Berge, L De Lathauwer and J Castaing "Generic and typical ranks of multi-way arrays" *Lin. Alg. Appl.* **430** 2997–3007, 2009

[5] P Comon, G Golub, L-H Lim and B Mourrain "Symmetric tensors and symmetric tensor rank" *SIAM J. Matrix Anal. Appl.* **30** 1254–1279, 2008

[6] P Comon, X Luciani, A L F de Almeida "Tensor decomposition, alternating least squares and other tales" *J. Chemometrics* **23** 393–405, 2009

[7] A L F de Almeida, G Favier and J C M Mota "A constrained factor decomposition with application to MIMO antenna systems" *IEEE Trans. Signal Process.* **56** 2429–2442, 2008

[8] L De Lathauwer "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization singular value decomposition" *SIAM J. Matrix Anal. Appl.* **28** 642–666, 2006

[9] L De Lathauwer and A de Baynast "Blind deconvolution of DS-CDMA signals by means of decomposition in rank-(1,L,L) terms" *IEEE Trans. Signal Process.* **56** 1562–1571, 2008

[10] L De Lathauwer and J Castaing "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization" *IEEE Trans. Signal Process.* **56** 1096–1105, 2008

[11] L De Lathauwer, B De Moor and J Vandewalle "A multilinear singular value decomposition" *SIAM J. Matrix Anal. Appl.* **21** 1253–1278, 2000

[12] L De Lathauwer, B De Moor and J Vandewalle "On the best rank-1 and rank-$(R_1, R_2, \ldots R_N)$ approximation of higher-order tensors" *SIAM J. Matrix Anal. Appl.* **21** 1324–1342, 2000

[13] L De Lathauwer, B De Moor and J Vandewalle "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition" *SIAM J. Matrix Anal. Appl.* **26** 295–327, 2004

[14] V De Silva and L-H Lim "Tensor rank and the ill-posedness of the best low-rank approximation problem" *SIAM J. Matrix Anal. Appl.* **30** 1084–1127, 2008

[15] L Eldén and B Savas "A Newton-Grassmann method for computing the best multi-linear rank-(r_1, r_2, r_3) approximation of a tensor" *SIAM J. Matrix Anal. Appl.* **31** 248–271, 2009

[16] M Haardt, F Roemer, and G Del Galdo "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems" *IEEE Trans. Signal Process.* **56** 3198–3213, 2008

[17] M Ishteva, P-A Absil, S van Huffel and L De Lathauwer " Best low rank multilinear rank approximation of higher-order tensors, basede on the Riemannian trust-region scheme" preprint, 2009

[18] T G Kolda and B W Bader "Tensor decompositions and applications" *SIAM Review* **3** 455–500, 2009

[19] J B Kruskal "Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics" *Lin. Alg. Appl.* **18** 95–138, 1977

[20] J B Kruskal "Rank, decomposition, and uniqueness for 3-way and N-way arrays" in *Multiway data analysis* (Eds. R Coppi and S Bolasco) 7–18, North-Holland, 1989

[21] L-H Lim and P Comon "Nonnegative approximations of nonnegative tensors" *J. Chemometrics* **23** 432–441, 2009

[22] D Nion and L De Lathauwer "A block component model-based blind DS-CDMA receiver" *IEEE Trans. Signal Process.* **56** 5567–5579, 2008

[23] D Nion and N D Sidiropoulos "Adaptive algorithms to track the PARAFAC decomposition of a third-order tensor" *IEEE Trans. Signal Process.* **57** 2299–2310, 2009

[24] F Roemer and M Haardt "A closed-form solution for parallel factor (PARAFAC) analysis" in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 2365–2368, 2008

[25] F Roemer and M Haardt "A closed-form solution for multilinear PARAFAC decompositions" in *Proc. 5-th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2008)* 487–491, 2008

[26] F Roemer and M Haardt "Tensor-based channel estimation (TENCE) for two-way relaying with multiple antennas and spatial reuse" in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)* 3641–3644, 2009

[27] F Roemer and M Haardt "Algebraic norm-maximizing (ANOMAX) transmit strategy for two-way relaying with MIMO amplify and relay orders" *IEEE Signal Process. Letters* **16** 909–912, 2009

[28] F Roemer and M Haardt "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse" *IEEE Trans. Signal Process.* **58** 5720–5735, 2010

[29] B Savas and L Eldén "Handwritten digit classification using higher order singular value decomposition" *Pattern Recogn.* **40** 993–1003, 2007

[30] B Savas and L-H Lim "Best multilinear rank approximation of tensors with quasi-Newton methods on Grassmannians" *SIAM J. Scientific Comput.* to appear

[31] N D Sidiropoulos, R Bro and G B Giannikis "Parallel factor analysis in sensor array processing" *IEEE Trans. Signal Process.* **48** 2377–2388, 2000

[32] N D Sidiropoulos, G B Giannikis and R Bro "Blind PARAFAC receivers for DS-CDMA systems" *IEEE Trans. Signal Process.* **48** 810–823, 2000

[33] A Stegeman and P Comon "Subtracting a best rank-1 approximation may increase tensor rank" *Lin. Alg. Appl.* **433** 1276–1300, 2010

[34] G Tomasi and R Bro "A comparison of algorithms for fitting the PARAFAC model" *Comput. Statist. Data Anal.* **50** 1700–1734, 2006

[35] M Vasilescu and D Terzopoulos "Multilinear (tensor) image synthesis, analysis and recognition" *IEEE Signal Process. Mag.* **6** 118–123, 2007

[36] Y Yu and A P Petropulu "PARAFAC-based blind estimation of possibly underdetermined convolutive MIMO systems" *IEEE Trans. Signal Process.* **56** 111–124, 2008