Numerische Mathematik manuscript No.
(will be inserted by the editor)

Well-posedness of convex maximization problems on Stiefel manifolds and orthogonal tensor product approximations

André Uschmajew

Received: date / Accepted: date

Abstract Problems of best tensor product approximation of low orthogonal rank can be formulated as maximization problems on Stiefel manifolds. The functionals that appear are convex and weakly sequentially continuous. It is shown that such problems are always well-posed, even in the case of non-compact Stiefel manifolds. As a consequence, problems of finding a best orthogonal, strong orthogonal or complete orthogonal low-rank tensor product approximation and problems of best Tucker format approximation to any given tensor are always well-posed, even in spaces of infinite dimension. (The best rank-one approximation is a special case of all of them.) In addition, the well-posedness of a canonical low-rank approximation with bounded coefficients can be shown. The proofs are non-constructive and the problem of computation is not addressed here.

Mathematics Subject Classification (2000) 15A69 · 41A46 · 41A63

1 Introduction

Tensor products are under intensive study in several fields of application. They seem to point the right direction for tackling real high dimensional problems in numerics or quantum mechanics [2,4,6,8,12,14,17,19]. Usually, they allow the exact or approximative separation of a high dimensional problem into a set of low dimensional problems. To be more specific, one ideally seeks a representation of a tensor f as a linear combination $f = \sum_{k=1}^{r} \phi_k^1 \otimes \cdots \otimes \phi_k^N$ of tensor products of order N with a rather low rank r. The rank of the tensor is, roughly speaking, the minimal number of summands that is needed to represent it as such a linear combination. One speaks of a

André Uschmajew
Technische Universität Berlin
Institut für Mathematik
Straße des 17. Juni 136
10623 Berlin, Germany
E-mail: uschmajew@math.tu-berlin.de

tensor product decomposition. But if the rank of f is very large or even infinite it cannot be decomposed into a small number of terms and one tries to approximate f by a tensor of low rank instead. It is our aim to prove some existence results concerning such *best low-rank approximation* problems under certain *orthogonality constraints*.
We will consider tensor products of Hilbert spaces and prove the existence of best low-rank approximations in the sense of least squares for several types of *orthogonal rank* and for *canonical rank with bounded coefficients*. While it has been shown that the unconstrained best low-rank approximation problem is ill-posed in many cases [19], even in finite dimension, it turns out that orthogonality constraints, which are required and useful for many other reasons in practice, provide enough additional structure to prove the existence of such best approximations.
The common theoretical framework for our problems is the maximization of convex and weakly continuous functionals over *Stiefel manifolds*. Our main result is that such problems are always well-posed. The first section of this article is devoted to the proof of this assertion. This is done with methods of nonlinear functional analysis. The second part of the paper shows how this result can be applied to several kinds of orthogonal low-rank tensor approximation problems.
Let us point out that all our results can be and have been easily derived in the case of finite-dimensional spaces since the underlying manifolds can be assumed to be compact then. This is not true for infinite-dimensional spaces, which appear for instance in quantum chemistry, and the existence of best approximations is not obvious. Thus, the paper should be read with focus on infinite-dimensional spaces. In addition, if the rank of the target tensor f is finite, one should be able to reformulate the problem of best approximation as a problem in finite dimensional tensor product spaces in which f is embedded, as it was demonstrated in [4] for the canonical low-rank approximation or Tucker approximation[1]. So the attention here lies on target tensors of infinite rank.
At the moment we are not able to give any quantitative results. We are only able to show existence and, moreover, our proofs are non-constructive. Although questions of existence might be considered as more or less irrelevant for practical issues by some reader (especially for spaces of infinite dimension, which do not occur in numerical practice), it is our belief, that the presented results are one step more towards an approximation theory for tensor products. In addition, they give theoretical justification for several approximation methods, e.g. the incremental rank-one approximation [6,20], and provide deeper insight into the structure of tensor product spaces.

[1] However, we were not able to show this for the orthogonal tensor product approximation presented later in this text, but we believe that it is true.

2 Convex maximization on Stiefel manifolds

2.1 Gram matrices and Stiefel manifolds

Orthogonality conditions can be elegantly formulated in terms of Gram matrices. We begin our work by collecting some properties of such matrices, which will be crucial for all our further considerations.

Let $\mathscr{H}$ be a Hilbert space[2] and $\phi = (\phi_1, \dots, \phi_r)$ a vector from $\mathscr{H}^r$, $r \in \mathbb{N}$ with $r \leq \dim \mathscr{H}$, which also becomes a Hilbert space if equipped with the inner product

$$\langle \phi, \psi \rangle_{\mathscr{H}^r} = \sum_{k=1}^{r} \langle \phi_k, \psi_k \rangle_{\mathscr{H}}.$$

The *Gram matrix* of $\phi \in \mathscr{H}^r$ is defined by

$$G(\phi) = G(\phi_1, \dots, \phi_r) = [\langle \phi_k, \phi_l \rangle_{\mathscr{H}}]_{k,l=1,\dots,r} \in \mathbb{C}^{r,r}.$$

For every ϕ this is a positive semi-definite matrix and it is well known that the rank of $G(\phi)$ is exactly the number of linearly independent elements in ϕ. For simplicity we will use the same symbol G for different spaces and vectors of different length.

The cone of positive semi-definite Hermitian matrices defines a partial order $\preceq$ on the real space of Hermitian matrices via $A \preceq B$ iff $B - A$ is positive semi-definite. This is often called the *Loewner partial order* [10].

Lemma 2.1 *The Gram matrix G is a convex function on $\mathscr{H}^r$ with respect to the Loewner partial order, i.e., for all $\phi, \psi \in \mathscr{H}^r$ and $0 \leq t \leq 1$ we have*

$$G(t\phi + (1-t)\psi) \preceq tG(\phi) + (1-t)G(\psi).$$

Proof Let $V = \operatorname{span}\{\phi_1, \dots, \phi_r, \psi_1, \dots, \psi_r\}$ and $d = \dim V$. Then we can find an orthonormal basis $\{\xi_1, \dots, \xi_d\}$ of V to write $\phi_k = \sum_{l=1}^{d} a_{kl} \xi_l$ and $\psi_k = \sum_{l=1}^{d} b_{kl} \xi_l$ for some matrices $A = [a_{kl}]$ and $B = [b_{kl}]$ from $\mathbb{C}^{r,d}$. The Gram matrices thus become $G(\phi) = AA^*$ and $G(\psi) = BB^*$. Moreover, for $0 \leq t \leq 1$ we have

$$\begin{aligned} G(t\phi + (1-t)\psi) &= (tA + (1-t)B)(tA + (1-t)B)^* \\ &= tAA^* + (1-t)BB^* - t(1-t)(A-B)(A-B)^*. \end{aligned}$$

Since the matrix $t(1-t)(A-B)(A-B)^*$ is positive semi-definite, the lemma is proved. □

The elements of a vector $\phi \in \mathscr{H}^r$ are orthonormal iff $G(\phi) = I$, where I denotes the identity matrix of dimension r. We will have to deal with maximization problems on manifolds of the form

$$\mathscr{S}_r(\mathscr{H}) = \{\phi \in \mathscr{H}^r \mid G(\phi) = I\}. \tag{2.1}$$

Definition 2.2 Manifolds of the form (2.1) are called *Stiefel manifolds of index r*.

[2] For the sake of completeness, we also call finite-dimensional spaces Hilbert spaces. All our Hilbert spaces are assumed to be separable.

The Stiefel manifold of index r consists here of all orthonormal bases (in form of vectors) of all r-dimensional subspaces of the underlying Hilbert space $\mathscr{H}$. It is practical for us to deal with it in such an explicit fashion as a submanifold of $\mathscr{H}^r$. For instance, if $\mathscr{H}$ is an r-dimensional complex space, then $\mathscr{S}_r(\mathscr{H})$ is isomorphic to the closed and bounded subset of unitary matrices in $\mathbb{C}^{r,r} \cong \mathscr{H}^r$. For some differential geometric analysis of the Stiefel manifolds and approaches to numerical algorithms on them see [3].

A Stiefel manifold is compact if and only if $\mathscr{H}$ is finite-dimensional. Moreover, this is even true for the weak topology. Therefore it is not surprising that questions of well-posedness of optimization problems on Stiefel manifolds are almost trivial in finite dimension. In contrast, our focus is the infinite-dimensional case.

An often used technique for non-compact optimization problems is to relax the constraints for a time being. In our case it is useful to consider the *relaxed constraint Stiefel manifolds*

$$\mathscr{R}_r(\mathscr{H}) = \{\phi \in \mathscr{H}^r \mid G(\phi) \preceq I\}.$$

The condition $G(\phi) \preceq I$ implies that all eigenvalues of $G(\phi)$ lie between 0 and 1, but note that $G(\phi) \preceq I$ together with $G(\phi) \neq I$ does not imply that all eigenvalues are strictly smaller than 1, i.e. $G(\phi) \prec I$.

Lemma 2.3 *The set $\mathscr{R}_r(\mathscr{H})$ is bounded, closed and convex in $\mathscr{H}^r$.*

Proof It is bounded because the diagonal elements $G(\phi)|_{kk} = \|\phi_k\|^2_{\mathscr{H}}$ are dominated by the largest eigenvalue of $G(\phi)$, and it is closed because the largest eigenvalue depends continuously on ϕ. The convexity follows from Lemma 2.1, since for $\phi, \psi \in \mathscr{R}_r(\mathscr{H})$ and $0 \le t \le 1$ we have

$$G(t\phi + (1-t)\psi) \preceq tG(\phi) + (1-t)G(\psi) \preceq tI + (1-t)I = I,$$

thus $t\phi + (1-t)\psi \in \mathscr{R}_r(\mathscr{H})$. □

Since, by the Weierstrass Existence Theorem [22], every weakly sequentially continuous function attains a minimum and a maximum value on a weakly sequentially compact set[3], the next lemma will be crucial.

Lemma 2.4 *The set $\mathscr{R}_r(\mathscr{H})$ is weakly sequentially compact in $\mathscr{H}^r$, that is, every sequence in $\mathscr{R}_r(\mathscr{H})$ contains a weakly convergent subsequence with limit point in $\mathscr{R}_r(\mathscr{H})$.*

Proof By Lemma 2.3 every sequence in $\mathscr{R}_r(\mathscr{H})$ is bounded and thus contains a weakly convergent subsequence, because $\mathscr{H}^r$ is a Hilbert space. It is known (Mazur Lemma) that the limit point of such a sequence belongs to the closed convex hull of its elements which, again by Lemma 2.3, is contained in $\mathscr{R}_r(\mathscr{H})$. Both facts are even true in the more general case of a reflexive Banach space, see [22]. □

[3] See Lemma 2.6. We assume the reader to be familiar with the concepts of weak convergence and weak sequential continuity. A good reference for a sufficient overview is [22].

The previous result originally came to our attention by the work [18] of Lieb and Simon on the existence of a Hartree-Fock ground state in quantum chemistry. They gave another interesting proof that is not based on convexity (or at least in a less explicit fashion).

Now as a final observation consider two Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$, their tensor product $\mathscr{H}_1 \otimes \mathscr{H}_2$ (see section 3.1) and vectors $\phi \in \mathscr{H}_1^r$ and $\psi \in \mathscr{H}_2^r$. Then due to (3.2) the Gram matrix of $\phi_1 \otimes \psi_1, \dots, \phi_r \otimes \psi_r$ is given by

$$\begin{aligned} G(\phi_1 \otimes \psi_1, \dots, \phi_r \otimes \psi_r) &= [\langle \phi_k \otimes \psi_k, \phi_l \otimes \psi_l \rangle_{\mathscr{H}_1 \otimes \mathscr{H}_2}]_{k,l=1,\dots,r} \\ &= [\langle \phi_k, \phi_l \rangle_{\mathscr{H}_1} \cdot \langle \psi_k, \psi_l \rangle_{\mathscr{H}_2}]_{k,l=1,\dots,r} \\ &= G(\phi) \circ G(\psi), \end{aligned} \tag{2.2}$$

where $\circ$ denotes the *Hadamard product* of two matrices defined by $A \circ B|_{kl} = a_{kl} b_{kl}$.

Remark 2.5 A Gram matrix is always positive semi-definite. On the other hand, every positive semi-definite matrix can be seen as a Gram matrix (see [10, p. 407]). So we incidentally have found a proof for the Schur Product Theorem for Hadamard products:

If $A, B \in \mathbb{C}^{r,r}$ are positive semi-definite matrices, then $A \circ B$ is also positive semi-definite. Moreover, if either A or B is positive definite, then $A \circ B$ is positive definite.

The second assertion can be derived from the fact that a set $\{\phi_1 \otimes \psi_1, \dots, \phi_r \otimes \psi_r\}$ of tensor products is linearly independent if one of the sets of directions $\{\phi_1, \dots, \phi_r\}$ or $\{\psi_1, \dots, \psi_r\}$ is linearly independent (see [7, p. 7f]). For different proofs of the Schur Product Theorem see [10, 11].

2.2 Convex maximization

Consider a convex functional $F\colon \mathscr{H}^r \to \mathbb{R}$, where again $\mathscr{H}$ is a Hilbert space and $r \in \mathbb{N}$, $r \le \dim \mathscr{H}$. We want to study the problem

$$F(\phi) = \max, \quad \phi \in \mathscr{S}_r(\mathscr{H}). \tag{2.3}$$

It is known that a continuous convex functional F will attain minimum values on closed, convex and bounded sets, e.g. $\mathscr{R}_r(\mathscr{H})$. Minimizers will even be unique if the functional is strictly convex. The situation is a little bit different in the case of maximization. Continuity and convexity are not sufficient for the existence of a maximizer in the infinite-dimensional case. For this, the stronger additional assumption is needed that F is *weakly sequentially continuous*[4], since a closed, convex and bounded set in a Hilbert space is, as we have stated above, indeed weakly sequentially compact. In general, we can use the Weierstrass Theorem in the following form.

Lemma 2.6 *Let $\mathscr{H}$ be a Hilbert space and $\mathscr{M} \subseteq \mathscr{H}$ be bounded and weakly sequentially closed. If $F\colon \mathscr{M} \to \mathbb{R}$ is weakly sequentially continuous, then the problem*

$$F(\phi) = \mathit{max}, \quad \phi \in \mathscr{M},$$

has a solution.

[4] In fact, it is sufficient that F is weakly upper sequentially continuous [22].

It is heuristically clear that the maximizer of a convex functional F over a convex set has to lie on the boundary of that set. Indeed, the only case in which this is not true is when F is constant.

Lemma 2.7 *Let $F : \mathscr{M} \to \mathbb{R}$ be convex, where $\mathscr{M}$ is a closed and convex subset of a normed space V. Assume $\phi^* \in \mathscr{M}$ is a maximizer of F, i.e., $F(\phi^*) = \sup\{F(\phi) \mid \phi \in \mathscr{M}\}$. If $\phi^* \notin \partial\mathscr{M}$, then F is constant. In particular, if F is strictly convex, every maximizer belongs to $\partial\mathscr{M}$.*

Proof Suppose $\phi^* \notin \partial\mathscr{M}$ and F is not constant. Then there exists a $\psi_1 \in \mathscr{M}$ with $F(\psi_1) < F(\phi^*)$. Since $\phi^* \notin \partial\mathscr{M}$ we have an $\varepsilon > 0$ such that $\psi_2 = \phi^* - \varepsilon(\psi_1 - \phi^*)$ also belongs to $\mathscr{M}$ and by assumption $F(\psi_2) \le F(\phi^*)$. Then $\phi^* = \frac{\varepsilon}{1+\varepsilon}\psi_1 + \frac{1}{1+\varepsilon}\psi_2$ and the convexity of F yields

$$F(\phi^*) \le \frac{\varepsilon}{1+\varepsilon}F(\psi_1) + \frac{1}{1+\varepsilon}F(\psi_2) < \frac{\varepsilon}{1+\varepsilon}F(\phi^*) + \frac{1}{1+\varepsilon}F(\phi^*) = F(\phi^*),$$

a contradiction. If F is strictly convex it cannot be constant. So in this case ϕ^* has to belong to $\partial\mathscr{M}$. □

The above result is a typical example how a relaxed constraint technique can work. In many cases a relaxed constraint yields a compact set and with that a solution to the problem. Additional considerations (here convexity) are then used to show that the solution lies on the boundary of that "relaxed" set. In this context we should return the attention to problem (2.3) again. By Lemma 2.3 we now know that the relaxed constraint problem

$$F(\phi) = \max, \quad \phi \in \mathscr{R}_r(\mathscr{H}) \ \text{(i.e. } G(\phi) \preceq I)$$

admits a solution if F is weakly sequentially continuous and convex. Lemma 2.7 tells us that the solution has to lie on the boundary of $\mathscr{R}_r(\mathscr{H})$. Now it seems intuitive to therefore replace $G(\phi) \preceq I$ by $G(\phi) = I$ and conclude $\phi \in \mathscr{S}_r(\mathscr{H})$ since this would have to be done in the case of real valued inequality constraints to describe the boundary. Unfortunately, the Loewner order is only a partial order and the boundary of the relaxed Stiefel manifold consists of all $\phi \in \mathscr{H}^r$ for which the largest eigenvalue of $G(\phi)$ equals 1. The Stiefel manifold $\mathscr{S}_r(\mathscr{H})$ thus only describes a small part of the boundary $\partial\mathscr{R}_r(\mathscr{H})$. Our first important result (Theorem 2.9) will show that it is nevertheless valid to replace $G(\phi) \preceq I$ by $G(\phi) = I$ in the case of convex functionals. For this purpose we first make a more specific statement on the convexity of $\mathscr{R}_r(\mathscr{H})$, which shows that only points from $\mathscr{S}_r(\mathscr{H})$ are "sharp" peak points of the boundary.

Lemma 2.8 *Let $\phi \in \mathscr{R}_r(\mathscr{H})$ and $G(\phi) \ne I$. Then there exists an $h \in \mathscr{H}^r$ such that the whole line segment connecting $\phi - h$ and $\phi + h$ belongs to $\mathscr{R}_r(\mathscr{H})$ and $G(\phi + h) = I$, i.e., $\phi + h \in \mathscr{S}_r(\mathscr{H})$. In particular, if $\phi \in \partial\mathscr{R}_r(\mathscr{H})$ then the whole line segment belongs to $\partial\mathscr{R}_r(\mathscr{H})$.*

Proof Let $\xi_1, \ldots, \xi_r$ be an orthonormal basis of an r-dimensional subspace of $\mathscr{H}$ such that $\phi_k = \sum_{l=1}^r a_{kl}\xi_l$ for $k = 1, \ldots, r$ with $A = [a_{kl}] \in \mathbb{C}^{r,r}$. Then we have $G(\phi) =$

$AA^* \neq I$. Let $A = PU$ be the polar decomposition of A with $P = (AA^*)^{1/2}$ and $U = [u_{ij}]$ unitary[5] (see e.g. [10]). Define $\psi \in \mathscr{H}^r$ by

$$\psi_k = \sum_{l=1}^{r} u_{kl}\xi_l, \quad k = 1, \ldots, r,$$

that is, the element ψ is, in the same basis ξ, represented by the matrix U. Set $h = \psi - \phi$, then clearly $h \neq 0$ and $G(\phi + h) = G(\psi) = UU^* = I$. Because $0 \preceq P \preceq I$ we have

$$G(\phi + th) = [A + t(U - A)][A + t(U - A)]^* = [(1-t)P + tI]^2 \preceq I$$

for $-1 \leq t \leq 1$ so that $\phi + th$ belongs to $\mathscr{R}_r(\mathscr{H})$. If $\phi \in \partial\mathscr{R}_r(\mathscr{H})$ then the largest eigenvalue of P is 1. The same is then true for $[(1-t)P + tI]$ with $-1 \leq t \leq 1$, which proves the second assertion. □

As we have seen in the proof, the natural *scaling*[6] tool for turning $G(\phi) = AA^* \preceq I$ into an equality is given by the polar decomposition PU of A. This is an interesting observation that generalizes the scaling of a complex number in radial direction. The "angular" matrix U remains unchanged while the "radial" matrix $P \preceq I$ is linearly scaled to the identity via $P + t(I - P)$, $0 \leq t \leq 1$. For a better understanding let us give another description of this scaling that might appear more familiar to some reader. If $G(\phi) \preceq I$, there is a unitary matrix V such that for $\widetilde{\psi} = V\phi$ we have $G(\widetilde{\psi}) = VG(\phi)V^* = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ with $0 \leq \lambda_k = \|\widetilde{\psi}_k\|^2_{\mathscr{H}} \leq 1$ for $k = 1, \ldots, r$. Now define $\widetilde{h}$ by $\widetilde{h}_k = \mu_k \widetilde{\psi}_k$ in a way that $\|\widetilde{\psi}_k + \widetilde{h}_k\|^2_{\mathscr{H}} = 1$. If $\widetilde{\psi}_k = 0$, pick as $\widetilde{h}_k$ any normed element from $\mathscr{H}$ which is orthogonal to the other $\widetilde{\psi}_l$. Then we have $G(\widetilde{\psi} + \widetilde{h}) = I$ and, moreover, $G(\widetilde{\psi} + t\widetilde{h}) \preceq I$ for $-1 \leq t \leq 1$. For $h = V^*\widetilde{h}$, which, in the case that $G(\phi)$ is invertible, is exactly the h from the proof above[7], we have $G(\phi + th) = G(V^*(\widetilde{\psi} + t\widetilde{h})) = V^*G(\widetilde{\psi} + t\widetilde{h})V \preceq I$ for $-1 \leq t \leq 1$. This shows how the eigenvalues of $G(\phi)$ are simultaneously scaled to 1. If an eigenvalue λ_k is already 1 (i.e. $\phi \in \partial\mathscr{R}_r(\mathscr{H})$), it remains unchanged ($\widetilde{h}_k = 0$) so that $\phi + th$ remains on the boundary. Finally, if $G(\phi)$ is singular (i.e. one $\widetilde{\psi}_k = 0$) the choice of h is not unique. This corresponds to the fact that in this case the matrix U in the polar decomposition is not unique (see [10]).

Theorem 2.9 *Let $\mathscr{H}$ be a Hilbert space and $r \leq \dim\mathscr{H}$. A weakly sequentially continuous functional $F : \mathscr{H}^r \to \mathbb{R}$ attains a maximum value on the relaxed Stiefel manifold $\mathscr{R}_r(\mathscr{H}) = \{\phi \in \mathscr{H}^r \mid G(\phi) \preceq I\}$ at a point ϕ^*. If F is convex and not constant, the maximizer ϕ^* belongs to the boundary $\partial\mathscr{R}_r(\mathscr{H})$. Moreover, if $G(\phi^*) \neq I$ there exists a $\psi \in \mathscr{H}^r$ with $G(\psi) = I$ such that F is constant on the line segment between ϕ^* and ψ, which then completely belongs to $\partial\mathscr{R}_r(\mathscr{H})$, i.e., ψ is also a maximizer. In*

[5] If the spaces are real, then a real version of the polar composition has to be considered, i.e., U orthogonal. This remark applies to all related points in the text where we might not have been precise enough.

[6] The point is that we do not want to involve rotations or reflections to diagonalize AA^* since the functional under consideration might not be invariant under such transformations. If this would be case, we could perform a unitary transformation, but in general we shall only exploit the convexity.

[7] If we represent ϕ by the matrix A from the proof and set $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, then $G(\phi) = AA^* = V^*\Lambda V$ and the here constructed h is represented by the matrix $V^*(\Lambda^{-1/2}VA - VA) = V^*\Lambda^{-1/2}VA - A = P^{-1}A - A = U - A$, which is the same as in the proof.

particular, if F is strictly convex, every maximizer satisfies $G(\phi^*) = I$, that is, belongs to the Stiefel manifold $\mathscr{S}_r(\mathscr{H})$.

Corollary 2.10 *Problem* (2.3),

$$F(\phi) = \ max, \quad \phi \in \mathscr{S}_r(\mathscr{H}),$$

has a solution if F is weakly sequentially continuous and convex on $\mathscr{H}^r$.

In general it is sufficient that F is weakly sequentially upper semicontinuous.

Proof of Theorem 2.9 The Lemmata 2.4 and 2.6 prove the existence of a maximizer ϕ^* in $\mathscr{R}_r(\mathscr{H})$, whereas the Lemmata 2.3 and 2.7 show that it belongs to the boundary $\partial\mathscr{R}_r(\mathscr{H})$ in the case that F is convex.
Suppose $G(\phi^*) \neq I$. Then Lemma 2.8 gives an $h \in \mathscr{H}^r$ such that the line segment between $\phi^* - h$ and $\phi^* + h$ completely belongs to $\partial\mathscr{R}_r(\mathscr{H})$ and $G(\phi^* + h) = I$. Therefore we set $\psi = \phi^* + h$. The element ϕ^* lies exactly in the middle of the line segment and maximizes the convex functional F. Lemma 2.7 applied to the line segment shows that this is only possible if F is constant on this segment. If F is strictly convex we get a contradiction, so $G(\phi^*) = I$. □

Remark 2.11 We are not able to give a uniqueness result and there seems to be no reason why a solution should be unique. Even the matrix U which appears in the proof of Lemma 2.8 is unique only if the matrix $G(\phi^*)$ has full rank, that is, if $\phi_1^*, \ldots, \phi_r^*$ are linearly independent. An answer to the question of uniqueness requires more a-priori information on the functional F.

2.3 Multivariate problems in products of different spaces

We extend our result to problems which are formulated in Cartesian products of different spaces. Let N Hilbert spaces $\mathscr{H}_1, \ldots, \mathscr{H}_N$ and positive integers $r_1, \ldots, r_N$ with $r_n \leq \dim \mathscr{H}_n$ be given. Define the product space

$$X = \mathscr{H}_1^{r_1} \times \cdots \times \mathscr{H}_N^{r_N}$$

and denote its elements by $\Phi = (\phi^1, \ldots, \phi^N)$ where each $\phi^n \in \mathscr{H}_n^{r_n}$, i.e., $\phi_{k_n}^n \in \mathscr{H}_n$ for $k_n = 1, \ldots, r_n$. The space X is clearly a Hilbert space equipped with the canonical inner product

$$\langle \Phi, \Psi \rangle_X = \sum_{n=1}^{N} \langle \phi^n, \psi^n \rangle_{\mathscr{H}_n^{r_n}} = \sum_{n=1}^{N} \sum_{k_n=1}^{r_n} \langle \phi_{k_n}^n, \psi_{k_n}^n \rangle_{\mathscr{H}_n}.$$

We now formulate our main theorem from which all our later applications will follow.

Theorem 2.12 *Consider the space X defined above. Let $F(\Phi) = F(\phi^1, \ldots, \phi^N)$ be a weakly sequentially continuous functional from X to $\mathbb{R}$ that is convex in each variable ϕ^n, $n = 1, \ldots, N$. Then the problem*

$$F(\Phi) = \ max, \quad \phi^1 \in \mathscr{S}_{r_1}(\mathscr{H}_1)\,,\ \ldots\ ,\ \phi^N \in \mathscr{S}_{r_N}(\mathscr{H}_N)$$

has a solution.

Proof The manifold[8]

$$\mathscr{R} = \{\Phi \in \mathscr{H}_1^{r_1} \times \cdots \times \mathscr{H}_N^{r_N} \mid G(\phi^1) \preceq I, \ldots, G(\phi^N) \preceq I\}$$

is weakly sequentially compact since it is clearly bounded and weakly sequentially closed as the Cartesian product of weakly sequentially closed sets (see Lemma 2.4). So a maximizer $\Phi = (\phi^1, \ldots, \phi^N)$ of F exists over $\mathscr{R}$. If we fix $\phi^2, \ldots, \phi^N$, then ϕ^1 is a maximizer of the problem

$$\widetilde{F}(\phi^1) = F(\phi^1, \phi^2, \ldots, \phi^N) = \max, \quad G(\phi^1) \preceq I$$

in $\mathscr{H}_1^{r_1}$. Assume $G(\phi^1) \neq I$. Since the functional $\widetilde{F}$ is convex, by Theorem 2.9 there is another maximizer $\widetilde{\phi}^1$ which satisfies $G(\widetilde{\phi}^1) = I$. With the same argument we can replace $\phi^2, \ldots, \phi^N$ if necessary and the theorem is proved. □

2.4 Least square approximation on Stiefel manifolds

Let us illustrate by a simple example how the previous results can be used in the context of least square approximation. Consequences for the well-posedness of best orthogonal low-rank tensor approximations are presented in the next section.

Let $\mathscr{H}$ be a Hilbert space, $r \in \mathbb{N}$ with $r \leq \dim \mathscr{H}$, and let $\psi \in \mathscr{H}^r$ be given. Consider the problem of least square approximation on the Stiefel manifold:

$$\|\psi - \phi\|_{\mathscr{H}^r} = \min, \quad G(\phi) = I. \tag{2.4}$$

We can take the square of the norm and obtain

$$\|\psi\|_{\mathscr{H}^r}^2 - 2\,\mathrm{Re}\langle \psi, \phi \rangle_{\mathscr{H}^r} + \|\phi\|_{\mathscr{H}^r}^2 = \min, \quad G(\phi) = I.$$

Since for such $\phi \in \mathscr{S}_r(\mathscr{H})$ we have $\|\phi\|_{\mathscr{H}^r}^2 = r$, the problem is equivalent to

$$F(\phi) = \mathrm{Re}\langle \psi, \phi \rangle_{\mathscr{H}^r} = \max, \quad \phi \in \mathscr{S}_r(\mathscr{H}). \tag{2.5}$$

The functional F is clearly convex on $\mathscr{H}^r$ (even linear) and weakly sequentially continuous, since, by definition and Riesz Representation Theorem, a sequence $(\phi_n)_n$ converges weakly to ϕ iff $\langle \psi, \phi_n \rangle_{\mathscr{H}^r} \to \langle \psi, \phi \rangle_{\mathscr{H}^r}$ for all $\psi \in \mathscr{H}^r$. With Corollary 2.10 we conclude that (2.5) and with that also (2.4) has a solution. The fact that we are dealing with a Hilbert space entered twice crucially: first in the reformulation (2.5) and second in the application of the Riesz Representation Theorem.

[8] Note that I in $G(\phi^n) \preceq I$ is the identity matrix of dimension r_n, which can be different for each $n = 1, \ldots, N$. The simplified notation was chosen to improve the readability of the text.

3 Application to orthogonal tensor rank approximation

The considerations of the previous section revealed some interesting properties of Stiefel manifolds. They can be used to prove the existence of of best low-rank tensor product approximations which consist of either *orthogonal* or *bounded* rank-one tensors. To our knowledge the well posedness of such approximations in spaces of infinite dimension has not been shown before. Our main goal here always lies in the reformulation of minimization problems into convex maximization problems on Stiefel manifolds for which Theorem 2.12 applies, but we shall present the necessary tools and notions first.

3.1 Tensor products

We present here an abstract construction of the tensor product of two vector spaces which is taken from [21]. Other constructions are possible, but they are all unique up to isomorphism. We will point out that the tensor product of spaces of functions (like $L^2(\mathbb{R}^d)$) can easily be realized as pointwise multiplication.

Let V and W be two vector spaces. Here and in the following it is always assumed that the spaces are defined over the same field $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ and will not be explicitely mentioned anymore. Define the vector space of *formal linear combinations*

$$\mathscr{F}(V,W) := \left\{ \sum_{i=1}^{m} c_i(x_i,y_i) \,\middle|\, x_i \in V,\ y_i \in W,\ c_i \in \mathbb{F},\ m \in \mathbb{N} \right\}$$

and its subspace $\mathscr{N}(V,W)$, which contains all combinations of the form

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j (x_i,y_j) - \left(\sum_{i=1}^{m} a_i x_i \,,\, \sum_{j=1}^{n} b_j y_j \right),$$

where $x_i \in V$, $y_j \in W$, $a_i, b_j \in \mathbb{F}$ and $m,n \in \mathbb{N}$.

Definition 3.1 The quotient space

$$V \otimes_a W := \mathscr{F}(V,W)/\mathscr{N}(V,W)$$

is called the *algebraic tensor product of V and W*. The equivalence class generated by (x,y) is denoted by $x \otimes y$. We call x and y the *directions* of the *tensor product $x \otimes y$*.

The equivalence classes are generated in a way, that $(x,y) \mapsto x \otimes y$ defines a bilinear mapping on $V \times W$ and any ambiguity in the representation of the formal linear combinations (when regarded as bilinear) is removed. The mapping $\otimes$ has all properties one would expect a multiplication to have (except commutativity).

In the following we will focus on the important case of Hilbert spaces. The appearing inner products are considered to be linear in the first argument and conjugate linear in the second if $\mathbb{F} = \mathbb{C}$.

Lemma 3.2 *If $\mathscr{H}_1$ and $\mathscr{H}_2$ are Hilbert spaces over $\mathbb{F}$ with inner products $\langle\cdot,\cdot\rangle_{\mathscr{H}_1}$ and $\langle\cdot,\cdot\rangle_{\mathscr{H}_2}$ respectively, then*

$$\left\langle \sum_{i=1}^{m} a_i(x_i \otimes y_i)\,,\, \sum_{j=1}^{n} b_j(x'_j \otimes y'_j) \right\rangle_{\mathscr{H}_1 \otimes \mathscr{H}_2} := \sum_{i=1}^{m}\sum_{j=1}^{n} a_i \overline{b_j}\, \langle x_i, x'_j\rangle_{\mathscr{H}_1} \langle y_i, y'_j\rangle_{\mathscr{H}_2} \tag{3.1}$$

defines an inner product in $\mathscr{H}_1 \otimes_a \mathscr{H}_2$.

The main property of this inner product is summarized in the formula

$$\langle x_1 \otimes y_1, x_2 \otimes y_2\rangle_{\mathscr{H}_1 \otimes \mathscr{H}_2} = \langle x_1, x_2\rangle_{\mathscr{H}_1} \cdot \langle y_1, y_2\rangle_{\mathscr{H}_2} \tag{3.2}$$

which is crucial for our later analysis. The space $\mathscr{H}_1 \otimes_a \mathscr{H}_2$ is not complete with respect to the norm induced by the product (3.1), it is only a pre-Hilbert space (see the example in Section 3.4).

Definition 3.3 The completion of $\mathscr{H}_1 \otimes_a \mathscr{H}_2$ with respect to the norm $\|\cdot\|_{\mathscr{H}_1 \otimes \mathscr{H}_2}$ induced by (3.1) is called the *complete tensor product of $\mathscr{H}_1$ and $\mathscr{H}_2$* and is denoted by $\mathscr{H}_1 \otimes \mathscr{H}_2$. Its elements are called *tensors*.

It is important to note that we will refer to the complete tensor product always as a normed spaced equipped with the norm $\|\cdot\|_{\mathscr{H}_1 \otimes \mathscr{H}_2}$ induced by definition (3.1) and not any other one. A completion of the algebraic tensor product with respect to another norm may result in a totally different space.

A proof of the next result is given in [21].

Lemma 3.4 *Let $\{\xi_1^1, \xi_2^1, \dots\}$ and $\{\xi_1^2, \xi_2^2, \dots\}$ be orthonormal basis sets of $\mathscr{H}_1$ and $\mathscr{H}_2$ respectively. Then*

$$\{\xi_i^1 \otimes \xi_j^2\}_{i,j}$$

is an orthonormal basis set of $\mathscr{H}_1 \otimes \mathscr{H}_2$.

A generalization to products of more than two spaces is straightforward. The N-fold tensor product of Hilbert spaces $\mathscr{H}_1, \dots, \mathscr{H}_N$ is denoted by $\bigotimes_{n=1}^N \mathscr{H}_n$. The number N is then called the *order* of the tensors under consideration.

For a vector $\phi = (\phi^1, \dots, \phi^N)$ with elements $\phi^n \in \mathscr{H}_n$, $n = 1, \dots, N$, we introduce the useful abbreviation

$$T(\phi) = T(\phi^1, \dots, \phi^N) = \phi^1 \otimes \cdots \otimes \phi^N.$$

We regard T as a mapping from $\mathscr{H}_1 \times \cdots \times \mathscr{H}_N$ to $\bigotimes_{n=1}^N \mathscr{H}_n$ and call $\phi^1, \dots, \phi^N$ the *directions* of the tensor $T(\phi)$. A crucial technical result that we will exploit concerns the weak sequential continuity of T with respect to the weak topology.

Lemma 3.5 *Let $\mathscr{H}_1, \dots, \mathscr{H}_N$ be Hilbert spaces and $(\phi_m) = ((\phi_m^1, \dots, \phi_m^N))$ a weakly convergent sequence in $\mathscr{H}_1 \times \cdots \times \mathscr{H}_N$ with limit point $\phi = (\phi^1, \dots, \phi^N)$. Then the sequence $(T(\phi_m))$ is weakly convergent in $\bigotimes_{n=1}^N \mathscr{H}_n$ with limit point $T(\phi)$, i.e.,*

$$\langle f, T(\phi_m)\rangle_{\otimes \mathscr{H}_n} \to \langle f, T(\phi)\rangle_{\otimes \mathscr{H}_n} \tag{3.3}$$

for every $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ as $m \to \infty$.

In other words this means that $\phi \mapsto \langle f, T(\phi)\rangle_{\otimes \mathscr{H}_n}$ is a weakly sequentially continuous mapping from $\mathscr{H}_1 \times \cdots \times \mathscr{H}_N$ to $\mathbb{F}$ for *every* choice of f.

Proof First of all notice that if ϕ_m converges weakly to ϕ in $\mathscr{H}_1 \times \cdots \times \mathscr{H}_N$, then ϕ_m^n converges weakly to ϕ^n in $\mathscr{H}_n$ for every $n = 1, \ldots, N$. This implies (see e.g. [15]) that the sequences (ϕ_m^n) are bounded in $\mathscr{H}_n$ and thus $(T(\phi_m))$ is bounded in $\bigotimes_{n=1}^N \mathscr{H}_n$. Now suppose that f is of the form $f = T(\psi) = \psi^1 \otimes \cdots \otimes \psi^N$, then clearly by (3.2)

$$\langle f, T(\phi_m)\rangle_{\otimes \mathscr{H}_n} = \langle T(\psi), T(\phi_m)\rangle_{\otimes \mathscr{H}_n} = \prod_{n=1}^N \langle \psi^n, \phi_m^n\rangle_{\mathscr{H}_n}$$
$$\to \prod_{n=1}^N \langle \psi^n, \phi^n\rangle_{\mathscr{H}_n} = \langle f, T(\phi)\rangle_{\otimes \mathscr{H}_n}$$

as $m \to \infty$. Since the span of tensor products $T(\psi)$ is the algebraic tensor product of the spaces $\mathscr{H}_1, \ldots, \mathscr{H}_N$, which is dense in $\bigotimes_{n=1}^N \mathscr{H}_n$, and since the sequence $(T(\phi_m))$ is bounded, the convergence (3.3) can be obtained for every $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ by a density argument, for which we refer to [15, p. 90]. □

3.2 Example: finite-dimensional spaces

If $\mathscr{H}_1 = \mathbb{R}^{d_1}$ and $\mathscr{H}_2 = \mathbb{R}^{d_2}$ with $d_1, d_2 \in \mathbb{N}$, then $\mathscr{H}_1 \otimes \mathscr{H}_2$ can be identified with the matrix space $\mathbb{R}^{d_1, d_2}$ via the usual matrix multiplication

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^T. \tag{3.4}$$

The inner product defined by (3.2) is the Frobenius inner product.

3.3 Example: L^2 spaces

As an important infinite-dimensional example we consider the function spaces $\mathscr{H}_1 = L^2(\mathbb{R}^{d_1})$ and $\mathscr{H}_2 = L^2(\mathbb{R}^{d_2})$, where d_1, d_2 are given dimensions. The L^2 spaces possess a natural tensor product structure: the tensor product $L^2(\mathbb{R}^{d_1}) \otimes L^2(\mathbb{R}^{d_2})$ can be identified with the space $L^2(\mathbb{R}^{d_1+d_2})$ by setting

$$(\phi \otimes \psi)(x,y) = \phi(x) \cdot \psi(y) \in L^2(\mathbb{R}^{d_1+d_2}) \tag{3.5}$$

for $\phi \in L^2(\mathbb{R}^{d_1}), \psi \in L^2(\mathbb{R}^{d_2})$. So the abstract construction of the tensor product presented above can be replaced by a simple pointwise multiplication in this setting. This of course has the same algebraic properties like the tensor product, but the more important observation is that $L^2(\mathbb{R}^{d_1+d_2})$ indeed has the correct inner product in the sense of Lemma 3.2. This is due to Fubini's Theorem:

$$\begin{aligned}\langle \phi_1 \otimes \psi_1, \phi_2 \otimes \psi_2 \rangle_{L^2(\mathbb{R}^{d_1+d_2})} &= \int \phi_1(x)\psi_1(y)\overline{\phi_2(x)\psi_2(y)}\,dxdy \\ &= \int \phi_1(x)\overline{\phi_2(x)}\,dx \cdot \int \phi_2(y)\overline{\psi_2(y)}\,dy \\ &= \langle \phi_1, \phi_2 \rangle_{L^2(\mathbb{R}^{d_1})} \cdot \langle \psi_1, \psi_2 \rangle_{L^2(\mathbb{R}^{d_2})}.\end{aligned}$$

Hence the identity $L^2(\mathbb{R}^{d_1}) \otimes L^2(\mathbb{R}^{d_2}) = L^2(\mathbb{R}^{d_1+d_2})$ is justified by the density of the span of functions of the form (3.5) in $L^2(\mathbb{R}^{d_1+d_2})$, which is a property of the Lebesgue measure itself.

3.4 Rank and Orthogonal Rank

We introduce several notions of tensor rank.

Definition 3.6 Let $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ be a tensor from the complete tensor product of Hilbert spaces $\mathscr{H}_1, \dots, \mathscr{H}_N$. The *canonical rank* of f is the minimal number $r \in \mathbb{N} \cup \{+\infty\}$ of summands that is needed to represent f as a linear combination of tensor products $f = \sum_{k=1}^r T(\phi_k)$, $\phi_k \in \mathscr{H}_1 \times \cdots \times \mathscr{H}_N$. This is then called a *canonical decomposition* of f. One writes $\operatorname{rank} f = r$.

This concept of tensor rank is in a certain sense the most general one can introduce. It is clear that all tensors in the algebraic tensor product space of $\mathscr{H}_1, \dots, \mathscr{H}_N$ have finite rank and that the completion process in Definition 3.3 adds tensors with infinite rank. As an example consider the tensor $g = \sum_{k=1}^\infty 2^{-k} \xi_k^1 \otimes \xi_k^2$ with $\{\xi_1^1, \xi_2^1, \dots\}$, $\{\xi_1^2, \xi_2^2, \dots\}$ being orthonormal basis sets of the Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ respectively. Then clearly $g \in \mathscr{H}_1 \otimes \mathscr{H}_2$ since it is the limit point of the the sequence $g_m = \sum_{k=1}^m 2^{-k} \xi_k^1 \otimes \xi_k^2$, but it cannot be represented by a finite linear combination of tensor products. This can be shown by some elementary comparison of coefficients, but it may be more enlightening to identify each tensor $f = \sum_{k=1}^r \phi_k^1 \otimes \phi_k^2$ ($r \in \mathbb{N} \cup \{+\infty\}$) with the linear mapping $A_f \colon \mathscr{H}_2 \to \mathscr{H}_1$ given by

$$A_f x = \sum_{k=1}^r \langle x, \phi_k^2 \rangle_{\mathscr{H}_2} \phi_k^1,$$

which does not depend on the representation of f. Note for instance that if this is done in the space $\ell^2 \otimes \ell^2$ then A_f can be seen as an infinite-dimensional matrix[9]. If f is a tensor of finite rank the range of A_f has finite dimension. Since in our above example the operator A_g spans the whole space $\mathscr{H}_1$ (it corresponds to an infinite-dimensional diagonal matrix with non-vanishing diagonal elements), it becomes clear that it can not be represented by a tensor of finite rank.

It is important to notice here that for every $f \in \mathscr{H}_1 \otimes \mathscr{H}_2$ the operator A_f is compact and therefore can be written in the form

$$A_f x = \sum_{k=1}^{\operatorname{rank} f} \sigma_k \langle x, v_k \rangle_{\mathscr{H}_2} u_k,$$

or the corresponding tensor as

$$f = \sum_{k=1}^{\operatorname{rank} f} \sigma_k u_k \otimes v_k,$$

[9] In finite dimension this is exactly the matrix which results from (3.4).

with rank $f \in \mathbb{N} \cup \{+\infty\}$ and positive numbers σ_k ("singular values") for certain orthonormal basis sets $\{u_1, u_2, \dots\}$ and $\{v_1, v_2, \dots\}$, see e.g. [5]. This generalizes the singular value decomposition of a matrix to infinite dimension, and shows that for tensors of order 2, the best low-rank approximation problem can be (theoretically) solved by truncation of this expansion, which even gives a complete orthogonal low-rank approximation (see the definition below). Apart from easy "hyper-diagonal" tensors this does not generalize to higher orders $N > 2$. Moreover, it is a hard task to determine the rank of a given tensor [9]. Note that a tensor given in the form

$$f = \sum_{k=1}^{r} \alpha_k T(\phi_k) \tag{3.6}$$

only has a rank less than or equal to r in general, since the representation might not be optimal.

Next we present definitions of *orthogonal rank* following Kolda [12]. The main idea is to generalize the singular value decomposition to tensors of higher order. This can be done in several different ways, none of them completely satisfying [12, 13]. Let $\phi_1 = (\phi_1^1, \dots, \phi_1^N)$ and $\phi_2 = (\phi_2^1, \dots, \phi_2^N) \in \mathscr{H}_1 \times \dots \times \mathscr{H}_N$. It is natural to call the tensor products $T(\phi_1)$ and $T(\phi_2)$ orthogonal iff

$$\langle T(\phi_1), T(\phi_2)\rangle_{\otimes \mathscr{H}_n} = \prod_{n=1}^{N} \langle \phi_1^n, \phi_2^n \rangle_{\mathscr{H}_n} = 0.$$

Observe that this is already the case if at least one pair ϕ_1^n, ϕ_2^n is orthogonal in its space $\mathscr{H}_n$. The aim of the next definition is to classify orthogonal tensor products more precisely.

Definition 3.7 Let $\mathscr{H}_1, \dots, \mathscr{H}_N$ be Hilbert spaces. Two tensor products $T(\phi_1)$ and $T(\phi_2)$ are called

(i) *orthogonal*, iff $\langle T(\phi_1), T(\phi_2)\rangle_{\otimes \mathscr{H}_n} = 0$,
(ii) *strongly orthogonal*, iff they are orthogonal and $\langle \phi_1^n, \phi_2^n \rangle_{\mathscr{H}_n} = 0$ or $\phi_1^n = \lambda \phi_2^n$ ($\lambda \in \mathbb{F}$) for $n = 1, \dots, N$,
(iii) *completely orthogonal*, iff $\langle \phi_1^n, \phi_2^n \rangle_{\mathscr{H}_n} = 0$ for $n = 1, \dots, N$.

Complete orthogonality implies strong orthogonality implies orthogonality. The specific types of orthogonality lead to corresponding notions of orthogonal rank.

Definition 3.8 Let $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ where $\mathscr{H}_1, \dots, \mathscr{H}_N$ are Hilbert spaces. The *orthogonal rank* of f is the minimal number $r \in \mathbb{N} \cup \{+\infty\}$ that is needed to write f as a linear combination $f = \sum_{k=1}^r T(\phi_k)$ with the $T(\phi_k)$ being mutually orthogonal. We write $\operatorname{rank}_\perp f = r$.
The *strong orthogonal rank* ($\operatorname{rank}_{\perp_s} f$) and *complete orthogonal rank* ($\operatorname{rank}_{\perp_c} f$) are defined in the same way, but with the $T(\phi_k)$ being mutually strongly orthogonal and completely orthogonal, respectively.

The orthogonal rank and the strong orthogonal rank are well defined terms, since at least the Fourier expansion of f into a strong orthogonal basis as in Lemma 3.4

has the desired form. With the complete orthogonal rank one has to be cautious: for order $N > 2$ not every tensor can be written as a sum of complete orthogonal tensor products, not even as an infinite series[10]. The complete orthogonal rank makes no sense in these cases. However, if $\operatorname{rank}_{\perp_c} f$ is defined for a tensor f it is clear that

$$\operatorname{rank}_{\perp} f \leq \operatorname{rank}_{\perp_s} f \leq \operatorname{rank}_{\perp_c} f.$$

For each $n = 1, \dots, N$ the elements $\phi_1^n, \dots, \phi_r^n$ in (3.6) span some r_n-dimensional subspace V_n of $\mathscr{H}_n$ with $r_n \in \mathbb{N} \cup \{+\infty\}$, $r_n \leq r$. We say that the *n-rank* of f is less than or equal to r_n and write $\operatorname{rank}_n f \leq r_n$. The values r_n are upper bounds for the number of dimensions needed for each direction n to represent f. The n-rank itself is defined as the least number to do so (see Remark 3.9 below). After choosing orthonormal bases for the spaces V_n every tensor can be written in *Tucker* $(r_1, \dots, r_N)$*-format*

$$f = \sum_{k_1=1}^{r_1} \cdots \sum_{k_N=1}^{r_N} \alpha_{k_1 \dots k_N} T(\phi_{k_1}^1, \dots, \phi_{k_N}^N) =: \sum_{\mathbf{k}} \alpha_{\mathbf{k}} T(\phi_{\mathbf{k}}) \tag{3.7}$$

with

$$G(\phi^1) = I \ , \ \dots \ , \ G(\phi^N) = I,$$

where $G(\phi^n)$ denotes the Gram matrix of dimension r_n. The sum on the right side of (3.7) will be used as an abbreviation for the expansion, that is, it runs over all multi-indices $\mathbf{k} = (k_1, \dots, k_N)$ with $1 \leq k_n \leq r_n$ for $n = 1, \dots, N$ and $T(\phi_{\mathbf{k}})$ means $T(\phi_{k_1}^1, \dots, \phi_{k_N}^N)$. The tensor $\alpha = (\alpha_{k_1 \dots k_N}) \in \bigotimes_{n=1}^N \mathbb{F}^{r_n}$ is called *core tensor*. The representation (3.7) is also called *subspace representation*. It is not unique, but shows that the strong orthogonal rank of f is at most $\prod_{n=1}^N r_n$.

Remark 3.9 If every r_n is chosen in a minimal way then $\operatorname{rank}_n f = r_n$ and one says that f has *Tucker rank* or *subspace rank* $(r_1, \dots, r_N)$. Indeed such a choice is possible and unique for every direction n, since the least number of dimensions needed in one direction is independent from all the others. For the n-ranks and the Tucker decomposition see also [14, 16, 17].

A tensor f with strong orthogonal rank r can be written in Tucker $(r, r, \dots, r)$-format with exactly r non-vanishing core coefficients $\alpha_{\mathbf{k}}$ by definition. In the special case of a complete orthogonal rank-r tensor the non-vanishing coefficients are precisely the ones on the diagonal, i.e., those with $\mathbf{k} = (k, k, \dots, k)$, $k = 1, \dots, r$. One speaks of a *(hyper)diagonal core tensor* then.

Let us emphasize that not only orthogonal, but also strong orthogonal, complete orthogonal and Tucker low-rank decompositions and -approximations of a given tensor f can be of large interest in practice, since they provide optimal subspaces in the direction spaces, in which one has to work to handle the complexity of high-order tensors. We will not pursue the topic of how to calculate such approximations. We will only prove that best approximations indeed exist. For the case $N = 2$ the canonical

[10] A counterexample is a tensor of the form $u \otimes v_1 \otimes w_1 + u \otimes v_2 \otimes w_2$ with v_1 orthogonal to v_2 and w_1 orthogonal to w_2.

rank and the three types of orthogonal rank coincide (consider the complete orthogonal singular value decomposition), but it is also not our purpose here to study their relationship in general. However, it is worth to mention that our results include the existence of a best canonical rank-one approximation to tensors of any order, because for $r = 1$ all types of rank also coincide.

Before we present specific results, we sketch out how orthogonality constraints can contribute to the well-posedness of a low-rank approximation problem. Given a tensor f, consider the best approximation

$$\left\| f - \sum_{k=1}^{r} \alpha_k T(\phi_k) \right\|_{\otimes \mathscr{H}_n} = \min, \quad \langle T(\phi_k), T(\phi_l) \rangle_{\otimes \mathscr{H}_n} = \delta_{kl} \tag{3.8}$$

and $\alpha_k \in \mathbb{F}$. For fixed $T(\phi_1), \dots, T(\phi_r)$ the square of the error is minimized only by the projection with $\alpha_k = \langle f, T(\phi_k) \rangle_{\otimes \mathscr{H}_n}$ and takes the value

$$\|f\|^2_{\otimes \mathscr{H}_n} - \sum_{k=1}^{r} |\langle f, T(\phi_k) \rangle_{\otimes \mathscr{H}_n}|^2.$$

Thus, the minimization problem (3.8) is equivalent to the maximization problem

$$\sum_{k=1}^{r} |\langle f, T(\phi_k) \rangle_{\otimes \mathscr{H}_n}|^2 = \max, \quad \langle T(\phi_k), T(\phi_l) \rangle_{\otimes \mathscr{H}_n} = \delta_{kl}. \tag{3.9}$$

The possibility to pass from (3.8) to (3.9) is crucial because in (3.9) we have to maximize a functional which is convex in each entry ϕ^n and weakly sequentially continuous by Lemma 3.5. So orthogonality constraints can be seen as a technical requirement for our method of proof. They are the essential reasons why the approximation problems considered in the next sections are indeed well-posed, in contrast to unconstrained canonical rank-r approximation problems. Also the consideration of Hilbert spaces is crucial for the reformulation (3.9).

3.5 Ill-Posedness of Canonical Low-Rank Approximation

For the remainder of this article we fix N Hilbert spaces $\mathscr{H}_1, \dots, \mathscr{H}_N$ over the same field $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$), integers $r, r_1, \dots, r_N$ with $0 < r_n \le \dim \mathscr{H}_n$ and a tensor $f \in \bigotimes_{n=1}^N \mathscr{H}_n$. As mentioned in the introduction it was shown by De Silva and Lim [19] that the approximation problem

$$\|f - g\|_{\otimes \mathscr{H}_n} = \min, \quad \operatorname{rank} g \le r \tag{3.10}$$

is ill-posed in many cases, even if the spaces $\mathscr{H}_1, \dots, \mathscr{H}_N$ are finite-dimensional. Problem (3.10) is called the *canonical rank-r approximation* of f and may be written in the form

$$\left\| f - \sum_{k=1}^{r} \alpha_k T(\phi_k) \right\|_{\otimes \mathscr{H}_n} = \min, \quad \|\phi_k^n\|_{\mathscr{H}_n} = 1, \quad \alpha_k \in \mathbb{F}.$$

De Silva and Lim give a counter-example of a sequence of rank-2 tensors of order 3 converging to a rank-3 tensor of order 3. So the infimum 0 of (3.10) is not attained for their choice of f and r. A closer look on the cited counter-example shows that the coefficients α_k in the approximating sequence are unbounded. This is due to an asymptotical degeneracy of the spanned subspaces $\text{span}\{T(\phi_1),\ldots,T(\phi_N)\}$. The tensors $T(\phi_k)$ in their approximating sequence become "almost linearly dependent", that is, their distance in the sense of subspaces tends to zero.

The need for additional constraints becomes evident. The two natural approaches are either to bound the coefficients α_k by a fixed constant or to impose angle constraints to the subspaces spanned by the $T(\phi_k)$ (we consider orthogonality). Both approaches lead to well-posed problems even in infinite dimension as we will show in the next sections. We start with the orthogonal approximations and consider bounded coefficients at the end. In all cases it is our aim to reformulate the problem as a convex maximization on a Stiefel manifold and to apply Theorem 2.12.

3.6 Best Low-Rank Approximation with Orthogonal Directions

We begin with the problem of best Tucker $(r_1,\ldots,r_N)$-format approximation, also known as optimal subspace approximation:

$$\Big\| f - \sum_{\mathbf{k}} \alpha_{\mathbf{k}} T(\phi_{\mathbf{k}}) \Big\|_{\otimes \mathscr{H}_n} = \min. \tag{3.11}$$

Here the sum runs over all multi-indices $\mathbf{k} = (k_1,\ldots,k_N)$ with $1 \le k_n \le r_n$ for $n = 1,\ldots,N$, see (3.7). It is quite remarkable that this problem, as we will show, has a solution, since no constraints appear at all. The trick is that there is no loss in generality to impose the additional constraints[11]

$$G(\phi^1) = I \,,\ \ldots\ ,\ G(\phi^N) = I$$

on the Gram matrices $G(\phi^n) \in \mathbb{C}^{r_n,r_n}$. If these conditions are satisfied the optimal choice of $\alpha_{\mathbf{k}}$ is given by projection as $\alpha_{\mathbf{k}} = \langle f, T(\phi_{\mathbf{k}})\rangle_{\otimes \mathscr{H}_n}$. Hence (3.11) is equivalent to

$$F(\Phi) = \sum_{\mathbf{k}} |\langle f, T(\phi_{\mathbf{k}})\rangle_{\otimes \mathscr{H}_n}|^2 = \max, \quad \phi^1 \in \mathscr{S}_{r_1}(\mathscr{H}_1) \,,\ \ldots\ ,\ \phi^N \in \mathscr{S}_{r_N}(\mathscr{H}_N).$$

Since the functional F is weakly sequentially continuous by Lemma 3.5 and convex in each entry ϕ^n the existence of a solution to this problem follows from Theorem 2.12.

In general, we can fix any subset $\mathscr{I} \subseteq \{\mathbf{k} = (k_1,\ldots,k_N) \mid 1 \le k_n \le r_n\}$ and consider the problem

$$\Big\| f - \sum_{\mathbf{k}\in\mathscr{I}} \alpha_{\mathbf{k}} T(\phi_{\mathbf{k}}) \Big\|_{\otimes \mathscr{H}_n} = \min, \quad G(\phi^1) = I \,,\ \ldots\ ,\ G(\phi^N) = I, \tag{3.12}$$

[11] This invariance with respect to the choice of basis shows that the best Tucker format approximation is a problem on Grassman manifolds rather than Stiefel manifolds, but it is better for us to formulate it on the latter ones, accepting additional redundancy. The true unknowns in this problem are the subspaces V_n spanned by the $\{\phi^n_1,\ldots,\phi^n_{r_n}\}$, $n = 1,\ldots,N$.

that is, we can impose $\alpha_{\mathbf{k}} = 0$ for $\mathbf{k} \notin \mathscr{I}$ in (3.11). This is then equivalent to

$$F(\Phi) = \sum_{\mathbf{k}\in\mathscr{I}} |\langle f, T(\phi_{\mathbf{k}})\rangle_{\otimes\mathscr{H}_n}|^2 = \max, \quad \phi^1 \in \mathscr{S}_r(\mathscr{H}_1), \ldots, \phi^N \in \mathscr{S}_r(\mathscr{H}_N),$$

and has a solution by Theorem 2.12. For reference we summarize the results.

Theorem 3.10 *Problem* (3.12) *has a solution* $g = \sum_{\mathbf{k}\in\mathscr{I}} \alpha_{\mathbf{k}} T(\phi_{\mathbf{k}})$.

Corollary 3.11 (Tucker Format Approximation) *To every* $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ *there exists a best Tucker* $(r_1,\ldots,r_N)$*-format approximation, that is,* (3.11) *has a solution* $g = \sum_{\mathbf{k}} \alpha_{\mathbf{k}} T(\phi_{\mathbf{k}})$ *for every choice of* $r_1,\ldots,r_N \in \mathbb{N}$.

Remark 3.12 In general this does not mean that the solution g indeed has Tucker rank $(r_1,\ldots,r_N)$ as defined in Remark 3.9, so there remains an uncertainty concerning the choice of the r_n. If f belongs to the tensor product $\bigotimes_{n=1}^N V_n$ of finite-dimensional subspaces $V_n \subseteq \mathscr{H}_n$ then so does g. In this case the existence of g alternatively can be derived from the compactness of the underlying Stiefel manifold.

Next we consider the complete orthogonal rank-r approximation. It is the solution g of the problem

$$\|f-g\|_{\otimes\mathscr{H}_n} = \min, \quad \operatorname{rank}_{\perp_c} g \le r. \tag{3.13}$$

As we have mentioned above a tensor with complete orthogonal rank at most r can be written in the orthogonal version of the Tucker format with diagonal core tensor. So (3.13) is equivalent to (3.12) with the special choice $r_1 = r_2 = \cdots = r_N = r$ and

$$\mathscr{I} = \{(1,1,\ldots,1),(2,2,\ldots,2),\ldots,(r,r,\ldots,r)\} \subseteq \mathbb{N}^N.$$

Hence Theorem 3.10 proves the existence of a solution.

Corollary 3.13 (Complete Orthogonal Approximation) *To every* $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ *there exists a best complete orthogonal rank-r approximation, that is,* (3.13) *has a solution* $g = \sum_{k=1}^r \alpha_k T(\phi_k)$ *for every choice of* $r \in \mathbb{N}$.

As a last application we show the well-posedness of the best strong orthogonal rank-r approximation

$$\|f-g\|_{\otimes\mathscr{H}_n} = \min, \quad \operatorname{rank}_{\perp_s} g \le r. \tag{3.14}$$

A tensor with strong orthogonal rank at most r can be written in othogonal Tucker $(r,r,\ldots,r)$-format with at most r non-vanishing core coefficients, that is, (3.14) is equivalent to (3.12) with $r_1 = r_2 = \cdots = r_N = r$ and $\mathscr{I}$ varying over all subsets with $|\mathscr{I}| = r$, where $|\mathscr{I}|$ denotes the number of elements in $\mathscr{I}$. Since the number of such subsets is finite, the existence of a solution to (3.14) also follows from Theorem 3.10.

Corollary 3.14 (Strong Orthogonal Approximation) *To every* $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ *there exists a best strong orthogonal rank-r approximation, that is,* (3.14) *has a solution* $g = \sum_{k=1}^r \alpha_k T(\phi_k)$ *for every choice of* $r \in \mathbb{N}$. *If* $\operatorname{rank}_{\perp_s} f \ge r$ *then* $\operatorname{rank}_{\perp_s} g = r$.

The addition about the rank of g follows from the fact that a set of strong orthonormal tensor products is a subset of a certain strong orthonormal tensor product basis of $\bigotimes_{n=1}^N \mathscr{H}_n$ (as in Lemma 3.4). If $\operatorname{rank}_{\perp_s} f \ge r$ then f has at least r non-vanishing Fourier coefficients in every such basis. Consequently, a best strong orthogonal rank-r approximation of f has to consist of r strongly orthogonal summands.

3.7 The More General Case of Orthogonal Approximation

The more general problem of a best orthogonal rank-r approximation is

$$\|f-g\|_{\otimes \mathscr{H}_n} = \min, \quad \operatorname{rank}_\perp g \le r, \tag{3.15}$$

which explicitly reads

$$\left\| f - \sum_{k=1}^{r} \alpha_k T(\phi_k) \right\|_{\otimes \mathscr{H}_n} = \min, \quad \alpha_k \in \mathbb{F}, \quad G(\phi^1) \circ \cdots \circ G(\phi^N) = I. \tag{3.16}$$

Recall that $G(\phi^1) \circ \cdots \circ G(\phi^N)$ denotes the N-fold Hadamard product of the Gram matrices $G(\phi^1), \ldots, G(\phi^N)$, which is precisely the Gram matrix of the tensor products $T(\phi_1), \ldots, T(\phi_r)$ (see (2.2)).

To prove the existence of a solution to (3.16) we have to argue with a little bit more sophistication. First of all notice that the set $\{\Phi \mid G(\phi^1) \circ \cdots \circ G(\phi^N) = I\}$ is not bounded in $\mathscr{H}_1^r \times \cdots \times \mathscr{H}_N^r$, but that for each Φ from that set we have

$$\|T(\phi_k)\|^2_{\otimes \mathscr{H}_n} = \prod_{n=1}^{N} \|\phi_k^n\|^2_{\mathscr{H}_n} = 1,$$

and thus $T(\phi_k) = T\left(\phi_k^1/\|\phi_k^1\|_{\mathscr{H}_1}, \ldots, \phi_k^N/\|\phi_k^N\|_{\mathscr{H}_N}\right)$ for every $k = 1, \ldots, r$. So we do not loose generality by adding the constraint $\|\phi_k^n\|_{\mathscr{H}_n} = 1$ to (3.16). If we set

$$\mathscr{M} = \{\Phi \mid G(\phi^1) \circ \cdots \circ G(\phi^N) = I, \, \|\phi_k^n\|_{\mathscr{H}_n} = 1 \quad (n = 1, \ldots, N; k = 1, \ldots, r)\},$$

and argue that for $\Phi \in \mathscr{M}$ the best choice for α_k is the projection $\langle f, T(\phi_k)\rangle_{\otimes \mathscr{H}_n}$, we obtain the equivalent problem

$$F(\Phi) = \sum_{k=1}^{r} |\langle f, T(\phi_k)\rangle_{\otimes \mathscr{H}_n}|^2 = \max, \quad \Phi \in \mathscr{M}. \tag{3.17}$$

We want to show that (3.17) has a solution. To apply the techniques developed above we need to somehow dispose of the coupling of the orthonormality conditions by the Hadamard product. This, in a sense, can be done in the following way: For each $\phi^n \in \mathscr{H}_n^r$ we choose $\xi^n \in \mathscr{H}_n^r$ consisting of orthonormal elements $\xi_1^n, \ldots, \xi_r^n$ such that $\operatorname{span}\{\phi_1^n, \ldots, \phi_r^n\} \subseteq \operatorname{span}\{\xi_1^n, \ldots, \xi_r^n\}$, and a matrix $A^{(n)} = [a_{kl}^{(n)}]_{kl} \in \mathbb{C}^{r,r}$ to write

$$\phi_k^n = \sum_{l=1}^{r} a_{kl}^{(n)} \xi_l^n. \tag{3.18}$$

The constraint $\Phi \in \mathscr{M}$ then reads

$$A^{(1)}(A^{(1)})^* \circ \cdots \circ A^{(N)}(A^{(N)})^* = I,$$

$$\sum_{l=1}^{N} (a_{kl}^{(n)})^2 = 1 \quad (n = 1, \ldots, N; k = 1, \ldots, r)$$

and

$$G(\xi^1) = I\,,\,\dots\,,\,G(\xi^N) = I.$$

We may summarize the first two conditions as $h(A^{(1)},\dots,A^{(N)}) = 0$ where h is a continuous mapping from $(\mathbb{C}^{r,r})^N$ into some other finite-dimensional space. If we express F in terms of $A^{(1)},\dots,A^{(N)}$ and $\Xi = (\xi^1,\dots,\xi^N)$ we obtain the problem

$$\begin{aligned} F(\Xi,A^{(1)},\dots,A^{(N)}) &= \max, \\ h(A^{(1)},\dots,A^{(N)}) &= 0, \\ \xi^1 \in \mathscr{S}_r(\mathscr{H}_1)\,,\,\dots\,,\,\xi^N &\in \mathscr{S}_r(\mathscr{H}_N), \end{aligned} \tag{3.19}$$

which is equivalent to (3.17). The desired solution to (3.16) can be obtained from this one via (3.18). The explicit representation of F in terms of $\Xi,A^{(1)},\dots,A^{(N)}$ reads

$$F(\Xi,A^{(1)},\dots,A^{(N)}) = \sum_{k=1}^{r} \left| \sum_{\mathbf{l}} \Big(\prod_{n=1}^{N} a_{kl_n}^{(n)} \Big) \langle f, T(\xi_{\mathbf{l}}) \rangle_{\otimes \mathscr{H}_n} \right|^2$$

with $\mathbf{l} = (l_1,\dots,l_N)$ being a multi-index.

Now consider a maximizing sequence $(\Xi_m,A_m^{(1)},\dots,A_m^{(N)})$ for (3.19). The condition $h(A_m^{(1)},\dots,A_m^{(N)}) = 0$ implies that the $A_m^{(n)}$ remain bounded and therefore can be assumed to converge to limit points $A^{(1)},\dots,A^{(N)}$. Since h is continuous, these limit points clearly satisfy $h(A^{(1)},\dots,A^{(N)}) = 0$. There is thus no loss of generality to fix those limit points a priori[12] and formulate the problem as

$$\widetilde{F}(\Xi) = \max, \quad \xi^1 \in \mathscr{S}_r(\mathscr{H}_1)\,,\,\dots\,,\,\xi^N \in \mathscr{S}_r(\mathscr{H}_N) \tag{3.20}$$

with $\widetilde{F}(\Xi) = F(\Xi,A^{(1)},\dots,A^{(N)})$. It is not hard to see that the functional $\widetilde{F}$ is convex in each ξ^n and weakly sequentially continuous. Consequently, the problem admits a solution by Theorem 2.12.

Theorem 3.15 (Orthogonal Approximation) *To every $f \in \bigotimes_{n=1}^N \mathscr{H}_n$ there exists a best orthogonal rank-r approximation, that is,* (3.15) *has a solution $g = \sum_{k=1}^r \alpha_k T(\phi_k)$ for every $r \in \mathbb{N}$.*

By now we do not know whether $\operatorname{rank}_\perp g = r$ if $\operatorname{rank}_\perp f \geq r$.

3.8 The Canonical Approximation with Bounded Coefficients

A broader and even more natural approach to fix the ill-posedness of the canonical low-rank approximation as described in Section 3.5 is to keep the coefficients α_k appearing therein bounded. This has been successfully done by Espig [4] to construct a numerical method for the calculation of low-rank approximations.

---

[12] This technique came to our attention by [1]. Equivalently one can argue that the set $\{(\Xi,A^{(1)},\dots,A^{(N)}) \mid G(\xi^1) \preceq I,\dots,G(\xi^N) \preceq I, h(A^{(1)},\dots,A^{(N)}) = 0)\}$ is weakly sequentially compact.

The problem of canonical rank-r approximation with coefficients bounded by the constant $C > 0$ reads

$$\Big\| f - \sum_{k=1}^{r} \alpha_k T(\phi_k) \Big\|_{\otimes \mathscr{H}_n} = \min, \quad \|\phi_k^n\|_{\mathscr{H}_n} = 1, \quad \alpha_k \in \mathbb{F}, \quad |\alpha_k| \leq C. \qquad (3.21)$$

To prove the well-posedness of this problem we apply a similar trick as for the orthogonal approximation. An expansion

$$\phi_k^n = \sum_{l=1}^{r} a_{kl}^{(n)} \xi_l^n$$

of each ϕ_k^n into an orthonormal basis $\xi_1^n, \ldots, \xi_r^n$ leads to

$$\Big\| f - \sum_{k=1}^{r} \alpha_k \sum_{\mathbf{l}} \Big(\prod_{n=1}^{N} a_{kl_n}^{(n)} \Big) T(\xi_{\mathbf{l}}) \Big\|_{\otimes \mathscr{H}_n} = \min$$

(with multi-indices $\mathbf{l} = (l_1, \ldots, l_N)$) as a problem in α_k, $a_{kl}^{(n)}$ and $\Xi = (\xi^1, \ldots, \xi^N)$, subject to the constraints

$$|\alpha_k| \leq C, \quad \sum_{l=1}^{r} (a_{kl}^{(n)})^2 = 1, \quad G(\xi^1) = I, \ldots, G(\xi^N) = I.$$

In a minimizing sequence for this problem the bounded coefficients α_k and $a_{kl}^{(n)}$ again can be assumed to converge and therefore may be fixed a priori. If we set $\beta_{\mathbf{l}} = \sum_{k=1}^{r} \alpha_k \prod_{n=1}^{N} a_{kl_n}^{(n)}$ then also

$$\Big\| \sum_{k=1}^{r} \alpha_k T(\phi_k) \Big\|_{\otimes \mathscr{H}_n}^2 = \Big\| \sum_{\mathbf{l}} \beta_{\mathbf{l}} T(\xi_{\mathbf{l}}) \Big\|_{\otimes \mathscr{H}_n}^2 = \sum_{\mathbf{l}} \beta_{\mathbf{l}}^2$$

is fixed. Taking the square of the norm in (3.21) one obtains the equivalent problem

$$\operatorname{Re} \sum_{\mathbf{l}} \langle f, \beta_{\mathbf{l}} T(\xi_{\mathbf{l}}) \rangle_{\otimes \mathscr{H}_n} = \max, \quad \xi^1 \in \mathscr{S}_r(\mathscr{H}_1), \ldots, \xi^N \in \mathscr{S}_r(\mathscr{H}_N).$$

This is a convex maximization problem on a Stiefel manifold in each ξ^n and satisfies the assumptions of Theorem 2.12. Hence it has a solution.

Theorem 3.16 (Canonical Approximation with Bounded Coefficients) *To every $f \in \bigotimes_{n=1}^{N} \mathscr{H}_n$ and $C > 0$ there exists a best canonical rank-r approximation with coefficients bounded by C, that is,* (3.21) *has a solution $g = \sum_{k=1}^{r} \alpha_k T(\phi_k)$ for every choice of $r \in \mathbb{N}$.*

This result has already been established in [4] for tensors f of finite rank.

References

1. Le Bris, C.: A general approach for multiconfiguration methods in quantum molecular chemistry. Ann. Inst. Henri Poincaré 11 No. 4, 441-484 (1994)
2. Chinnamsetty, S.R. , Espig, M., Flad, H.-J., Khoromskij, B.N., Hackbusch. W.: Tensor Product Approximation with Optimal Rank in Quantum Chemistry. Jour. Chem. Phys. 127, 084110 (2007)
3. Edelman, A., Arias, T., Smith, S.T.: The Geometry of Algorithms with Orthogonality Constraints. SIAM J. Matrix. Anal. Appl. 20, 303-353 (1999)
4. Espig, M.: Effiziente Bestapproximation mittels Summen von Elementartensoren in hohen Dimensionen. Phd. Thesis, MPI-MIS Leipzig, Germany 2007
5. Gohberg, I.C., Krein, M.G.: Introduction to the Theory of Linear Nonselfadjoint Operators. American Mathematical Society, Providence, RI (1969)
6. Golub, G.H., Zhang, T.: Rank-One Approximation to High Order Tensors. SIAM J. Matrix Anal. Appl. 23, 534-550 (2001)
7. Greub, W.H.: Multilinear Algebra. Springer-Verlag, Berlin (1967)
8. Hackbusch, W., Khoromskij, B.N.: Tensor-product approximation to operators and functions in high dimensions. Journal of Complexity 23, 697-714 (2007)
9. Håstad, J.: Tensor Rank Is NP-Complete. Journal of Algorithms 11, 644-654 (1990)
10. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1990)
11. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1994)
12. Kolda, T.G.: Orthogonal Tensor Decompositions. Siam J. Matrix Anal. Appl. 23 No. 1, 243-255 (2001)
13. Kolda, T.G.: A counterexample to the possibility of an extension of the Eckart-Young low-rank approximation theorem for the orthogonal rank tensor decomposition. Siam J. Matrix Anal. Appl. 24 No. 3, 762-767 (2003)
14. Kolda, T.G., Bader, B.W.: Tensor Decompositions and Applications. Siam Review 51 No. 3 (2009)
15. Kolmogorov, A.N., Fomin, S.V.: Elements of the Theory of Functions and Functional Analysis. Volume 1 Metric and Normed Spaces. Graylock Press, Rochester, NY (1957)
16. De Lathauwer, L., De Moor, B., Vandewalle, J.: A Multilinear Singular Value Decomposition. Siam J. Matrix Anal. Appl. 21 No. 4, 1253-1278 (2000)
17. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the Best Rank-1 and Rank-$(r_1, r_2, \ldots, r_N)$ Approximation of High-Order Tensors. Siam J. Matrix Anal. Appl. 21 No. 4, 1324-1342 (2000)
18. Lieb, E.H., Simon, B.: The Hartree-Fock Theory for Coulomb Systems. Comm. Math. Phys. 53, 185-194 (1977)
19. De Silva, V., Lim, L.-H.: Tensor Rank and The Ill-Posedness of the Best Low-Rank Approximation Problem. Siam J. Matrix. Amal. Appl. 30 No. 3 (2008) 1084-1127
20. Temlyakov, V.N.: Greedy Algortihms and M-Term Approximation with Regard to Redundant Dictionaries. J. Approx. Theory 98, 117-145 (1999)
21. Weidmann, J.: Lineare Operatoren in Hilberträumen, Teil 1. Teubner Verlag, Stuttgart (2000)
22. Zeidler, E.: Applied Functional Analysis. Main Principles and Their Applications. Springer-Verlag, New York (1995)