arXiv:0805.4220v1 [math.NA] 27 May 2008

Best subspace tensor approximations

S. Friedland* V. Mehrmann†

27 May, 2008

Abstract

In many applications such as data compression, imaging or genomic data analysis, it is important to approximate a given tensor by a tensor that is sparsely representable. For matrices, i.e. 2-tensors, such a representation can be obtained via the singular value decomposition which allows to compute the best rank k approximations. For t-tensors with $t > 2$ many generalizations of the singular value decomposition have been proposed to obtain low tensor rank decompositions. In this paper we will present a different approach which is based on best subspace approximations, which present an alternative generalization of the singular value decomposition to tensors.

2000 Mathematics Subject Classification. 15A18, 15A69, 65D15, 65H10, 65K05

Key words. singular value decomposition, rank k approximation, least squares, tensor rank, best subspace tensor approximation.

1 Introduction

In this paper we will consider data sparse approximations of tensors. We will discuss a generalization of the singular value decomposition from matrices to tensors that is an alternative to the Tucker decomposition [8, 10]. In order not to overload the paper with technical we will mainly discuss 3-tensors, but our approach will work for arbitrary tensors.

Let $\mathbb{F}$ be either the field of real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$. Denote by $\mathbb{F}^{m_1\times\ldots\times m_d} := \otimes_{i=1}^{d}\mathbb{F}^{m_j}$ the tensor products of $\mathbb{F}^{m_1},\ldots,\mathbb{F}^{m_d}$. $\mathcal{T} = [t_{i_1,\ldots,i_d}] \in \mathbb{F}^{m_1\times\ldots\times m_d}$ is called a *d-tensor* in the given tensor product. Note that the number of coordinates of $\mathcal{T}$ is $N = m_1 \ldots m_d$. A tensor $\mathcal{T}$ is called a *sparsely representable tensor* if it can represented with a number of coordinates that is much smaller than N.

The best known example of a sparsely representable 2-tensor is a low rank approximation of a matrix $A \in \mathbb{F}^{m_1\times m_2}$. A rank k approximation of A is given by $A_{\mathrm{appr}} := \sum_{i=1}^{k} \mathbf{u}_i\mathbf{v}_i^{\top}$, which can be identified with $\sum_{i=1}^{k} \mathbf{u}_i \otimes \mathbf{v}_i$. To store A_{appr} we

*Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607-7045, USA, `friedlan@uic.edu`, Partially supported by *Berlin Mathematical School*, Berlin, Germany.

†Institut für Mathematik, TU Berlin, Str. des 17. Juni 136, D-10623 Berlin, FRG. `mehrmann@math.tu-berlin.de`. Partially supported by *Deutsche Forschungsgemeinschaft*, through the DFG Research Center Matheon *Mathematics for Key Technologies* in Berlin

need only the $2k$ vectors $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{F}^{m_1}$, $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{F}^{m_2}$. The best rank k approximation of $A \in \mathbb{F}^{m_1 \times m_2}$ can be computed via the *singular value decomposition*, abbreviated here as SVD, [4].

The computation of the SVD requires $\mathcal{O}(m_1 m_2^2 + m_2^2)$ operations and at least $\mathcal{O}(m_1 m_2)$ storage. Thus, if the dimensions m_1 and m_2 are very large, then the computation of the SVD is often infeasible. In this case other type of low rank approximations are considered, see e.g. [2, 3, 5].

For d-tensors with $d > 2$, however the situation is rather unsatisfactory. It is a major theoretical and computational problem to formulate good generalizations of low rank approximation for tensors and to give efficient algorithms to compute these approximations, see e.g. [8, 9, 10]. It is the goal of this paper to present and analyze an alternative generalization of the SVD to tensors.

A tensor $\mathcal{T} = [t_{i,j,k}] \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ is called a *rank* 1 tensor, and denoted by $\mathcal{T} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$, if $t_{i,j,k} = u_i v_j w_k$, where $\mathbf{u} = (u_1, \dots, u_{m_1})^\top, \mathbf{v} = (v_1, \dots, v_{m_2})^\top, \mathbf{w} = (w_1, \dots, w_{m_3})^\top$. A tensor $\mathcal{T} \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ is said to have *rank* k if $\mathcal{T}$ can be represented as a sum of k rank 1 tensors, and cannot be represented as a sum of $k-1$ rank 1 tensors. Note that if $\mathcal{T}$ is a sum of k rank 1 tensors, then $\mathcal{T}$ can be represented with at most $\mathcal{O}(k(\ell + m + n))$ storage.

We denote by $\mathcal{R}(k; m_1, m_2, m_3)$ the set of tensors in $\mathbb{F}^{m_1 \times m_2 \times m_3}$ of rank k at most. It is easy to show that $\mathcal{R}(1; m_1, m_2, m_3)$ is a closed set, more precisely an algebraic variety, in $\mathbb{F}^{m_1 \times m_2 \times m_3}$. However, it is well known, see e.g. [1], that for some values of $k \geq 2$, $\mathcal{R}(k; m_1, m_2, m_3)$ is not a closed set. ($\mathcal{R}(k; m_1, m_2, m_3)$ *is called a quasi-algebraic variety.*)

Let $\|\cdot\|$ be a norm on $\mathbb{F}^{m_1 \times m_2 \times m_3}$. Then for $k \geq 2$ it is possible that the minimization problem

$$\min_{\mathcal{X} \in \mathcal{R}(k; m_1, m_2, m_3)} \|\mathcal{T} - \mathcal{X}\| \tag{1.1}$$

does not have a minimal solution. This will happen if $\mathcal{T}$ has rank greater than k and $\mathcal{T}$ lies in the closure of $\mathcal{R}(k; m_1, m_2, m_3)$. Hence, any algorithm which tries to find a solution to the minimization problem (1.1) will fail for certain tensors $\mathcal{T}$. Since $\mathcal{R}(k; m_1, m_2, m_3)$ is a closed set, for $k = 1$, i.e. for the best approximation by a rank 1 tensor, (1.1) will always have a minimal solution.

The object of this paper to introduce a new family of sparsely representable approximations to tensors, which we call *best subspace tensor approximation (BSTA)* of a given tensor $\mathcal{T}$. As for the best rank 1 approximation, we will show that the BSTA always exists. Due to this fact, we think that in the case that the norm $\|\cdot\|$ on $\mathbb{F}^{m_1 \times m_2 \times m_3}$ is the norm induced by the inner products on the vector spaces $\mathbb{F}^{m_1}$, $\mathbb{F}^{m_2}$, $\mathbb{F}^{m_3}$, the BSTA is an appropriate generalization of the SVD, see [8] for other generalizations of the SVD for tensors. Similar approach was suggested recently by Khoromskij [7]. We will also present a numerical algorithm to compute the best subspace tensor approximation that is based on the computation of singular value decompositions for matrices.

Unfortunately this numerical algorithm is extremely expensive. In order to reduce the complexity, in the last section we consider a procedure that is based on the recently suggested fast SVD [3].

2 Notation and preliminary results

We denote by a bold capital letter a finite dimensional vector space $\mathbf{U}$ over the field $\mathbb{F}$. A vector $\mathbf{u} \in \mathbf{U}$ is denoted by a bold face lower case letter. A matrix $A \in \mathbb{F}^{m_1 \times m_2}$ denoted by a capital letter A, and we let either $A = [a_{i,j}]_{i=j=1}^{m_1 \times m_2}$ or simply $A = [a_{i,j}]$. A *3-tensor array* $\mathcal{T} \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ will be denoted by a capital calligraphic letter. So either $\mathcal{T} = [t_{i,j,k}]_{i=j=k=1}^{m_1,m_2,m_3}$ or simply $\mathcal{T} = [t_{i,j,k}]$. For a positive integer n we also use the convenient notation $\langle n \rangle := \{1, 2, \ldots, n\}$.

Let $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$ be three vectors spaces over $\mathbb{F}$ with $m_j := \dim \mathbf{U}_j$, $j = 1, 2, 3$ and let $\mathbf{u}_{1,j}, \ldots, \mathbf{u}_{m_j,j}$ be a basis of $\mathbf{U}_j$ for $j = 1, 2, 3$. Then $\mathbf{U} := \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3$ is the *tensor product* of $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_3$; $\mathbf{U}$ is a vector space of dimension $m_1m_2m_3$, and

$$\mathbf{u}_{i_1,1} \otimes \mathbf{u}_{i_2,2} \otimes \mathbf{u}_{i_3,3}, \quad i_j = 1, \ldots, m_j,\ j = 1, 2, 3, \tag{2.1}$$

is a basis of $\mathbf{U}$.

A 3-*tensor* τ is a vector in $\mathbf{U}$ and it has a representation

$$\tau = \sum_{i_1=i_2=i_3=1}^{m_1,m_2,m_3} t_{i_1,i_2,i_3} \mathbf{u}_{i_1,1} \otimes \mathbf{u}_{i_2,2} \otimes \mathbf{u}_{i_3,3}, \tag{2.2}$$

in the basis (2.1). If the basis (2.1) is fixed then τ is identified with $\mathcal{T} = [t_{i_1,i_2,i_3}] \in \mathbb{F}^{m_1 \times m_2 \times m_3}$.

Recall that $\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$, were $\mathbf{x}_i \in \mathbf{U}_i$, $i = 1, 2, 3$, is called a *rank* 1 tensor.(Usually one assumes that all $\mathbf{x}_i \neq \mathbf{0}$. Otherwise $\mathbf{0} = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$ is called a rank 0 tensor.) Then (2.2) is a decomposition of τ as a sum of at most $m_1m_2m_3$ rank 1 tensors, as $t_{i_1,i_2,i_3}\mathbf{u}_{i_1,1} \otimes \mathbf{u}_{i_2,2} \otimes \mathbf{u}_{i_3,3} = (t_{i_1,i_2,i_3}\mathbf{u}_{i_1,1}) \otimes \mathbf{u}_{i_2,2} \otimes \mathbf{u}_{i_3,3}$. A decomposition of $\tau \in \mathbf{U} \backslash \{\mathbf{0}\}$ as a sum of rank 1 tensors is given by

$$\tau = \sum_{i=1}^{k} \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i, \quad \mathbf{x}_i \in \mathbf{U}_1,\ \mathbf{y}_i \in \mathbf{U}_2,\ \mathbf{z}_i \in \mathbf{U}_3,\ i = 1, \ldots, k. \tag{2.3}$$

The minimal k for which the above equality holds is called the *rank* of the tensor τ. This definition is completely analogous to the definition of the rank for a matrix $A = [a_{i_1,i_2}] \in \mathbb{F}^{m_1 \times m_2}$, which can be identified with 2-tensor in $\sum_{i_1=i_2=1}^{m_1,m_2} a_{i_1,i_2}\mathbf{u}_{i_1,1} \otimes \mathbf{u}_{i_2,2} \in \mathbf{U}_1 \otimes \mathbf{U}_2$.

For $j \in \{1, 2, 3\}$ denote by $j^c := \{p, q\} = \{1, 2, 3\} \backslash \{j\}$, where $1 \le p < q \le 3$, and set $\mathbf{U}_{j^c} = \mathbf{U}_{\{p,q\}} := \mathbf{U}_p \otimes \mathbf{U}_q$.

A tensor $\tau \in \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3$ induces a linear transformation $\tau(j) : \mathbf{U}_{j^c} \to \mathbf{U}_j$ as follows. Suppose that $\mathbf{u}_{1,\ell}, \ldots, \mathbf{u}_{m_\ell,\ell}$ is a basis in $\mathbf{U}_\ell$ for $\ell = 1, 2, 3$. Then any $\mathbf{v} \in \mathbf{U}_{j^c}$ is of the form

$$\mathbf{v} = \sum_{i_p=i_q=1}^{m_p,m_q} v_{i_p,i_q} \mathbf{u}_{i_p,p} \otimes \mathbf{u}_{i_q,q}$$

and the application of $\tau(j)$ is given by

$$\tau(j)\ \mathbf{v} = \sum_{i_j=1}^{m_j} \Big(\sum_{i_p,i_q=1}^{m_p,m_q} t_{i_1,i_2,i_3} v_{i_p,i_q} \Big) \mathbf{u}_{i_j,j}. \tag{2.4}$$

Then $\mathrm{rank}_j(\tau)$ *is the rank of the operator* $\tau(j)$. Equivalently, let $A(j) = [a_{\ell,i_j}] \in \mathbb{R}^{m_p m_q \times m_j}$, where each integer $\ell \in \langle m_p m_q \rangle$ corresponds to a pair (i_p, i_q), for $i_p = 1, \ldots, m_p$, $i_q = 1, \ldots, m_q$, and $i_j \in \langle m_j \rangle$. (For example we may arrange the pairs (i_p, i_q) in the lexicographical order. Then $i_p = \lceil \frac{\ell}{m_q} \rceil$ and $i_q = \ell - (i_p - 1)m_q$.) Set $a_{\ell,i_j} = t_{i_1,i_2,i_3}$. Then $\mathrm{rank}_j(\tau) = \mathrm{rank}\, A(j)$.

The following proposition is straightforward.

Proposition 2.1 *Let* $\tau \in \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3$ *be given by (2.2). Fix* $j \in \{1, 2, 3\}$ *and set* $j^c = \{p, q\}$. *Let* $T_{i_j,j} := [t_{i_1,i_2,i_3}]_{i_p=i_q=1}^{m_p,m_q} \in \mathbb{F}^{m_p \times m_q}$, $i_j = 1, \ldots, m_j$. *Then* $\mathrm{rank}_j(\tau)$ *is the dimension of subspace of* $m_p \times m_q$ *matrices spanned by* $T_{1,j}, \ldots, T_{m_j,j}$.

Assume that each $\mathbf{U}_j$ is an inner product space, with the inner product $\langle \cdot, \cdot \rangle_j$ for $j = 1, 2, 3$. Let $\mathbf{u}_{1,j}, \ldots, \mathbf{u}_{m_j,j}$, $j = 1, 2, 3$ be an orthonormal basis in $\mathbf{U}_j$ with respect to $\langle \cdot, \cdot \rangle_j$. Define an inner product on $\mathbf{U}$, denoted by $\langle \cdot, \cdot \rangle$, by assuming that the basis (2.1) is an orthonormal basis in $\mathbf{U}$. It is straightforward to show that the above inner product does not depend on the choice of the orthonormal bases in $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$. The so defined inner product in $\mathbf{U}$ is called the *induced inner product* and we have identity

$$\langle \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}, \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} \rangle = \langle \mathbf{x}, \mathbf{u} \rangle_1 \langle \mathbf{y}, \mathbf{v} \rangle_2 \langle \mathbf{z}, \mathbf{w} \rangle_3.$$

On $\mathbb{F}^{m_1 \times m_2 \times m_3}$ the standard inner product $\langle \mathcal{X}, \mathcal{Y} \rangle$ is given by $\sum_{i=j=k}^{m_1,m_2,m_3} x_{i,j,k} \bar{y}_{i,j,k}$, where $\mathcal{X} = [x_{i,j,k}], \mathcal{Y} = [y_{i,j,k}]$. This inner product is induced by the standard inner products on $\mathbb{F}^{m_1}, \mathbb{F}^{m_2}, \mathbb{F}^{m_3}$. So $\|\mathcal{X}\| = (\sum_{i=j=k=1}^{m_1,m_2,m_3} |x_{i,j,k}|^2)^{\frac{1}{2}}$ is the *Hilbert-Schmidt norm* on $\mathbb{F}^{m_1 \times m_2 \times m_3}$.

We denote by $\mathrm{Gr}(p, \mathbb{F}^n)$ the set of all p-dimensional subspaces of $\mathbb{F}^n$. It is well known that $\mathrm{Gr}(p, \mathbb{F}^n)$ is a closed set, more precisely an algebraic variety, called the *Grassmannian* of $\mathbb{F}^n$ [6].

Definition 2.2 *Let* $p \in \langle m_1 \rangle, q \in \langle m_2 \rangle, r \in \langle m_3 \rangle$. *Denote by* $\mathrm{Gr}(p, \mathbb{F}^{m_1}) \otimes \mathrm{Gr}(q, \mathbb{F}^{m_2}) \otimes \mathrm{Gr}(r, \mathbb{F}^{m_3}) \subseteq \mathrm{Gr}(pqr, \mathbb{F}^{m_1 \times m_2 \times m_3})$ *the set of all* pqr*-dimensional subspaces in* $\mathbb{F}^{m_1 \times m_2 \times m_3}$ *of the form* $\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z}$, *where* $\mathbf{X} \in \mathrm{Gr}(p, \mathbb{F}^{m_1}), \mathbf{Y} \in \mathrm{Gr}(q, \mathbb{F}^{m_2}), \mathbf{Z} \in \mathrm{Gr}(r, \mathbb{F}^{m_3})$.

Clearly, $\mathrm{Gr}(p, \mathbb{F}^{m_1}) \otimes \mathrm{Gr}(q, \mathbb{F}^{m_2}) \otimes \mathrm{Gr}(r, \mathbb{F}^{m_3})$ is a closed subvariety of $\mathrm{Gr}(pqr, \mathbb{F}^{m_1 \times m_2 \times m_3})$. Define by $\mathrm{dist}(\mathcal{T}, \mathrm{S}, \| \, \|) := \inf_{\mathcal{X} \in \mathrm{S}} \|\mathcal{T} - \mathcal{X}\|$ the distance of $\mathcal{T}$ to a set $\mathrm{S} \subset \mathbb{F}^{m_1 \times m_2 \times m_3}$ with respect to the norm $\| \, \|$. Then the *best* (p, q, r) *subspace approximation of* $\mathcal{T} \in \mathbb{F}^{l \times m \times n}$ is given by

$$\min_{\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \in \mathrm{Gr}(p, \mathbb{F}^{m_1}) \otimes \mathrm{Gr}(q, \mathbb{F}^{m_2}) \otimes \mathrm{Gr}(r, \mathbb{F}^{m_3})} \mathrm{dist}(\mathcal{T}, \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z}, \| \, \|), \tag{2.5}$$

and we denote the subspace where the minimum is achieved by $\mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*$ and the minimal tensor by $\mathcal{X}^* \in \mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*$, i.e. we have

$$\mathrm{dist}(\mathcal{T}, \mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*, \| \, \|) = \|\mathcal{T} - \mathcal{X}^*\|. \tag{2.6}$$

Let $\ell_1 \in \langle m_1 \rangle, \ell_2 \in \langle m_2 \rangle, \ell_3 \in \langle m_3 \rangle$ and suppose that $\mathbf{U}_1 \in \mathrm{Gr}(\ell_1, \mathbb{F}^{m_1}), \mathbf{U}_2 \in \mathrm{Gr}(\ell_2, \mathbb{F}^{m_2}), \mathbf{U}_3 \in \mathrm{Gr}(\ell_3, \mathbb{F}^{m_3})$. Choose

$$\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{\ell_1,1} \in \mathbb{F}^{m_1}, \; \mathbf{u}_{1,2}, \ldots, \mathbf{u}_{\ell_2,2} \in \mathbb{F}^{m_2}, \; \mathbf{u}_{1,3}, \ldots, \mathbf{u}_{\ell_3,3} \in \mathbb{F}^{m_3},$$

such that $\mathbf{u}_{1,j}, \ldots, \mathbf{u}_{\ell_j,j}$ is an orthonormal basis in $\mathbf{U}_j$ for $j = 1, 2, 3$. Then for $\tau \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ let

$$t_{i,j,k} = \langle \tau, \mathbf{u}_{i,1} \otimes \mathbf{u}_{j,2} \otimes \mathbf{u}_{k,3} \rangle, \quad i = 1, \ldots, l,\ j = 1, \ldots, m,\ k = 1, \ldots, n. \tag{2.7}$$

So $\mathcal{T} = [t_{i,j,k}]$ is the representation of τ in the orthonormal basis. Then

$$P_{\mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3}(\tau) = \xi = \sum_{(i,j,k) \in \langle \ell_1 \rangle \times \langle \ell_2 \rangle, \times \langle \ell_3 \rangle} t_{i,j,k} \mathbf{u}_{i,1} \otimes \mathbf{u}_{j,2} \otimes \mathbf{u}_{k,3} \tag{2.8}$$

is the orthogonal projection of τ on the subspace $\mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3$. Thus

$$\operatorname{dist}(\tau, \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3) = \|\tau - \xi\| = \Big(\sum_{(i,j,k) \in \langle m_1 \rangle \times \langle m_2 \rangle \times \langle m_3 \rangle \setminus \langle \ell_1 \rangle \times \langle \ell_2 \rangle \times \langle \ell_3 \rangle} |t_{i,j,k}|^2 \Big)^{\frac{1}{2}} \tag{2.9}$$

is the distance with respect to the Hilbert-Schmidt norm on $\mathbb{F}^{m_1 \times m_2 \times m_3}$. Clearly, we have

$$\|\tau\|^2 = \|P_{\mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3}(\tau)\|^2 + \operatorname{dist}(\tau, \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \mathbf{U}_3)^2. \tag{2.10}$$

3 The SVD as best subspace tensor approximation

In this section we will illustrate that the SVD allows to compute the best subspace tensor approximation for 2-tensors.

Let us view $m_1 \times m_2$ matrices as 2-tensors. Here $\mathbf{x} \otimes \mathbf{y}$ corresponds to the matrix $\mathbf{x}\mathbf{y}^\top$. A tensor $\tau \in \mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2}$ can be viewed as a linear transformation $\tau : \mathbb{F}^{m_1} \to \mathbb{F}^{m_2}$ as follows. First observe that a rank 1 tensor $\mathbf{x} \otimes \mathbf{y}$ gives rise to the linear transformation $(\mathbf{x} \otimes \mathbf{y})(\mathbf{z}) = \langle \mathbf{z}, \bar{\mathbf{y}} \rangle \mathbf{x}$. Now extend this notion to any $\tau \in \mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2}$, which is a sum of rank 1 tensors.

We claim that the best rank k approximation of τ is obtained as the solution to the minimization problem

$$\min_{\mathbf{X} \in \operatorname{Gr}(k, \mathbb{F}^{m_1}), \mathbf{Y} \in \operatorname{Gr}(k, \mathbb{F}^{m_2})} \operatorname{dist}(\tau, \mathbf{X} \otimes \mathbf{Y}) = \operatorname{dist}(\tau, \mathbf{X}^* \otimes \mathbf{Y}^*), \tag{3.1}$$

where $\mathbf{X}^*, \mathbf{Y}^*$ are the subspaces spanned by the k left and right singular vectors of τ associated with the largest k singular values.

Indeed, suppose that the minimum in (3.1) is achieved for some tensor $\alpha \in \mathbf{X}^* \otimes \mathbf{Y}^*$, so rank $\alpha \le k$. Hence the best approximation by a rank k tensor is not worse than the minimum of (3.1). On the other hand, any rank k tensor is an element of sum $\mathbf{X} \otimes \mathbf{Y}$ for some $\mathbf{X} \in \operatorname{Gr}(k, \mathbb{F}^{m_1}), \mathbf{Y} \in \operatorname{Gr}(k, \mathbb{F}^{m_2})$. So the minimum in (3.1) is not bigger than the best rank k approximation. But the best rank k approximation to a given 2-tensor is obtained by the SVD [4].

We now consider the following approximation problems for 2-tensors, which is equivalent to the corresponding matrix problem.

Lemma 3.1 *Let $\mathbf{Y} \subset \mathbb{F}^{m_2}$ be a given $\ell_1 \in \langle m_1 \rangle$ dimensional subspace. For $i \in \langle m_1 \rangle$ and $\tau \in \mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2}$ consider the minimization problem of finding $X \in$* $\operatorname{Gr}(i, \mathbb{F}^{m_1})$ *such that*

$$\min_{\mathbf{X} \in \operatorname{Gr}(i, \mathbb{F}^{m_1})} \operatorname{dist}(\tau, \mathbf{X} \otimes \mathbf{Y}) = \operatorname{dist}(\tau, \mathbf{X}^* \otimes \mathbf{Y}). \tag{3.2}$$

View τ as a linear mapping from $\mathbb{F}^{m_1}$ to $\mathbb{F}^{m_2}$. If $\dim(\tau\mathbf{Y}) \le i$ then $\mathbf{X}^*$ is any subspace that contains $\tau\mathbf{Y}$. If $\dim(\tau\mathbf{Y}) > i$ then $\mathbf{X}^*$ is the subspace spanned by the left singular vectors associated with the i largest singular values of $\tau|_{\mathbf{Y}}$ (which is a linear map $\tau : \mathbf{Y} \to \mathbb{F}^{m_1}$).

Proof. Choose the standard orthonormal basis $\mathbf{e}_1, \ldots, \mathbf{e}_{m_1} \in F^{m_1}$ and an orthonormal basis $\mathbf{y}_1, \ldots, \mathbf{y}_{m_2} \in \mathbb{F}^{m_2}$ such that $\mathbf{Y} = \mathrm{span}(\mathbf{y}_1, \ldots, \mathbf{y}_\ell)$. Let $\mathbf{Y}^\perp = \mathrm{span}(\mathbf{y}_{\ell+1}, \ldots, \mathbf{y}_{m_2})$. Then $\mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2} = \mathbb{F}^{m_1} \otimes \mathbf{Y} \oplus \mathbb{F}^{m_1} \otimes \mathbf{Y}^\perp$ is an orthogonal decomposition of $\mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2}$. This means that we can write τ as

$$\tau = \phi + \psi,\ \phi = P_{\mathbb{F}^{m_1} \otimes \mathbf{Y}}(\tau),\ \psi = P_{\mathbb{F}^{m_1} \otimes \mathbf{Y}^\perp}(\tau),\ \|\tau\|^2 = \|\phi\|^2 + \|\psi\|^2.$$

Since we require $\mathbf{X} \otimes \mathbf{Y} \subset \mathbb{F}^{m_1} \otimes \mathbf{Y}$ it follows that the minimization problem (3.2) is equivalent to the minimization problem

$$\min_{\mathbf{X} \in \mathrm{Gr}(i, \mathbb{F}^{m_1})} \mathrm{dist}(\phi, \mathbf{X} \otimes \mathbf{Y}) = \mathrm{dist}(\tau, \mathbf{X}^* \otimes \mathbf{Y}). \tag{3.3}$$

Observe next that ϕ, viewed as a linear transformation $\phi : \mathbb{F}^{m_1} \to \mathbf{Y}$ is equal to $\tau|\mathbf{Y}$. The classical result for matrices implies that the best rank i approximation of ϕ is given via the left singular vectors associated to the largest i singular values of ϕ. $\square$

In this section we have shown that the best subspace tensor approximation for 2-tensors is obtained via the singular value decomposition. This immediately suggest to use it as a generalization of the SVD for higher tensors.

4 Best subspace tensor approximations for 3-tensors

I n this section we study the best subspace tensor approximation for 3-tensors. Let $\tau \in \mathbb{F}^{m_1 \times m_2 \times m_3}$ and assume that $p \in \langle m_1 \rangle, q \in \langle m_2 \rangle, r \in \langle m_3 \rangle$ and consider the minimization problem

$$\min_{\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \in \mathrm{Gr}(p, \mathbb{F}^{m_1}) \otimes \mathrm{Gr}(q, \mathbb{F}^{m_2}) \otimes \mathrm{Gr}(r, \mathbb{F}^{m_3})} \mathrm{dist}(\tau, \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z}) \tag{4.1}$$

and suppose that is minimum is achieved for the subspace $\mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*$ with the tensor ξ, i.e.

$$\mathrm{dist}(\tau, \mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*) = \|\tau - \xi^*\|,\ \xi^* \in \mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*.$$

In view of (2.10) this minimization problem is equivalent to the maximization problem

$$\max_{\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \in \mathrm{Gr}(p, \mathbb{F}^{m_1}) \otimes \mathrm{Gr}(q, \mathbb{F}^{m_2}) \otimes \mathrm{Gr}(r, \mathbb{F}^{m_3})} \|P_{\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z}}(\tau)\|^2 = \|P_{\mathbf{X}^* \otimes \mathbf{Y}^* \otimes \mathbf{Z}^*}(\tau)\|^2. \tag{4.2}$$

To simplify our exposition we state our results for $\mathbb{F} = \mathbb{R}, \mathbb{C}$, but we give the proofs only for $\mathbb{F} = \mathbb{R}$.

To solve the minimization problem, we study the critical points (i.e. the points of vanishing gradient) of $\|P_{\mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z}}(\tau)\|^2$ on $\mathrm{Gr}(p, \mathbb{F}^{m_1}) \otimes \mathrm{Gr}(q, \mathbb{F}^{m_2}) \otimes \mathrm{Gr}(r, \mathbb{F}^{m_3})$. To do that we need the following lemma which follows from the Courant-Fischer theorem, see e.g. [4]. In the following, we use $\mathrm{Fr}(i, \mathbb{F}^{m_1})$ to denote the manifold of all sets of i orthonormal vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_i\} \subset \mathbb{F}^{m_1}$.

Lemma 4.1 *Let $B \in \mathbb{F}^{m_1 \times m_1}$ be a Hermitian matrix. Let a linear functional $g_B : \mathrm{Fr}(i, \mathbb{R}^{m_1}) \to \mathbb{R}$ be given by $g_B(\mathbf{x}_1, \ldots, \mathbf{x}_i) = \sum_{l=1}^{i} \mathbf{x}_l^\top B \mathbf{x}_l$. Then the critical points of g_B are all sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_i\}$ such that $\mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i)$ contains i linearly independent eigenvectors of B.*

Proof. We prove the lemma by induction on i. For $i = 1$ we have $g_B(x) = \mathbf{x}^\top B\mathbf{x}$ (note that $\|\mathbf{x}\| = 1$). Then by the Courant-Fischer Min-Max characterization, see e.g. [4], $\mathbf{x} \neq \mathbf{0}$ is a critical point if and only if $\mathbf{x}$ is an eigenvector of B.

If $\mathbf{x}_1, \ldots, \mathbf{x}_i$ are eigenvectors of B it is straightforward to see that $\{x_1, \ldots, \mathbf{x}_i\}$ is a critical point of g_B. Indeed, consider a variation $x_\ell(t) = \mathbf{x}_\ell + t\mathbf{u}_\ell + t\mathbf{v}_\ell + O(t^2), \ell = 1, \ldots, i$, where $\mathbf{u}_l \in \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i)$, $\mathbf{v}_l \in \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i)^\perp$. Then the contribution involving $\mathbf{u}_1, \ldots, \mathbf{u}_i$ is quadratic in t. Since $\mathbf{v}_\ell^\top \mathbf{x}_l = 0,\ \ell = 1, \ldots, i$ it follows that the contribution in $\mathbf{v}_1, \ldots, \mathbf{v}_i$ is also quadratic in t. It remains to show that if $\{\mathbf{x}_1, \ldots, \mathbf{x}_i\}$ is a critical point of g_B then $\mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i)$ is spanned by i eigenvectors of B.

Suppose that the assertion holds for $i = k - 1$ and assume that $i = k \leq m_1$. If $k = m_1$ then the assertion is clear because the whole space is spanned by eigenvectors of B. So let $k < m$. Note that if $\{\mathbf{y}_1, \ldots, \mathbf{y}_i\} \in \mathrm{Fr}(i, \mathbb{R}^{m_1})$ and $\mathrm{span}(\mathbf{y}_1, \ldots, \mathbf{y}_i) = \mathrm{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i)$ then $g_B(\mathbf{x}_1, \ldots, \mathbf{x}_i) = g_B(\mathbf{y}_1, \ldots, \mathbf{y}_i)$. So we may assume w.l.o.g. that the matrix

$$C = [\mathbf{x}_s^\top B \mathbf{x}_t]_{s,t=1}^{i}$$

is diagonal. Furthermore, we may assume that $\mathbf{x}_s = \mathbf{e}_s, s = 1, \ldots, i$. The induction hypothesis states that for any $k \in \{i+1, \ldots, m_1\}$ the symmetric matrix B_k, obtained by erasing k rows and columns of B is a direct sum of C and the corresponding other block. Hence $B = C \oplus C'$ and the assertion follows. □

We immediately have the following corollary.

Corollary 4.2 *Let $\alpha \in \mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2}$ and suppose that i is an integer in the interval $[1, m_1]$. Then $\mathbf{U} \in \mathrm{Gr}(i, \mathbb{F}^{m_1})$ is a critical point of the linear functional $\|P_{\mathbf{X} \otimes \mathbb{F}^{m_2}}(\alpha)\|^2 : \mathrm{Gr}(i, \mathbb{F}^{m_1}) \to [0, \infty)$ if and only if $\mathbf{U}$ is spanned by some i left singular vectors of the induced dual operator $\tilde{\alpha} : \mathbb{F}^{m_2} \to \mathbb{F}^{m_1}$. (Here some singular vectors may correspond to the singular value 0.)*

Proof. Represent $\tilde{\alpha}$ by $A \in \mathbb{R}^{m_1 \times m_2}$ and let $B = AA^\top$. Let $\mathbf{X} \in \mathrm{Gr}(i, \mathbb{R}^{m_1})$ and suppose that $\{\mathbf{x}_1, \ldots, \mathbf{x}_i\} \in \mathrm{Fr}(i, \mathbb{R}^m)$ is a basis of $\mathbf{X}$. Then $\|P_{\mathbf{X} \otimes \mathbb{F}^{m_2}}(\alpha)\|^2 = g_B(\mathbf{x}_1, \ldots, \mathbf{x}_i)$, and the result follows from Lemma 4.1. □

We will now construct projections of 3-tensors to 2-tensors, which we can use to compute best subspace approximations.

Let $\tau \in \mathbb{F}^{m_1} \otimes \mathbb{F}^{m_2} \otimes \mathbb{F}^{m_3}$ and $\mathbf{X} \in \mathrm{Gr}(p, \mathbb{F}^{m_1}), \mathbf{Y} \in \mathrm{Gr}(q, \mathbb{F}^{m_2}), \mathbf{Z} \in \mathrm{Gr}(r, \mathbb{F}^{m_3})$. Suppose that $\mathbf{e}_1, \ldots, \mathbf{e}_{m_1},\ \mathbf{f}_1, \ldots, \mathbf{f}_{m_2},\ \mathbf{g}_1, \ldots, \mathbf{g}_{m_3}$ are orthonormal bases in $\mathbb{F}^{m_1}, \mathbb{F}^{m_2}, \mathbb{F}^{m_3}$ respectively, such that $\mathbf{e}_1, \ldots, \mathbf{e}_p,\ \mathbf{f}_1, \ldots, \mathbf{f}_q,\ \mathbf{g}_1, \ldots, \mathbf{g}_r$ are bases of $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, respectively. Then we can express τ as $\tau = \sum_{i=j=k=1}^{m_1,m_2,m_3} t_{i,j,k} \mathbf{e}_i \otimes \mathbf{f}_j \otimes \mathbf{g}_k$ and consider the following linear operators.

1. The first operator $\tau(\mathbf{Y}, \mathbf{Z}) : \mathbb{F}^{m_1} \to \mathbf{Y} \otimes \mathbf{Z}$ is constructed as follows. View

$P_{\mathbb{F}^{m_1}\otimes\mathbf{Y}\otimes\mathbf{Z}}(\tau)$ as a tensor in $\mathbb{F}^{m_1}\otimes\mathbf{Y}\otimes\mathbf{Z}$, i.e.

$$P_{\mathbb{F}^{m_1}\otimes\mathbf{Y}\otimes\mathbf{Z}}(\tau)=\sum_{i=j=k=1}^{m_1,q,r} t_{i,j,k}\mathbf{e}_i\otimes\mathbf{f}_j\otimes\mathbf{g}_k$$

and then define for $\mathbf{x}\in\mathbb{F}^{m_1}$ the operator via

$$\tau(\mathbf{Y},\mathbf{Z})(\mathbf{x})=\sum_{i=j=k=1}^{m_1,q,r} t_{i,j,k}\langle\mathbf{x},\mathbf{e}_i\rangle_1\ \mathbf{f}_j\otimes\mathbf{g}_k,$$

where as before $\langle\cdot,\cdot\rangle_1$ denotes the inner product in $\mathbb{F}^{m_1}$.

2. Analogously we proceed for $\tau(\mathbf{X},\mathbf{Z}):\mathbb{F}^{m_2}\to\mathbf{X}\otimes\mathbf{Z}$. We view $P_{\mathbf{X}\otimes\mathbb{F}^{m_2}\otimes\mathbf{Z}}(\tau)$ as a tensor in $\mathbf{X}\otimes\mathbb{F}^{m_2}\otimes\mathbf{Z}$, i. e.,

$$P_{\mathbf{X}\otimes\mathbb{F}^{m_2}\otimes\mathbf{Z}}(\tau)=\sum_{i=j=k=1}^{p,m_2,r} t_{i,j,k}\mathbf{e}_i\otimes\mathbf{f}_j\otimes\mathbf{g}_k$$

and then for any $\mathbf{y}\in\mathbb{F}^{m_2}$ we define the operator via

$$\tau(\mathbf{X},\mathbf{Z})(\mathbf{y})=\sum_{i=j=k=1}^{p,m_2,r} t_{i,j,k}\langle\mathbf{y},\mathbf{f}_j\rangle_2\ \mathbf{e}_i\otimes\mathbf{g}_k.$$

3. Finally $\tau(\mathbf{X},\mathbf{Y}):\mathbb{F}^{n}\to\mathbf{X}\otimes\mathbf{Y}$ is given as follows. View $P_{\mathbf{X}\otimes\mathbf{Y}\otimes\mathbb{F}^{m_3}}(\tau)$ as a tensor in $\mathbf{X}\otimes\mathbf{Y}\otimes\mathbb{F}^{m_3}$, i. e.,

$$P_{\mathbf{X}\otimes\mathbf{Y}\otimes\mathbb{F}^{m_3}}(\tau)=\sum_{i=j=k=1}^{p,q,m_3} t_{i,j,k}\mathbf{e}_i\otimes\mathbf{f}_j\otimes\mathbf{g}_k.$$

Then for any $\mathbf{z}\in\mathbb{F}^{m_3}$, we define the operator via

$$\tau(\mathbf{X},\mathbf{Y})(\mathbf{z})=\sum_{i=j=k=1}^{p,q,m_3} t_{i,j,k}\langle\mathbf{z},\mathbf{g}_k\rangle_3\ \mathbf{e}_i\otimes\mathbf{f}_j.$$

We have the following theorem.

Theorem 4.3 *Let* $0\neq\tau\in\mathbb{F}^{m_1}\otimes\mathbb{F}^{m_2}\otimes\mathbb{F}^{m_3}$. *Let* $p\in\langle m_1\rangle, q\in\langle m_2\rangle, r\in\langle m_3\rangle$. *Then* $\mathbf{U}\in\mathrm{Gr}(p,\mathbb{F}^{m_1}),\mathbf{V}\in\mathrm{Gr}(q,\mathbb{F}^{m_2}),\mathbf{W}\in\mathrm{Gr}(r,\mathbb{F}^{m_3})$ *is a critical point of* $\|P_{\mathbf{X}\otimes\mathbf{Y}\otimes\mathbf{Z}}(\tau)\|^2$ *on* $\mathrm{Gr}(p,\mathbb{F}^{m_1})\otimes\mathrm{Gr}(q,\mathbb{R}^{m_2})\otimes\mathrm{Gr}(k,\mathbb{F}^{m_3})$ *if and only if the following conditions hold*

1. $\mathbf{U}$ *is spanned by some* p *left singular vectors of* $\tau(\mathbf{V},\mathbf{W})$.

2. $\mathbf{V}$ *is spanned by some* q *left singular vectors of* $\tau(\mathbf{U},\mathbf{W})$.

3. $\mathbf{W}$ *is spanned by some* r *left singular vectors of* $\tau(\mathbf{U},\mathbf{V})$.

Proof. Since the critical points are the zeros of the first derivative, it is enough to prove the necessary conditions for the function $\|P_{\mathbf{X}\otimes\mathbf{V}\otimes\mathbf{W}}(\tau)\|^2$. Considering this as function on $\mathrm{Gr}(p,\mathbb{R}^{m_1})$, Condition 1. then follows immediately by Corollary 4.2. The other conditions follow analogously. □

In the following we will describe an iterative procedure to compute the best subspace tensor approximation. In order to find good starting values for $\mathbf{U}=\mathbf{X}_0, \mathbf{V}=\mathbf{Y}_0, \mathbf{W}=\mathbf{Z}_0$ we make use of the SVD. As explained in §2 we can *unfold* τ as a matrix A_1, say $m_1 \times (m_2 n_3)$, by considering $\tau(1)$ as defined in (2.4). Then we perform the SVD and use as approximation the corresponding p-dimensional $\mathbf{X}_0 \in \mathrm{Gr}(p,\mathbb{F}^{m_1})$ spanned the left singular vectors of A_1 associated with the p largest singular values. In a similar way we determine $\mathbf{Y}_0 \in \mathrm{Gr}(q,\mathbb{F}^{m_2}), \mathbf{Z}_0 \in \mathrm{Gr}(r,\mathbb{F}^{m_3})$.

To find the maximum in (4.2) we then apply a relaxation method.

Algorithm 4.4 *Let* $\tau \in \mathbb{F}^{m_1}\otimes\mathbb{F}^{m_2}\otimes\mathbb{F}^{m_3}$, $p\in\langle m_1\rangle, q\in\langle m_2\rangle, r\in\langle m_3\rangle$ *and starting values* $\mathbf{X}_0 \in \mathrm{Gr}(p,\mathbb{F}^{m_1}), \mathbf{Y}_0 \in \mathrm{Gr}(q,\in\mathbb{F}^{m_2}), \mathbf{Z}_0 \in \mathrm{Gr}(r,\mathbb{F}^{m_3})$ *be given.*

Suppose that $(\mathbf{X}_i,\mathbf{Y}_i,\mathbf{Z}_i)$ *have been computed. Then*

1. $\mathbf{X}_{i+1}$ *is obtained as the* p*-dimensional subspace corresponding to left singular vectors of* $\tau(\mathbf{Y}_i,\mathbf{Z}_i)$ *associated with the* p *largest singular values.*

2. $\mathbf{Y}_{i+1}$ *is obtained as the* q*-dimensional subspace corresponding to the left singular vectors of* $\tau(\mathbf{X}_{i+1},\mathbf{Z}_i)$ *associated with the* q *largest singular values.*

3. $\mathbf{Z}_{i+1}$ *is obtained as the* r*-dimensional subspace corresponding to the left singular vectors of* $\tau(\mathbf{X}_{i+1},\mathbf{Y}_{i+1})$ *associated with the* r *largest singular values.*

We have the following convergence result.

Corollary 4.5 *The subspaces* $\mathbf{X}_i,\mathbf{Y}_i,\mathbf{Z}_i, i=0,1,\ldots$ *defined in Algorithm 4.4 converge to subspaces* $\mathbf{U},\mathbf{V},\mathbf{W}$ *which give a critical point of* $\|P_{\mathbf{X}\otimes\mathbf{Y}\otimes\mathbf{Z}}(\tau)\|^2$. *Moreover, this critical point is a maximal point, with respect to any one variable, when the other variables are fixed. Furthermore the following conditions hold.*

1. $\mathbf{U}$ *is spanned by the left singular vectors of* $\tau(\mathbf{V},\mathbf{W})$ *associated with the* p *largest values.*

2. $\mathbf{V}$ *is spanned by the left singular vectors of* $\tau(\mathbf{U},\mathbf{W})$ *associated with the* q *largest values.*

3. $\mathbf{W}$ *is spanned by the left singular vectors of* $\tau(\mathbf{U},\mathbf{V})$ *associated with the* r *largest singular values.*

In this section we have shown that the best subspace tensor approximation for 3-tensors is a a generalization of the singular value decomposition. It is obvious how this procedure can be extended to arbitrary k tensors.

Unfortunately the described procedure is extremely expensive, since in every step a singular value decomposition of a very large full matrix has to be performed. In order to reduce the complexity, in the next section we consider a procedure that is based on the recently suggested fast SVD [3].

5 Fast low rank 3-tensors approximations

In this section we generalize the algorithm outlined in [3] to the fast low rank tensor approximation, abbreviated as FLRTA, to 3-tensors. Let $\mathcal{A} = [a_{i_1,i_2,i_3}] \in \mathbb{R}^{l_1 \times l_2 \times l_3}$ be a 3-tensor, where the dimensions l_1, l_2, l_3, are large. For each $j = 1, 2, 3$ we read subtensors of $\mathcal{A}$ denoted by $\mathcal{C}_j = [c^{(j)}_{i_{1,j} i_{2,j} i_{3,j}}] \in \mathbb{R}^{l_{1,j} \times l_{2,j} \times l_{3,j}}$. We assume that $\mathcal{C}_j$ has the same number of coordinates as $\mathcal{A}$ in j-th direction, and a small number of coordinates in the other two directions. That is, $l_{j,j} = l_j$ and the other two indices $l_{s,j},\ s \in \{1, 2, 3\}\backslash\{j\}$ are of order $\mathcal{O}(k)$, for $j = 1, 2, 3$. So $\mathcal{C}_j$ corresponds to the j-*section* of the tensor $\mathcal{A}$. The *small* dimensions of $\mathcal{C}_j$ are $(l_{s_j,j}, l_{t_j,j})$ where $\{s_j, t_j\} = \{1, 2, 3\}\backslash\{j\}$ for $j = 1, 2, 3$. Let $m_j := l_{s_j,j} l_{t_j,j}$ for $j = 1, 2, 3$.

To determine an approximation, we then look for a 6-tensor

$$\mathcal{V} = [v_{q_1,q_2,q_3,q_4,q_5,q_6}] \in \mathbb{R}^{l_{2,1} \times l_{3,1} \times l_{1,2} \times l_{3,2} \times l_{1,3} \times l_{2,3}}$$

and approximate the given tensor $\mathcal{A}$ by a tensor

$$\mathcal{B} = [b_{i_1,i_2,i_3}] := \mathcal{V} \cdot \mathcal{C}_1 \cdot \mathcal{C}_2 \cdot \mathcal{C}_3 \in \mathbb{R}^{\ell_1 \times \ell_2 \times \ell_3},$$

where we contract the 6 indices in $\mathcal{V}$ and the corresponding two indices $\{1, 2, 3\}\backslash\{j\}$ in $\mathcal{C}_j$ for $j = 1, 2, 3$, i.e., our approximation has the entries

$$b_{i_1,i_2,i_3} = \sum_{q_1=1}^{\ell_{2,1}} \sum_{q_2=1}^{\ell_{3,1}} \sum_{q_3=1}^{\ell_{1,2}} \sum_{q_4=1}^{\ell_{3,2}} \sum_{q_5=1}^{\ell_{1,3}} \sum_{q_6=1}^{\ell_{2,3}} v_{q_1,q_2,q_3,q_4,q_5,q_6} c^{(1)}_{i_1,q_1,q_2} c^{(2)}_{q_3,i_2,q_4} c^{(3)}_{q_5,q_6,i_3}. \tag{5.1}$$

This approximation is equivalent to a so-called *Tucker approximation* [10]. Indeed, if we represent each tensor $\mathcal{C}_j$ by a matrix $C_j \in \mathbb{R}^{m_j \times l_j}$ that has the same number of columns as the range of the j-th index of the tensor $\mathcal{A}$ and as number of rows the product of the ranges of the remaining two *small* indices of $\mathcal{C}_j$, i.e. $C_j = [c^{(j)}_{r,i_j}]^{m_j \cdot \ell_j}_{r,i_j=1}$. Then $c^{(j)}_{r,i_j}$ is equal to the corresponding entry $c^{(j)}_{i_1,i_2,i_3}$, where the value of r corresponds to the double index (i_s, i_t) for $\{s, t\} = \{1, 2, 3\}\backslash\{j\}$.

Now with $\mathcal{U} = [u_{j_1,j_2,j_3}] \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, the equivalent Tucker representation of $\mathcal{B} = [b_{i_1,i_2,i_3}]$ is given by the entries

$$b_{i_1,i_2,i_3} = \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} \sum_{j_3=1}^{m_3} u_{j_1,j_2,j_3} c^{(1)}_{j_1,i_1} c^{(2)}_{j_2,i_2} c^{(3)}_{j_3,i_3}, \quad (i_1, i_2, i_3) \in \langle \ell_1 \rangle \times \langle \ell_2 \rangle \times \langle \ell_3 \rangle. \tag{5.2}$$

This formula is expressed commonly as

$$\mathcal{B} = \mathcal{U} \times_1 C_1 \times_2 C_2 \times_3 C_3. \tag{5.3}$$

We now choose three subsets of the rows, columns and heights of $\mathcal{A}$

$$I \subset \langle \ell_1 \rangle,\ \#I = p, \quad J \subset \langle \ell_2 \rangle,\ \#J = q, \quad K \subset \langle \ell_3 \rangle,\ \#K = r. \tag{5.4}$$

Let

$$\begin{aligned}
\mathcal{C}_1 &= \mathcal{A}_{\langle \ell_1 \rangle, J, K} := [a_{i,j,k}] \in \mathbb{R}^{\ell_1 \times q \times r},\ i \in \langle \ell_1 \rangle, j \in J, k \in K,\\
\mathcal{C}_2 &= \mathcal{A}_{I, \langle \ell_2 \rangle, K} := [a_{i,j,k}] \in \mathbb{R}^{p \times \ell_2 \times r},\ i \in I, j \in \langle \ell_2 \rangle, k \in K,\\
\mathcal{C}_3 &= \mathcal{A}_{I, J, \langle \ell_3 \rangle} := [a_{i,j,k}] \in \mathbb{R}^{p \times q \times \ell_3},\ i \in I, j \in J, k \in \langle \ell_3 \rangle,\\
\mathcal{S} &= (\langle \ell_1 \rangle \times J \times K) \cup (I \times \langle \ell_2 \rangle \times K) \cup (I \times J \times \langle \ell_3 \rangle).
\end{aligned} \tag{5.5}$$

We define $\mathcal{U}_b$ and $\mathcal{U}_{\text{opt}}$ as in [3].

$$\mathcal{U}_b = \arg\min_{\mathcal{U}\in\mathbb{R}^{m_1\times m_2\times m_3}} \sum_{(i,j,k)\in\langle l_1\rangle\times\langle l_2\rangle\times\langle l_3\rangle} (a_{i,j,k} - (\mathcal{U}\times_1 C_1 \times_2 C_2 \times_3 C_3)_{i,j,k})^2, \quad (5.6)$$

$$\mathcal{U}_{\text{opt}} = \arg\min_{\mathcal{U}\in\mathbb{R}^{m_1\times m_2\times m_3}} \sum_{(i,j,k)\in\mathcal{S}} (a_{i,j,k} - (\mathcal{U}\times_1 C_1 \times_2 C_2 \times_3 C_3)_{i,j,k})^2. \quad (5.7)$$

Instead of computing $\mathcal{U}_{\text{opt}}$ we do the following approximations, as suggested in [3] for the case $q = p, r = p^2$. Unfold the tensor $\mathcal{A} = [a_{i,j,k}]$ in the direction 3 to obtain the matrix $E = [e_{s,k}] \in \mathbb{R}^{(\ell_1\cdot\ell_2)\times\ell_3}$. So $e_{s,k} = a_{i,j,k}$ for the corresponding pair of indices $(i,j) \in \langle\ell_1\rangle \times \langle\ell_2\rangle$. Then the set of indices $(i,j) \in I \times J$ corresponds to the set of indices $L \subset \langle\ell_1 \cdot \ell_2\rangle$, where $\#L = pq$. Denote by $E_{L,K}$ the submatrix of E which has row indices in L and column indices in K. Let $E^\dagger_{L,K} \in \mathbb{R}^{r\times(pq)}$ be the Moore-Penrose inverse of $E_{L,K}$. As in [3] we approximate the tensor $\mathcal{A}$ by

$$\mathcal{A}_{\langle\ell_1\rangle,\langle\ell_2\rangle,K} E^\dagger_{L,K} \mathcal{A}_{I,J,\langle\ell_3\rangle}. \quad (5.8)$$

For each $k \in K$ consider the matrix

$$F_k := \mathcal{A}_{\langle\ell_1\rangle,\langle\ell_2\rangle,k} = [a_{i,j,k}]_{i,j=1}^{\ell_1,\ell_2} \in \mathbb{R}^{\ell_1\times\ell_2}.$$

Next we approximate F_k by $G_k := (F_k)_{\langle\ell_1\rangle,J}(F_k)^\dagger_{I,J}(F_k)_{I,\langle\ell_2\rangle}$. As in [3] we try several random choices of I, J, K with the cardinalities p, q, r respectively, with the best preset conditions numbers for the matrices $E_{L,K}$ and $(F_k)_{I,J}$ for $k \in K$.

Equivalently, we have that

$$\mathcal{A}_{\langle\ell_1\rangle,J,k} \mathcal{A}^\dagger_{I,J,k} \mathcal{A}_{I,\langle\ell_2\rangle,k}, \quad (5.9)$$

is an approximation of $\mathcal{A}_{\langle\ell_1\rangle,\langle\ell_2\rangle,k}$. Replacing $\mathcal{A}_{\langle\ell_1\rangle,\langle\ell_2\rangle,k}$ appearing in (5.8) with the expression that appears in (5.9), we obtain the approximation $\mathcal{B}$ of the form (5.3).

References

[1] S. Friedland, On the generic rank of 3-tensors, arXiv: 0805.3777v1.

[2] S. Friedland, M. Kaveh, A. Niknejad and H. Zare, Fast Monte-Carlo low rank approximations for matrices, *Proc. IEEE Conference SoSE*, Los Angeles, 2006, 218-223.

[3] S. Friedland, V. Mehrmann, A. Miedlar and M. Nkengla, Fast low rank approximations of matrices and tensors, *submitted*, www.matheon.de/preprints/4903.

[4] G.H. Golub and C.F. Van Loan, *Matrix Computation*, John Hopkins Univ. Press, 3rd Ed., 1996.

[5] S.A. Goreinov and E.E. Tyrtyshnikov, The maximum-volume concept in approximation by low-rank matrices, *Contemporary Mathematics* 280 (2001), 47-51.

[6] J. Harris, *Algebraic Geometry*, A First Course, 1992, Springer, New York.

[7] B.N. Khoromskij, Methods of Tensor Approximation for Multidimensional Operators and Functions with Applications, *Lecture at the workschop*, Schnelle Löser für partielle Differentialgleichungen, Oberwolfach, 18.-23.05, 2008.

[8] L. de Lathauwer, B. de Moor and J. Vandewalle, A Multilinear singular value decomposition, *SIAM J. Matrix Anal. Appl.* 21 (2000), 1253-1278.

[9] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," to appear in SIAM Journal on Matrix Analysis and Applications.

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