Numerische
Mathematik

CrossMark

On the interconnection between the higher-order singular values of real tensors

Wolfgang Hackbusch[1] · **André Uschmajew[2]**

Abstract A higher-order tensor allows several possible matricizations (reshapes into matrices). The simultaneous decay of singular values of such matricizations has crucial implications on the low-rank approximability of the tensor via higher-order singular value decomposition. It is therefore an interesting question which simultaneous properties the singular values of different tensor matricizations actually can have, but it has not received the deserved attention so far. In this paper, preliminary investigations in this direction are conducted. While it is clear that the singular values in different matricizations cannot be prescribed completely independent from each other, numerical experiments suggest that sufficiently small, but otherwise arbitrary perturbations preserve feasibility. An alternating projection heuristic is proposed for constructing tensors with prescribed singular values (assuming their feasibility). Regarding the related problem of characterising sets of tensors having the same singular values in specified matricizations, it is noted that orthogonal equivalence under multilinear matrix multiplication is a sufficient condition for two tensors to have the same singular values in all principal, Tucker-type matricizations, but, in contrast to the matrix case, not necessary. An explicit example of this phenomenon is given.

Mathematics Subject Classification 15A18 · 15A21 · 15A69

✉ Wolfgang Hackbusch
wh@mis.mpg.de

André Uschmajew
uschmajew@ins.uni-bonn.de

[1] Max-Planck-Institut Mathematik in den Naturwissenschaften, Inselstraße 22, 04103 Leipzig, Germany

[2] Hausdorff Center for Mathematics & Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany

1 Introduction and problem statement

A space $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ of higher-order tensors is isomorphic to many different matrix spaces of the form $(\bigotimes_{j \in t} \mathbb{R}^{n_j}) \otimes (\bigotimes_{i \notin t} \mathbb{R}^{n_i})$ where $t \subsetneq \{1, \ldots, d\}$, $|t| \geq 1$. Concretely, when identifying tensors with d-dimensional arrays of coordinates with respect to an orthonormal tensor product basis, such an isomorphism is realized by reshaping the array into a matrix. The directions in t indicate the multi-indices for the rows of the resulting matrix, while the other directions are used for the columns. All these different *matricizations* (also called *unfoldings* or *reshapes* in the literature) of the tensor carry some spectral information in form of their singular value decompositions.

For subsets of t that are part of a dimension partition tree, the column spaces of the corresponding matricizations satisfy certain nestedness properties that form the basis for important subspace based low-rank tensor decompositions like the Tucker format [22], the hierarchical Tucker (HT) format [7,9], or the tensor train (TT) format [18,19]. As a by-product, the ranks r_t of the corresponding matricizations, that is, the number of nonzero singular values, are estimated as $r_t \leq r_{t_1} \cdots r_{t_s}$, where $t = t_1 \cup \cdots \cup t_s$ is a disjoint partition. In contrast, the interconnections between the singular values themselves have not been studied so far.

At first sight, the singular values of different matricizations could be considered as unnatural or artificial characteristics for tensors, as they ignore their multilinear nature. However, as it turns out, they provide crucial measures for the approximability of tensors in the aforementioned low-rank subspace formats. In the pioneering work [3] the higher-order singular value decomposition has been defined, and it has been shown how it can be practically used to obtain quasi-optimal low-rank approximations in the Tucker format with full error control. The approximation is obtained by an orthogonal projection on the tensor product of subspaces spanned by the dominating singular vectors of the corresponding matricizations in $\mathbb{R}^{n_j} \otimes (\bigotimes_{i \neq j} \mathbb{R}^{n_i})$ (i.e. corresponding to the choices $t = \{j\}$ for $j = 1, \ldots, d$). An upper bound of the squared error is then given by the sum of squares of all deleted singular values in all directions. Later, variants of such truncation procedures have been obtained for the TT format [16,17] and the HT format [7] with similar error bounds, but involving singular values of some other matricizations of the tensor.

Building on these available, quasi-optimal bounds for low-rank approximations via higher-order versions of SVD truncation, it is understandable that quite some theorems have been stated making simultaneous assumptions on the singular values of certain matricizations of a tensor. This concerns stability of low-rank ODE integrators [12,14], local convergence of optimization algorithms [20], or approximability by low-rank tensor formats [1], to name a few. Assumed properties of interest are decay rate of and gaps between the singular values, for instance. A principal task would then be to give alternative descriptions of classes of tensors satisfying such assumptions to prevent tautological results or, in worst case, void statements. But this task has turned out to be notoriously difficult. For tensors arising from function discretization, some qualitative statements about the decay of singular values can be obtained from their regularity using explicit analytic approximation techniques by tensor products of (trigonometric) polynomials or wavelets, exponential sum, or cross approximation; see [8,21] and references therein. But these qualitative implications on the decay of singular values

obtained from explicit separable approximations can rarely be made quantitatively precise, for instance, if they contain unknown constants, and also provide little insight on the actual interconnection between different matricizations.

In its purest form the question we are interested in is very simple to state. Given prescribed singular values for some matricizations (having, e.g., some favourable properties), does there exist at all a tensor having these singular values? For a matrix this is of course very easy to answer by simply constructing a diagonal matrix. For tensors it turns out to be quite difficult, and seems to depend on how many matricizations are simultaneously considered.

In the present paper we study this and related questions for the singular values related to the classical higher-order SVD, that is, the singular values of the principal, Tucker-type matricizations that separate single indices $t = \{j\}$. We call the collection of the d corresponding singular value vectors the *higher-order singular values*, see Definition 1.1 below. Compared to other systems of matricizations, this framework is the historically most important. It also appears to be the simplest, partly because it is, at least to some extent, very conveniently possible to manipulate the principal matricizations simultaneously via multilinear matrix multiplications. Yet, even in this case, the obtained results remain fragmentary and far from complete. Nevertheless, we consider them as valuable first steps toward future investigations of this important and fascinating subject. Our contributions are as follows.

- We show that not all configurations of higher-order singular values can be feasible. The proof is nonconstructive (Sect. 3.1).
- However, conducted numerical experiments suggest that the singular values for different matricizations are, except for degenerate situations, *locally* independent from each other. That is, in the neighbourhood of a tensor it is possible to slightly perturb, say, only the singular values of the first matricization, while maintaining the singular values of the other ones. This is fundamentally different from the matrix case, since the singular values of a matrix are always the same as the ones of its transpose. However, currently this remains an unproved conjecture Sect. 3.2).
- We propose the method of alternating projections as a heuristic to construct (approximately) tensors with prescribed singular values in certain matricizations (Sect. 3.3).
- The higher-order SVD (HOSVD) is a generalization of the SVD from matrices to tensors. The role of the diagonal matrix of singular values is replaced by the core tensor in the HOSVD, representing the normal form under orthogonal equivalence, and characterized by slice-wise orthogonality properties. We show manifold properties of the set of these core tensors (called HOSVD tensors) in the case of strictly decreasing and positive higher-order singular values (Sect. 2.2).
- We provide an example of two $2 \times 2 \times 2$ tensors having the same singular values in all three principal matricizations without being orthogonally equivalent (Sect. 2.4).

In this paper we consider real tensors for convenience. Although most concepts seem to generalize to the complex case, some care would be required, e.g., when switching from smooth manifolds to analytic ones.

The rest of this section is devoted to the precise statements of the considered problems. They require some amount of definitions and notation, which will be introduced first.

1.1 Preliminaries, definitions, notation

Let $d \geq 3$ and $n_1, \ldots, n_d \in \mathbb{N}$. We consider the d-fold tensor product space $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ as isomorphic to the space $\mathbb{R}^{n_1 \times \cdots \times n_d}$ of real $n_1 \times \cdots \times n_d$ arrays (sometime called *hyper-matrices*). The entries $X_{i_1,\ldots,i_d}$ of $\mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ will be indexed by multi-indices $(i_1, \ldots, i_d)$, with every i_j taking values between 1 and n_j. For convenience, it will be assumed

$$n_1 \geq n_2 \geq \cdots \geq n_d$$

throughout the paper. Furthermore, we set

$$n_j^c = \prod_{i \neq j} n_i.$$

A tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ admits d *principal matricizations* [10,11]

$$M_{\mathbf{X}}^{(j)} \in \mathbb{R}^{n_j \times n_j^c}$$

in which the i_j-th row contains all entries $X_{i_1,\ldots,i_d}$ with fixed i_j, arranged in some fixed ordering with respect to the remaining multi-indices. The choice of that ordering is not important for our purposes. The matricizations realize the isomorphism between the tensor space $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ and the matrix spaces $\mathbb{R}^{n_j \times n_j^c}$. Note that for tensors $\mathbf{X}$ of order $d = 2$ (that is, matrices), the matricizations are given by $M_{\mathbf{X}}^{(1)} = \mathbf{X}$ and $M_{\mathbf{X}}^{(2)} = \mathbf{X}^{\mathsf{T}}$ (up to permutation).

It will further be convenient to have a notation for the Gram matrix of $M_{\mathbf{X}}^{(j)}$, which will be

$$G_{\mathbf{X}}^{(j)} = M_{\mathbf{X}}^{(j)} \left(M_{\mathbf{X}}^{(j)} \right)^{\mathsf{T}}.$$

By $O(n)$ we denote the group of real orthogonal $n \times n$ matrices, by I_n the $n \times n$ identity matrix. Each of the matrices $M_{\mathbf{X}}^{(j)}$ admits a singular value decomposition

$$M_{\mathbf{X}}^{(j)} = U_{\mathbf{X}}^{(j)} \Sigma_{\mathbf{X}}^{(j)} \left(V_{\mathbf{X}}^{(j)} \right)^{\mathsf{T}}, \tag{1.1}$$

where $U_{\mathbf{X}}^{(j)} \in O(n_j)$, $(V_{\mathbf{X}}^{(j)})^{\mathsf{T}} V_{\mathbf{X}}^{(j)} = I_{n_j}$, and $\Sigma_{\mathbf{X}}^{(j)}$ is a diagonal matrix containing the *mode-j singular values* $\sigma_1^{(j)} \geq \sigma_2^{(j)} \geq \ldots \sigma_{n_j}^{(j)} \geq 0$ as diagonal elements. We denote $\Lambda_{\mathbf{X}}^{(j)} = (\Sigma_{\mathbf{X}}^{(j)})^2$. The diagonal entries of $\Lambda_{\mathbf{X}}^{(j)}$ are the ordered eigenvalues of $G_{\mathbf{X}}^{(j)}$.

Definition 1.1 Let $\mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

1. For $j = 1, \ldots, d$, the vector

$$\sigma_{\mathbf{X}}^{(j)} = \operatorname{diag}\left(\Sigma_{\mathbf{X}}^{(j)}\right) = \left(\sigma_1^{(j)}, \ldots, \sigma_{n_j}^{(j)}\right) \in \mathbb{R}_+^{n_j}$$

is called the *vector of mode-j singular values*. The tuple

$$\Sigma_{\mathbf{X}} = \left(\sigma_{\mathbf{X}}^{(1)}, \ldots, \sigma_{\mathbf{X}}^{(d)}\right) \in \mathbb{R}_+^{n_1} \times \cdots \times \mathbb{R}_+^{n_d}$$

is called the set of *higher-order singular values* of the tensor $\mathbf{X}$.

2. Correspondingly, for $j = 1, \ldots, d$, the vector

$$\lambda_{\mathbf{X}}^{(j)} = \operatorname{diag}\left(\Lambda_{\mathbf{X}}^{(j)}\right) = \left(\left(\sigma_1^{(j)}\right)^2, \ldots, \left(\sigma_{n_j}^{(j)}\right)^2\right) \in \mathbb{R}_+^{n_j}$$

is called the *vector of mode-j Gramian eigenvalues*. The tuple

$$\Lambda_{\mathbf{X}} = \left(\lambda_{\mathbf{X}}^{(1)}, \ldots, \lambda_{\mathbf{X}}^{(d)}\right)$$

is called the set of *higher-order Gramian eigenvalues* of the tensor $\mathbf{X}$.

3. The *multilinear rank* of the tensor $\mathbf{X}$ is the tuple $\mathbf{r_X} = (r^{(1)}, \ldots, r^{(d)})$ with $r^{(j)} = \operatorname{rank}(M_{\mathbf{X}}^{(j)}) = \operatorname{rank}(G_{\mathbf{X}}^{(j)})$ being equal to the number of nonzero entries of $\sigma_{\mathbf{X}}^{(j)}$.

4. The tensor $\mathbf{X}$ is called *non-singular*, if $\mathbf{r_X} = (n_1, \ldots, n_d)$.

We note that for matrices the definition of 'non-singular' coincides with the usual definition (in particular, it enforces $n_1 = n_2$). In general, the following is true.

Proposition 1.2 *There exists a non-singular tensor in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ if and only if the following compatibility conditions hold:*

$$n_j \leq n_j^c, \quad j = 1, 2, \ldots, d. \tag{1.2}$$

In this case the set of non-singular tensors is open and dense in $\mathbb{R}^{n_1 \times \cdots \times n_d}$.

Proof Consider j fixed. By isomorphy and known results on matrices, it is clear that the set of all $\mathbf{X}$ with $M_{\mathbf{X}}^{(j)}$ being of rank n_j is not empty, open, and dense if and only if $n_j \leq n_j^c$. The set of non-singular tensors is the intersection of these sets for $j = 1, \ldots, d$. As such, it is also open and dense. $\qquad\square$

Let $\| \cdot \|_F$ denote the Frobenius norm of matrices and tensors, and $\| \cdot \|_2$ the standard Euclidean norm for vectors. Since matricization of a tensor is an isometric isomorphism in Frobenius norm, and since it holds $\|M_{\mathbf{X}}^{(j)}\|_F = \|\sigma_{\mathbf{X}}^{(j)}\|_2$, an obvious observation for higher-order singular values is

$$\|\sigma_{\mathbf{X}}^{(1)}\|_2 = \cdots = \|\sigma_{\mathbf{X}}^{(d)}\|_2 = \|\mathbf{X}\|_F.$$

Therefore, we can focus in the following on tensors $\mathbf{X}$ on the unit sphere

$$\mathscr{S} = \left\{ \mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d} : \|\mathbf{X}\|_F = 1 \right\},$$

and hence higher-order singular values $\boldsymbol{\Sigma}$ in the set

$$\mathfrak{S}_{\geq} = \mathfrak{S}_{\geq}(n_1, \ldots, n_d) = S_{\geq}^{(1)} \times \cdots \times S_{\geq}^{(d)},$$

where $S_{\geq}^{(j)}$ denotes the set of all nonnegative, decreasingly ordered vectors on the Euclidean unit sphere in $\mathbb{R}^{n_j}$. For most results, however, it will be necessary to further restrict to the set $\mathscr{S}^*$ of non-singular tensors having strictly decreasing mode-j singular values in every direction j. Therefore, we also introduce the notation

$$\mathfrak{S}_{>} = \mathfrak{S}_{>}(n_1, \ldots, n_d) = S_{>}^{(1)} \times \cdots \times S_{>}^{(d)} \subset \mathfrak{S}_{\geq},$$

where each $S_{>}^{(j)}$ contains the unit norm vectors in $\mathbb{R}^{n_j}$ with strictly decreasing *and* strictly positive entries. Then

$$\mathscr{S}^* = \{\mathbf{X} \in \mathscr{S} : \boldsymbol{\Sigma}_{\mathbf{X}} \in \mathfrak{S}_{>}\}.$$

Note that we do not introduce a notation for the slightly larger set of all non-singular tensors in $\mathscr{S}$. The main technical advantage of tensors in $\mathscr{S}^*$ is that all principal unfoldings admit essentially unique singular value decompositions.

The following two facts are useful to know, and follow immediately from the matrix case.

Proposition 1.3 *The function $\mathbf{X} \mapsto \boldsymbol{\Sigma}_{\mathbf{X}}$ is continuous on $\mathscr{S}$. Assuming (1.2), the set $\mathscr{S}^*$ is relatively open and dense in $\mathscr{S}$.*

Proof The continuity of $\boldsymbol{\Sigma}_{\mathbf{X}}$ as a function of $\mathbf{X}$ follows by isomorphy to $\mathbb{R}^{n_j \times n_j^c}$ from the continuity of each $\sigma_{\mathbf{X}}^{(j)}$ as a function of $M_{\mathbf{X}}^{(j)}$. The proof that $\mathscr{S}^*$ is relatively open and dense in $\mathscr{S}$ is analogous to the proof of Proposition 1.2. $\square$

1.2 Problem statement

Regarding the higher-order singular values of tensors a principle question of interest is the following one.

Problem 1.4 (*Feasible higher-order singular values*) Given $\boldsymbol{\Sigma} \in \mathfrak{S}_{\geq}$, does there exist a tensor $\mathbf{X} \in \mathscr{S}$ such that $\boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}$?

Such $\boldsymbol{\Sigma}$ will be called a *feasible*. We define

$$\mathfrak{F} = \mathfrak{F}(n_1, \ldots, n_d) = \{\boldsymbol{\Sigma} \in \mathfrak{S}_{\geq} : \boldsymbol{\Sigma} \text{ is feasible}\}.$$

In this generality, Problem 1.4 appears to be quite hard, and will not be satisfactorily solved in this article. At least, as a first result, we are able to show that not all $\boldsymbol{\Sigma} \in \mathfrak{S}_{\geq}$ are feasible: $\mathfrak{F} \neq \mathfrak{S}_{\geq}$. The argument, however, is non-constructive, see Sect. 3.1.

A relaxed question of a more qualitative nature is the following one.

Problem 1.5 (*Properties of $\mathfrak{F}$*) What are the topological properties of the set $\mathfrak{F}$ as a subset of $\mathfrak{S}_{\geq}$? Does it, for instance, have positive (relative) Lebesgue measure?

Numerical experiments with random tensors seem to indicate that the answer to the second question could be positive when $d \geq 3$, but we are not able to prove it. So it remains a conjecture. In fact, we conjecture that for every $\mathbf{X}$ in $\mathscr{S}^*$ it holds that $\boldsymbol{\Sigma_X}$ is an interior point of $\mathfrak{F}$, see Sect. 3.2. A striking implication of this conjecture is that given $\mathbf{X} \in \mathscr{S}^*$, its high-order singular values in different directions can be perturbed independently from each other without loosing feasibility (local independence of high-order singular values). In Sect. 3.3 we will present a heuristic approach to do this using an alternating projection method, which seems to work quite reliably for small perturbations, although we are currently neither able to prove its convergence nor that limit points must have the desired property.

To approach Problems 1.4 and 1.5, it seems useful to also study the following problem, which is of some interest in itself.

Problem 1.6 (*Tensors with same higher-order singular values*) Given $\mathbf{X} \in \mathscr{S}$, characterize sets of tensors having the same singular values $\boldsymbol{\Sigma_X}$ as $\mathbf{X}$.

The corresponding equivalence classes for tensors in $\mathscr{S}$ and $\mathscr{S}^*$ are denoted by

$$\mathscr{S}_\mathbf{X} = \{\mathbf{Y} \in \mathscr{S} \colon \boldsymbol{\Sigma_Y} = \boldsymbol{\Sigma_X}\}, \quad \mathscr{S}_\mathbf{X}^* = \{\mathbf{Y} \in \mathscr{S}^* \colon \boldsymbol{\Sigma_Y} = \boldsymbol{\Sigma_X}\}.$$

The next Sect. 2 provides some results related to Problem 1.6. It is observed that orbits of orthogonally equivalent tensors provide trivial examples of subsets of tensors having the same higher-order singular values. However, other than in the matrix case, their dimension is too small to provide a complete description. Via the tool of HOSVD tensors, which serve as normal forms in the orbits of orthogonally equivalent tensors, we are able to construct an example of two tensors with the same higher-order singular values that are not orthogonally equivalent.

2 Tensors with the same higher-order singular values

In this section we focus on equivalence classes of tensors having the same higher-order singular values.

2.1 Orthogonally equivalent tensors

We recall a fact from matrices: Two rectangular matrices $X, Y \in \mathbb{R}^{n \times n'}$, $n \leq n'$, have the same singular values, if and only if they are orthogonally equivalent, that is, if there exists $U \in \mathrm{O}(n)$ and $V \in \mathrm{O}(n')$ such that

$$UXV^\mathsf{T} = Y.$$

This definition of orthogonal equivalence can be generalized to tensors using the multilinear matrix multiplication, see, e.g., [5,13]. We consider the product unitary group

$$\mathrm{O}(n_1 \times \cdots \times n_d) = \mathrm{O}(n_1) \times \mathrm{O}(n_2) \times \cdots \times \mathrm{O}(n_d).$$

The left action $(U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}$ of this group on $\mathbb{R}^{n_1 \times \cdots \times n_d}$ is defined as the canonical action of the tensor product operator $U^{(1)} \otimes \cdots \otimes U^{(d)}$ on $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ in the sense that

$$(U^{(1)}, \ldots, U^{(d)}) \cdot \left(\sum_{k=1}^{r} x_k^{(1)} \otimes \cdots \otimes x_k^{(d)} \right) = \sum_{k=1}^{r} U^{(1)} x_k^{(1)} \otimes \cdots \otimes U^{(d)} x_k^{(d)}.$$

In terms of matricizations, in a slight abuse of notation, we note that

$$M_{(U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}}^{(j)} = U^{(j)} M_{\mathbf{X}}^{(j)} \left(U^{(1)} \otimes \cdots \otimes U^{(j-1)} \otimes U^{(j+1)} \otimes \cdots \otimes U^{(d)} \right)^\mathsf{T}, \tag{2.1}$$

cf. [8, Lemma 5.6]. In particular, since $U^{(1)} \otimes \cdots \otimes U^{(j-1)} \otimes U^{(j+1)} \otimes \cdots \otimes U^{(d)} \in \mathrm{O}(n_j^c)$, it holds

$$G_{(U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}}^{(j)} = U^{(j)} G_{\mathbf{X}}^{(j)} (U^{(j)})^\mathsf{T} \tag{2.2}$$

for $j = 1, \ldots, d$. For matrices ($d = 2$), these formulas define orthogonal equivalence, which motivates the following generalization.

Definition 2.1 (*see* [5]) Two tensors $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ are called *orthogonally equivalent*, if there exists $(U^{(1)}, \ldots, U^{(d)}) \in \mathrm{O}(n_1 \times \cdots \times n_d)$ such $\mathbf{Y} = (U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}$.

From (2.2), we draw a trivial but important conclusion.

Proposition 2.2 *If two tensors are orthogonally equivalent, then they have the same higher-order singular values.*

In particular, the *orbit* of each $\mathbf{X}$ under the group action contains only tensors with identical higher-order singular values.

Proposition 2.3 *Let* $\mathbf{X} \in \mathscr{S}^*$. *Then the* orbit $\mathrm{O}(n_1 \times \cdots \times n_d) \cdot \mathbf{X}$ *is a locally smoothly embedded submanifold of* $\mathscr{S}^*$ *of dimension*

$$\dim(\mathrm{O}(n_1 \times \cdots \times n_d) \cdot \mathbf{X}) = \dim(\mathrm{O}(n_1 \times \cdots \times n_d)) = \sum_{j=1}^{d} \frac{1}{2} n_j (n_j - 1).$$

Proof We write O instead of $O(n_1 \times \cdots \times n_d)$. Consider the canonical map $\theta_{\mathbf{X}} \colon O \to \mathscr{S}^*$, $(U^{(1)}, \ldots, U^{(d)}) \mapsto (U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}$, whose image is $O \cdot \mathbf{X}$. Since $\theta_{\mathbf{X}}$ is of constant rank [6, §16.10.2] and easily shown to be locally injective (uniqueness of left singular vectors up to sign flipping for $\mathbf{X} \in \mathscr{S}^*$), it is already an immersion [6, §16.8.8.(iv)]. The result is now standard, see, e.g., [6, §16.8.8.(ii)]. □

For $\mathbf{X} \in \mathscr{S} \setminus \mathscr{S}^*$ the dimension of $O(n_1 \times \cdots \times n_d) \cdot \mathbf{X}$ can be smaller than $\dim O(n_1 \times \cdots \times n_d)$. Note that we did not attempt to prove or disprove that the orbits are globally embedded submanifolds.

2.2 HOSVD tensors

The compact Lie group $O(n_1 \times \cdots \times n_d)$ acts freely on $\mathscr{S}^*$. It also acts properly (since it is compact and acts continuously). By a general theorem (e.g. [6, § 16.10.3]), the quotient manifold $\mathscr{S}^*/O(n_1 \times \cdots \times n_d)$ of equivalence classes exists, and the canonical mapping $\mathscr{S}^* \to \mathscr{S}^*/O(n_1 \times \cdots \times n_d)$ is a submersion. A concrete realization of this abstract quotient manifold is the set $\mathscr{H}^*$ of regular HOSVD tensors which is now introduced.

Given $\mathbf{Y}$, let $U_{\mathbf{Y}}^{(j)}$ denote a matrix of left singular vectors of $M_{\mathbf{Y}}^{(j)}$ as in (1.1). By (2.1), $\mathbf{X} = ((U_{\mathbf{Y}}^{(1)})^{\mathsf{T}}, \ldots, (U_{\mathbf{Y}}^{(d)})^{\mathsf{T}}) \cdot \mathbf{Y}$ has the matricizations

$$M_{\mathbf{X}}^{(j)} = \Sigma_{\mathbf{Y}}^{(j)} \left(V_{\mathbf{Y}}^{(j)} \right)^{\mathsf{T}} \left(U_{\mathbf{Y}}^{(1)} \otimes \cdots \otimes U_{\mathbf{Y}}^{(j-1)} \otimes U_{\mathbf{Y}}^{(j+1)} \otimes \cdots \otimes U_{\mathbf{Y}}^{(d)} \right).$$

In particular, the rows of $(V_{\mathbf{Y}}^{(j)})^{\mathsf{T}}(U_{\mathbf{Y}}^{(1)} \otimes \cdots \otimes U_{\mathbf{Y}}^{(j-1)} \otimes U_{\mathbf{Y}}^{(j+1)} \otimes \cdots \otimes U_{\mathbf{Y}}^{(d)})$ are right singular vectors of $M_{\mathbf{X}}^{(j)}$, the left singular vectors are unit vectors, and the singular values are the same as of $\mathbf{Y}$, that is, $\Sigma_{\mathbf{X}}^{(j)} = \Sigma_{\mathbf{Y}}^{(j)}$. Hence $\mathbf{X}$ has the specific property that

$$G_{\mathbf{X}}^{(j)} = M_{\mathbf{X}}^{(j)} \left(M_{\mathbf{X}}^{(j)} \right)^{\mathsf{T}} = \left(\Sigma_{\mathbf{X}}^{(j)} \right)^2 = \Lambda_{\mathbf{X}}^{(j)}, \quad j = 1, \ldots, d, \tag{2.3}$$

is a diagonal matrix of decreasing eigenvalues. The reverse relation

$$\mathbf{Y} = \left(U_{\mathbf{Y}}^{(1)}, \ldots, U_{\mathbf{Y}}^{(d)} \right) \cdot \mathbf{X}$$

between $\mathbf{X}$ and $\mathbf{Y}$ is called the *higher-order singular value decomposition (HOSVD)* of $\mathbf{Y}$ and has been introduced by De Lathauwer et al. [3].

Definition 2.4 Tensors satisfying (2.3) are called *HOSVD tensors*. The subset of HOSVD tensors in $\mathscr{S}$ is denoted by $\mathscr{H}$, and the subset of HOSVD tensors in $\mathscr{S}^*$ by $\mathscr{H}^* = \mathscr{H} \cap \mathscr{S}^*$.

HOSVD tensors can be regarded as representatives of orbits $O(n_1 \times \cdots \times n_d) \cdot \mathbf{X}$ of orthogonally equivalent tensors. For $\mathbf{X} \in \mathscr{S}^*$, the representatives are essentially unique as stated next. Here it is instructive to note that for square matrices, the set $\mathscr{H}^*$ consists of regular diagonal matrices with strictly decreasing diagonal entries.

Proposition 2.5 *Let* $\mathbf{X}, \mathbf{Y} \in \mathscr{H}^*$ *be two HOSVD tensors. If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are orthogonally equivalent, that is,* $\mathbf{Y} = (U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}$, *then the* $U^{(j)}$ *must be diagonal orthogonal matrices (i.e. with values* ± 1 *on the diagonal).*

The proof is immediate from (2.2), (2.3), and the uniqueness of orthogonal diagonalization up to sign flipping in the case of mutually distinct eigenvalues. Comparing with the explicit form (2.1), we see that the action of $(U^{(1)}, \ldots, U^{(d)}) \cdot \mathbf{X}$ with diagonal $U^{(j)}$ with ± 1 entries results in *some* sign flipping pattern for the entries of $\mathbf{X}$. This provides the following, sometimes useful necessary condition.

Proposition 2.6 *If two HOSVD tensors* $\mathbf{X}, \mathbf{Y} \in \mathscr{H}^*$ *are orthogonally equivalent, then* $|X_{i_1, \ldots, i_d}| = |Y_{i_1, \ldots, i_d}|$. *In particular,* $\mathbf{X}$ *and* $\mathbf{Y}$ *have the same zero pattern.*

We now turn to the manifold properties of $\mathscr{H}^*$.

Theorem 2.7 *The set* $\mathscr{H}^*$ *is a smooth embedded submanifold of* $\mathscr{S}^*$ *of dimension*

$$\dim(\mathscr{H}^*) = n_1 \cdots n_d - \sum_{j=1}^{d} \frac{1}{2} n_j (n_j - 1) = \dim(\mathscr{S}^*) - \dim(O(n_1 \times \cdots \times n_d)).$$

Proof The formal setting is as follows. We denote by $\mathbb{R}_{\mathrm{sym}}^{n_j \times n_j}$ the space of symmetric $n_j \times n_j$ matrices, by $\mathbb{R}_{\mathrm{sym},0}^{n_j \times n_j}$ the subspace with zeros on the diagonal, by $\pi_j(Z) = Z - \mathrm{diag}(Z)$ the orthogonal projection from $\mathbb{R}_{\mathrm{sym}}^{n_j \times n_j}$ onto $\mathbb{R}_{\mathrm{sym},0}^{n_j \times n_j}$, and $p = \pi_1 \times \cdots \times \pi_d$. Consider

$$g: \mathscr{S}^* \to \mathbb{R}_{\mathrm{sym}}^{n_1 \times n_1} \times \cdots \times \mathbb{R}_{\mathrm{sym}}^{n_d \times n_d}, \quad \mathbf{X} \mapsto (G_{\mathbf{X}}^{(1)}, \ldots, G_{\mathbf{X}}^{(d)}). \tag{2.4}$$

Then $f = p \circ g$ is a smooth map from $\mathscr{S}^*$ to $\mathbb{R}_{\mathrm{sym},0}^{n_1 \times n_1} \times \cdots \times \mathbb{R}_{\mathrm{sym},0}^{n_d \times n_d}$, and, by Definition 2.4, we have

$$\mathscr{H}^* = f^{-1}(0).$$

Since $\dim(\mathbb{R}_{\mathrm{sym},0}^{n_1 \times n_1} \times \cdots \times \mathbb{R}_{\mathrm{sym},0}^{n_d \times n_d}) = \dim(O(n_1 \times \cdots \times n_d)) = \sum_{j=1}^{d} \frac{1}{2} n_j (n_j - 1)$, the assertion will follow from the regular value theorem, if we show that $f'(\mathbf{X})$ is surjective for every $\mathbf{X} \in \mathscr{H}^*$. To prove the latter, we show that the range of $f'(\mathbf{X})$ contains the spaces $\mathscr{W}_j = \{0\} \times \cdots \times \{0\} \times \mathbb{R}_{\mathrm{sym},0}^{n_j \times n_j} \times \{0\} \times \cdots \times \{0\}$ for $j = 1, \ldots, d$. We demonstrate this for $j = 1$. Consider the map

$$\varphi: O(n_1) \to \mathscr{S}^*, \quad U \mapsto (U, I_{n_2}, \ldots, I_{n_d}) \cdot \mathbf{X}.$$

For brevity, we set $I = I_{n_1}$. Since $\varphi(I) = \mathbf{X}$, it follows from the chain rule that the range of $f'(\mathbf{X})$ contains the range of $(f \circ \varphi)'(I)$. We show that the latter equals $\mathscr{W}_1$. Since $\mathbf{X} \in \mathscr{H}^*$, we have $(f \circ \varphi)(U) = p(U \Lambda_{\mathbf{X}}^{(1)} U^{\mathsf{T}}, \Lambda_{\mathbf{X}}^{(2)}, \ldots, \Lambda_{\mathbf{X}}^{(d)})$. Further noting

that the tangent space to $O(n_1)$ at I is the space $\mathbb{R}^{n_1 \times n_1}_{\text{skew}}$ of skew-symmetric $n_1 \times n_1$ matrices, we see that

$$(f \circ \varphi)'(I) \colon \mathbb{R}^{n_1 \times n_1}_{\text{skew}} \to \mathscr{W}_1, \quad H \mapsto p(H\Lambda_{\mathbf{X}}^{(1)} + \Lambda_{\mathbf{X}}^{(1)} H^{\mathsf{T}}, 0, \dots, 0).$$

As $\dim(\mathscr{W}_1) = \frac{1}{2}n_1(n_1 - 1) = \dim(\mathbb{R}^{n_1 \times n_1}_{\text{skew}})$, it is enough to show that $(f \circ \varphi)'(I)$ is injective in order to finish the proof. This now follows from the fact that, by definition of $\mathscr{H}^*$, the diagonal entries of $\Lambda_{\mathbf{X}}^{(1)}$ are strictly decreasing, as it implies that $H\Lambda_{\mathbf{X}}^{(1)} + \Lambda_{\mathbf{X}}^{(1)} H^{\mathsf{T}} = H\Lambda_{\mathbf{X}}^{(1)} - \Lambda_{\mathbf{X}}^{(1)} H$ cannot be diagonal for skew-symmetric $H \neq 0$. This, however, is equivalent to injectivity of $(f \circ \varphi)'(I)$ as given above. $\qquad \square$

Remark 2.8 In our definition (2.3) of HOSVD tensors we required the diagonal elements of $G_{\mathbf{X}}^{(j)}$ to be decreasing. This has advantages and drawbacks. One advantage are the narrower uniqueness properties leading to the practical condition in Proposition 2.6. A disadvantage is that it is more difficult to design HOSVD tensors "by hand" as in Sect. 2.4. Alternatively, one may define a set $\tilde{\mathscr{H}}$ by just requiring the $G_{\mathbf{X}}^{(j)}$ to be diagonal. Then for every $\mathbf{X} \in \tilde{\mathscr{H}}$ we have $(P^{(1)}, \dots, P^{(d)}) \cdot \mathbf{X} \in \mathscr{H}$, where $P^{(j)}$ are permutation matrices that sort the diagonal entries of $G_{\mathbf{X}}^{(j)}$ accordingly. For mutually distinct eigenvalues the choice of $P^{(j)}$ is unique. The corresponding set $\tilde{\mathscr{H}}^*$ is therefore the finite disjoint union of sets $(P^{(1)}, \dots, P^{(d)}) \cdot \mathscr{H}^*$ over all $P^{(j)}$, and as such also an embedded submanifold of $\mathscr{S}^*$.

2.3 Degrees of freedom

A principal challenge in understanding the interconnection between higher-order singular values of tensors arises from the fact that, in contrast to the matrix case, the converse statement of Proposition 2.2 is in general not true when $d \geq 3$. Tensors may have the same higher-order singular values without being orthogonally equivalent. This can be seen from the following heuristic.

The set $\mathscr{S}^*$ is open and dense in $\mathscr{S}$ by Proposition 1.3, and therefore is a smooth manifold of dimension

$$\dim(\mathscr{S}^*) = (n_1 \dots n_d) - 1.$$

The set $\mathfrak{S}_>$ is an open subset of Cartesian products of spheres and hence of dimension

$$\dim(\mathfrak{S}_>) = (n_1 + \dots + n_d) - d.$$

Therefore, given $\mathbf{X} \in \mathscr{S}^*$, we expect the set $\mathscr{S}_{\mathbf{X}}^*$ of tensors having the same higher-order singular values as $\mathbf{X}$ to be at least of "dimension"

$$\dim(\mathscr{S}^*) - \dim(\mathfrak{S}_>) = (n_1 \cdots n_d) - (n_1 + \dots + n_2) + (d - 1).$$

When $d \geq 3$, by Proposition 2.3, this set cannot only consist of tensors that are orthogonally equivalent to $\mathbf{X}$.[1] In fact, for large d, the orthogonally equivalent tensors will only be a very "low-dimensional" subset of $\mathscr{S}_{\mathbf{X}}^*$.

2.4 A non-equivalent example

The previous considerations suggest that there must exist tensors having the same higher-order singular values without being orthogonally equivalent. We construct here an example of size $2 \times 2 \times 2$ using Proposition 2.6. Let us shortly count the degrees of freedom in this situation. The Euclidean unit sphere $\mathscr{S}$ is of dimension seven, the set $\mathfrak{S}_\geq$ of potential tuples of higher-order singular values is of dimension three, while orbits $O(2 \times 2 \times 2) \cdot \mathbf{X}$ of orthogonally equivalent tensors are of dimension at most three, too. This indicates for every $\mathbf{X} \in \mathscr{S}$ an at least one-dimensional set of non-equivalent tensor with same higher-order singular values.

Using a common slice-wise notation of tensors, we consider (currently not normalized)

$$\mathbf{X} = \begin{bmatrix} -2 & 1 & | & -1 & -1 \\ 1 & 1 & | & -1 & 0 \end{bmatrix}.$$

The three matricizations are $M_{\mathbf{X}}^{(1)} = \begin{bmatrix} -2 & 1 & -1 & -1 \\ 1 & 1 & -1 & 0 \end{bmatrix}$, $M_{\mathbf{X}}^{(2)} = \begin{bmatrix} -2 & 1 & -1 & -1 \\ 1 & 1 & -1 & 0 \end{bmatrix}$, and $M_{\mathbf{X}}^{(3)} = \begin{bmatrix} -2 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 \end{bmatrix}$. In all three matricizations the rows are orthogonal, and the norm of the first row is larger than the norm of the second one. This shows that $\mathbf{X}$ is a HOSVD tensor. Its squared higher-order singular values are

$$\Lambda_{\mathbf{X}} = \left[\lambda_{\mathbf{X}}^{(1)}, \lambda_{\mathbf{X}}^{(2)}, \lambda_{\mathbf{X}}^{(3)} \right] = \left[\begin{pmatrix} 7 \\ 3 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \end{pmatrix} \right].$$

In particular, $\mathbf{X}/\|\mathbf{X}\|_F \in \mathscr{S}^*$. As a second tensor consider

$$\mathbf{Y} = \begin{bmatrix} 3/\sqrt{2} & -1 & | & 1 & 1/\sqrt{2} \\ -1 & -1/\sqrt{2} & | & 1/\sqrt{2} & 1 \end{bmatrix}.$$

One checks again that all three matricizations $M_{\mathbf{Y}}^{(1)} = \begin{bmatrix} 3/\sqrt{2} & -1 & 1 & 1/\sqrt{2} \\ -1 & -1/\sqrt{2} & 1/\sqrt{2} & 1 \end{bmatrix}$, $M_{\mathbf{Y}}^{(2)} = \begin{bmatrix} 3/\sqrt{2} & -1 & 1 & 1/\sqrt{2} \\ -1 & -1/\sqrt{2} & 1/\sqrt{2} & 1 \end{bmatrix}$, and $M_{\mathbf{Y}}^{(3)} = \begin{bmatrix} 3/\sqrt{2} & -1 & -1 & -1/\sqrt{2} \\ 1 & 1/\sqrt{2} & 1/\sqrt{2} & 1 \end{bmatrix}$

---

[1] For $d = 2$, i.e., matrices, it is the case: as (1.2) is assumed, we have $n_1 = n_2 = n$, and two square matrices have the same singular values if and only if they are orthogonally equivalent. The formula gives $n^2 - 2n + 1$ which, however, only equals $n^2 - n - (n - 1) = \dim(O(n \times n)) - (n - 1)$. The reason is that in the matrix case we know that the singular values of $\mathbf{X}$ and $\mathbf{X}^\top$ are the same. Hence the feasible set is only of dimension $n - 1$, and not of dimension $2(n - 1)$ (the argument will be repeated in Sect. 3.1). For tensors, however, we conjecture that the dimension of $\mathfrak{F}$ is indeed $(n_1 + \cdots + n_d) - d$, see Sect. 3.2.

have orthogonal rows with squared row norms

$$\mathbf{\Lambda_Y} = \left[\lambda_{\mathbf{Y}}^{(1)}, \lambda_{\mathbf{Y}}^{(2)}, \lambda_{\mathbf{Y}}^{(3)}\right] = \left[\binom{7}{3}, \binom{7}{3}, \binom{7}{3}\right].$$

This shows that $\mathbf{X}/\|\mathbf{X}\|_F$ and $\mathbf{Y}/\|\mathbf{Y}\|_F$ are two HOSVD tensors in $\mathscr{S}^*$ with the same set of higher-order singular values. By Proposition 2.6, they are not orthogonally equivalent.

3 The set of feasible configurations

The set $\mathfrak{F} = \mathfrak{F}(n_1, \ldots, n_d) \subseteq \mathfrak{S}_\geq$ of feasible configurations has been defined in (1.2). In this section we investigate this set. A simple observation worth to mention is that $\mathfrak{F}$ is closed. This follows from Proposition 1.3 and the compactness of $\mathscr{S}$.

3.1 Not all configurations are feasible

When $d = 2$, we know that the singular values of a matrix and its transpose are the same, so trivially not all configurations for $\sigma_{\mathbf{X}}^{(1)}$ and $\sigma_{\mathbf{X}}^{(2)}$ are possible. The formal statement, using the introduced notation, reads, with $n_1 \geq n_2$,

$$\mathfrak{F}(n_1, n_2) = \left\{\mathbf{\Sigma} = (\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}) \in \mathfrak{S}_\geq(n_1, n_2) \colon \boldsymbol{\sigma}^{(1)} = (\boldsymbol{\sigma}^{(2)}, 0, \ldots, 0)\right\} \neq \mathfrak{S}_\geq.$$

In fact, $\mathfrak{F}$ is an n_2-dimensional subset in the $(n_1 + n_2)$-dimensional set $\mathfrak{S}_\geq(n_1, n_2)$.

This known phenomenon in the matrix case can be used to give a qualitative proof that also for higher-order tensors not all configurations are feasible. To start, we recall a fact on the HOSVD from the literature. Let $\mathbf{X}$ have the left singular vector matrices $U_{\mathbf{X}}^{(j)}$ (column-wise ordered by decreasing singular values), and multilinear rank $\mathbf{r} = (r_1, \ldots, r_d)$. Then we can write the "economic" HOSVD as

$$\mathbf{X} = \left(\hat{U}_{\mathbf{X}}^{(1)}, \ldots, \hat{U}_{\mathbf{X}}^{(d)}\right) \cdot \mathbf{C},$$

where $\hat{U}_{\mathbf{X}}^{(j)}$ contains only the first r_j columns of $U_{\mathbf{X}}^{(j)}$, and the *core tensor* $\mathbf{C}$ is of size $r_1 \times \cdots \times r_d$. The multilinear matrix product here corresponds to the action of the tensor product operator $\hat{U}_{\mathbf{X}}^{(1)} \otimes \cdots \otimes \hat{U}_{\mathbf{X}}^{(d)}$ on $\mathbb{R}^{r_1} \otimes \cdots \otimes \mathbb{R}^{r_d}$, the explicit formulas are similar to (2.1). Note that if $\mathbf{X}$ is non-singular, $\mathbf{C}$ is just an HOSVD tensor in the orthogonally equivalent orbit of $\mathbf{X}$ as defined above. The key observation in the general case is that $\mathbf{C}$ is non-singular in $\mathbb{R}^{r_1 \times \cdots \times r_d}$, and its higher-order singular values in every direction are given by the nonzero higher-order singular values of $\mathbf{X}$ [3].

Based on this fact, we can first give trivial examples of singular tensors for which the nonzero singular values in different directions are not independent of each other.

Lemma 3.1 *Let $\mathbf{X}$ have multilinear rank $\mathbf{r} = (r_1, \ldots, r_d)$. Assume $r_j = 1$ for $j \geq 3$. Then $r_1 = r_2$ and $\{\sigma_1^{(1)}, \ldots, \sigma_{r_1}^{(1)}\} = \{\sigma_1^{(2)}, \ldots, \sigma_{r_2}^{(2)}\}$.*

Proof Let $\mathbf{C} \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ be the economic HOSVD core tensor of $\mathbf{X}$. The matriciza-tions $M_{\mathbf{C}}^{(j)}$ for $j \geq 3$ are just row vectors and have only one singular value which equals the Frobenius norm of $\mathbf{X}$. On the other hand, we have $M_{\mathbf{C}}^{(1)} = (M_{\mathbf{C}}^{(2)})^{\mathsf{T}}$ (up to possible permutations), which implies the result. □

Since tensors with $r_j = 1$ for $j \geq 3$ considered in the previous lemma are naturally identified as elements of $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, that is, as matrices, the previous statement may appear rather odd at first. However, using a perturbation argument, it leads to a non-constructive proof that non-feasible configurations for higher-order singular values do exist even in the non-singular case. In fact, these configurations are of positive volume within $\mathfrak{S}_{\geq}$.

Theorem 3.2 *For $n_1 \geq n_2$, consider $\boldsymbol{\sigma}^{(1)} \in S_{\geq}^{(1)}$ and $\boldsymbol{\sigma}^{(2)} \in S_{\geq}^{(2)}$ such that*

$$\boldsymbol{\sigma}^{(1)} \neq (\boldsymbol{\sigma}^{(2)}, 0, \ldots, 0), \tag{3.1}$$

where the number of appended zeros on the right side equals $n_1 - n_2$. Let further $\mathscr{O}_{\epsilon}^{(j)}$, $j = 3, \ldots, d$, denote neighbourhoods of $\mathbf{e}_1^{(j)} = (1, 0, \ldots, 0)$ in $S_{\geq}^{(j)}$ of diameter at most $\epsilon > 0$ (w.r.t. some norm). For $j \in \{1, 2\}$, let $\mathscr{O}_{\epsilon}^{(j)}$ be similar neighbourhoods, but of $\boldsymbol{\sigma}^{(1)}$ and $\boldsymbol{\sigma}^{(2)}$, respectively. Then there exists $\epsilon > 0$ such that

$$\mathfrak{F} \cap \left(\mathscr{O}_{\epsilon}^{(1)} \times \cdots \times \mathscr{O}_{\epsilon}^{(d)} \right) = \emptyset,$$

that is, no $\boldsymbol{\Sigma} \in \mathscr{O}_{\epsilon}^{(1)} \times \cdots \times \mathscr{O}_{\epsilon}^{(d)}$ is feasible.

Proof Assume to the contrary that for every n there exists a tensor $\mathbf{X}_n \in \mathscr{S}$ such that $\boldsymbol{\Sigma}_{\mathbf{X}_n} \in \mathscr{O}_{1/n}^{(1)} \times \cdots \times \mathscr{O}_{1/n}^{(d)}$. The sequence of $\mathbf{X}_n$ has a convergent subsequence with a limit $\mathbf{X} \in \mathscr{S}$. By Lemma 1.3, $\mathbf{X}$ has higher-order singular values $\boldsymbol{\Sigma}_{\mathbf{X}} = (\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}, \mathbf{e}_1^{(3)}, \ldots, \mathbf{e}_1^{(d)})$. Now Lemma 3.1 applies, but is in contradiction to (3.1). □

Remark 3.3 The condition (3.1) can hold in two cases: (i) the number of nonzero singular values in direction one and two are the same ($r_1 = r_2$), but the singular values themselves are not, or (ii) $r_1 \neq r_2$. The second case has some interesting implications for rectangular tensors. Assume for instance $n_1 \neq n_2$. Then by Theorem 3.2 there cannot exist normalized non-singular tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ for which the singular value vectors $\boldsymbol{\sigma}^{(j)}$ in directions $j = 3, \ldots, d$ are arbitrarily close to the corresponding unit vector $\mathbf{e}^{(j)}$. This surprising connection between mode sizes of the tensor and location of the singular value vectors is not obvious, especially given the fact that almost every tensor is non-singular (assuming (1.2)).

3.2 A conjecture on interior points

For $d = 2$ we have seen that $\mathfrak{F}(n_1, n_2)$ is a set of measure zero within $\mathfrak{S}_{\geq}(n_1, n_2)$, even when $n_1 = n_2$. One question is whether this is also true for higher-order tensors. Remarkably, the following experiment suggests that this does not need to be the case.
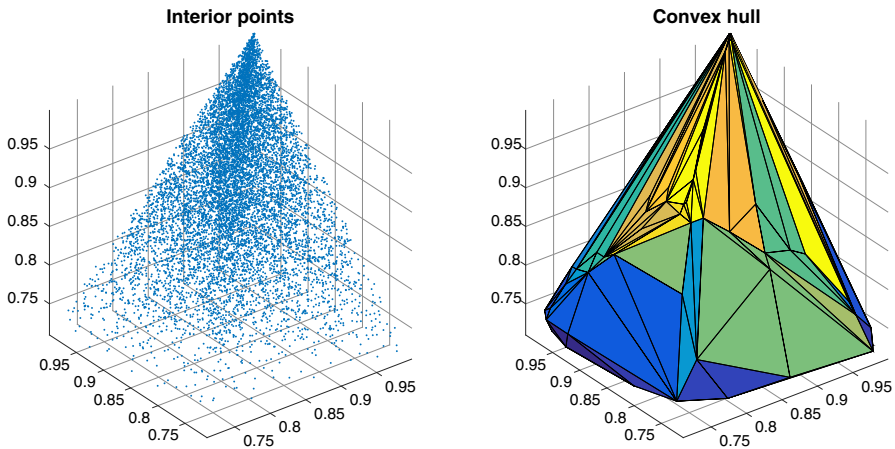
Fig. 1 Visualizing higher-order singular values in the $2 \times 2 \times 2$ case. Plotted are the vectors $(\sigma_1^{(1)}, \sigma_1^{(2)}, \sigma_1^{(3)})$ containing the largest singular values of the three directions for 10,000 random tensors $\mathbf{X} \in \mathscr{S}$. They seem to form a three-dimensional connected set. Hence, the corresponding set of $\mathbf{\Sigma_X} = (\sigma_{\mathbf{X}}^{(1)}, \sigma_{\mathbf{X}}^{(1)}, \sigma_{\mathbf{X}}^{(1)})$ should be of positive volume in the three-dimensional set $\mathfrak{S}_{\geq}$

We generate random $2 \times 2 \times 2$ tensors $\mathbf{X}$ of Frobenius norm one.[2] With probability one, the higher-order singular values $\mathbf{\Sigma_X} = (\sigma_{\mathbf{X}}^{(1)}, \sigma_{\mathbf{X}}^{(2)}, \sigma_{\mathbf{X}}^{(3)})$ are elements of $\mathfrak{S}_{>}(2, 2, 2)$, which is a set of three dimensions and therefore can be visualized. We simply make the identification,

$$\mathbf{\Sigma_X} \mapsto \left(\sigma_1^{(1)}, \sigma_1^{(2)}, \sigma_1^{(3)} \right) \in \left[\frac{1}{\sqrt{2}}, 1 \right]^3,$$

that is, we project on the first coordinate of each singular value vector. In Fig. 1 we see these projected points for 10,000 random examples, and their convex hull computed with a Matlab integrated Delauney triangulation.

As the resulting point cloud appears three-dimensional, we suppose that the set of feasible configurations is also three-dimensional. But one can also verify in the plot that not all configurations are feasible. Above we made use of the fact that $\sigma^{(j)} = (1, 0)$ (Tucker rank in the direction j equals one) implies $\sigma^{(i)} = \sigma^{(k)}$ for $i, k \neq j$. This can be seen in the picture as the convex polytope intersects the hyperplanes $x = 1$, $y = 1$ and $z = 1$ in single one-dimensional facets of 45 degree.

We are led to the following conjecture.

Conjecture 3.4 *When $d \geq 3$, and given the compatibility condition (1.2), the set $\mathfrak{F}(n_1, \ldots, n_d)$ has positive (relative) volume in $\mathfrak{S}_{\geq}(n_1, \ldots, n_d)$.*

In fact, the following seems likely (under the same assumptions).

Conjecture 3.5 *For generic $\mathbf{X} \in \mathscr{S}$, $\mathbf{\Sigma_X}$ is a (relative) interior point of $\mathfrak{F}(n_1, \ldots, n_d)$ within $\mathfrak{S}_{\geq}(n_1, \ldots, n_d)$.*

---

[2] For our experiments we made use of the Tensor Toolbox [2] in Matlab.

Remark 3.6 During revision of the paper, a possible strategy to prove this conjecture has been revealed. It is based on the observation that $\boldsymbol{\Sigma}_{\mathbf{X}}$ is a relative interior point of $\mathfrak{F}(n_1, \ldots, n_d)$ if and only if the map $g(\mathbf{X}) = (G_{\mathbf{X}}^{(1)}, \ldots, G_{\mathbf{X}}^{(d)})$ (that has already been considered in (2.4)) is locally surjective when regarded as a map from the unit sphere $\mathscr{S}$ to the Cartesian product of hyperplanes $\{A^{(j)} \in \mathbb{R}_{\mathrm{sym}}^{n_j \times n_j} : \mathrm{tr}(A^{(j)}) = 1\}$. In other words, one has to show that the rank of the derivative $g'(\mathbf{X})$, when restricted to the tangent space $T_{\mathbf{X}}\mathscr{S}$, equals the maximum possible value $\alpha = \left(\sum_{j=1}^{d} \frac{1}{2} n_j (n_j + 1)\right) - d$. A sufficient condition for this is that $g'(\mathbf{X})$ is of rank $\alpha + 1$ on $\mathbb{R}^{n_1 \times \cdots \times n_d}$. However, as $g'(\mathbf{X})$ depends polynomially on the entries of $\mathbf{X}$, the function $\mathbf{X} \mapsto \mathrm{rank}(g'(\mathbf{X}))$ achieves its maximum value for almost all $\mathbf{X}$. Since it is bounded by $\alpha + 1$, it is therefore enough to find a single tensor $\mathbf{X}$ for which rank $\alpha + 1$ is achieved. In this way, one can validate Conjecture 3.5 for different configurations of $n_1, \ldots, n_d$ by constructing random $\mathbf{X}$ and evaluating the rank of $g'(\mathbf{X})$ numerically. A rigorous proof would have to confirm this numerical rank for "simple" candidates $\mathbf{X}$, which we were able to do for $2 \times 2 \times 2$ tensors so far. This approach shall be subject of a future work.

3.3 Alternating projection method

Even in the case that one would be given the information that a configuration $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(d)}) \in \mathfrak{S}_{\geq}$ is feasible, the question remains how to construct a corresponding tensor. Note that the suggested strategy to prove Conjecture 3.5 by showing full rank of (2.4) may not provide an explicit way for perturbing singular values in single directions.

A (currently) heuristic approach can be taken via the method of alternating projections. It is based on an alternative viewpoint on Problem 1.4: Given $\boldsymbol{\sigma}^{(j)} \in S_{\geq}^{(j)}$ for $j = 1, \ldots, d$, the configuration $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(d)})$ is feasible, if and only if there exists a tensor $\mathbf{X}$ such that

$$\mathbf{X} \in \bigcap_{j=1}^{d} \mathscr{M}_{\boldsymbol{\sigma}^{(j)}}^{(j)}, \tag{3.2}$$

where $\mathscr{M}_{\boldsymbol{\sigma}}^{(j)}$ denotes the set of all tensors with mode-j singular values $\boldsymbol{\sigma}$. More concretely,

$$\mathscr{M}_{\boldsymbol{\sigma}}^{(j)} = \left\{\mathbf{X} : M_{\mathbf{X}}^{(j)} = U \, \mathrm{diag}(\boldsymbol{\sigma}) V^{\mathsf{T}}, \ U \in \mathrm{O}(n_j), \ V^{\mathsf{T}} V = I_{n_j}\right\}.$$

The method of alternating projections tries to find $\mathbf{X}$ satisfying (3.2) by successively projecting on the sets $\mathscr{M}_{\boldsymbol{\sigma}^{(j)}}^{(j)}$. It hence takes the form

$$\mathbf{X}_{k+1} = \left(\Pi_{\boldsymbol{\sigma}^{(d)}}^{(d)} \circ \cdots \circ \Pi_{\boldsymbol{\sigma}^{(1)}}^{(1)}\right)(\mathbf{X}_k), \tag{3.3}$$

where $\Pi_{\boldsymbol{\sigma}}^{(j)} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathscr{M}_{\boldsymbol{\sigma}}^{(j)}$ is a metric projection on the set $\mathscr{M}_{\boldsymbol{\sigma}}^{(j)}$, that is, $\Pi_{\boldsymbol{\sigma}}^{(j)}(\mathbf{Y})$ returns a best approximation of $\mathbf{Y}$ in the set $\mathscr{M}_{\boldsymbol{\sigma}}^{(j)}$. A best approximation in

Frobenius norm can be obtained by simply replacing the singular values of $M_{\mathbf{Y}}^{(j)}$ with $\boldsymbol{\sigma}$:

$$M_{\Pi_{\boldsymbol{\sigma}}^{(j)}(\mathbf{Y})}^{(j)} = U_{\mathbf{Y}}^{(j)} \operatorname{diag}(\boldsymbol{\sigma}) (V_{\mathbf{Y}}^{(j)})^{\mathsf{T}}. \tag{3.4}$$

Moreover, if $\boldsymbol{\sigma}_{\mathbf{Y}}^{(j)} \in S_>^{(j)}$, this best approximation is unique. To prove these assertions, note that the best approximation problem in Frobenius norm is equivalent to maximizing the trace of $U \operatorname{diag}(\boldsymbol{\sigma}) V^{\mathsf{T}} (M_{\mathbf{Y}}^{(j)})^{\mathsf{T}}$ over all $U \in O(n_j)$ and $V \in \mathbb{R}^{n_j^c}$, $V^{\mathsf{T}} V = I_{n_j}$. The von Neumann trace inequality [15,23] states that the upper bound for this quantity is $\boldsymbol{\sigma}^{\mathsf{T}} \boldsymbol{\sigma}_{\mathbf{Y}}^{(j)}$. Moreover, equality is achieved at U, V if and only if $M_{\mathbf{Y}}^{(j)} = U \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{Y}}^{(j)}) V^{\mathsf{T}}$, see [4, Remark 1.2]. Hence $U = U_{\mathbf{Y}}^{(j)}$ and $V = V_{\mathbf{Y}}^{(j)}$ are unique in the case that $\boldsymbol{\sigma}_{\mathbf{Y}}^{(j)} \in S_>^{(j)}$.

Although the interpretation as an alternating projection method is nice, we remark that the multiplication by $U_{\mathbf{Y}}^{(j)}$ in (3.4) could be omitted in practice. It is an easy induction to show that in this case an orthogonally equivalent sequence of tensors would be produced.

Even assuming that intersection points exist, we are currently not able to provide local or global convergence results for the alternating projection method (3.3). Instead, we confine ourselves with three numerical illustrations.

Recovering a feasible configuration

To obtain a feasible configuration $\boldsymbol{\Sigma}$, we create a norm-one tensor $\mathbf{X}$ and take its higher-order singular values, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathbf{X}}$. Then we run the iteration (3.3) starting from a random initialization, and measure the errors $\|\boldsymbol{\sigma}_{\mathbf{X}_{k+1}}^{(j)} - \boldsymbol{\sigma}_{\mathbf{X}}^{(j)}\|_2$ (Euclidean norm) after every full cycle of projections. Since $\Pi_{\boldsymbol{\sigma}^{(3)}}^{(3)}$ is applied last, the singular values in direction three are always correct at the time the error is measured. The question is whether also the singular values in the other directions converge to the desired target values. The left plot in Fig. 2 shows one typical example of error curves observed in this kind of experiment in $\mathbb{R}^{30 \times 30 \times 30}$. We see that the sequence $\boldsymbol{\Sigma}_{\mathbf{X}_k}$ converges to $\boldsymbol{\Sigma}$, hence every cluster point of the sequence $\mathbf{X}_k$ will have the desired higher-order singular values. So far, we have no theoretical explanation for the shifted peaks occurring in the curves.

Since our initial guess is random, we do not expect that the generating $\mathbf{X}$ or an orthogonally equivalent tensor will be recovered. To verify this, we make use of Proposition 2.6 and measure the error $\max_{i_1,\dots,i_d} ||\hat{X}_{i_1,\dots,i_d}| - |(\hat{X}_{k+1})_{i_1,\dots,i_d}||$ after every loop, where $\hat{\mathbf{X}}_{k+1}$ and $\hat{\mathbf{X}}$ are HOSVD representatives in the corresponding orbits of orthogonal equivalence. The right plot in Fig. 2 shows this error curve, and we can see it does not tend to zero. By Proposition 2.6, the limiting tensor is hence not orthogonally equivalent to $\mathbf{X}$. Since this behaviour was observed being typical, the alternating projection method can be suggested as a practical procedure to construct tensors having the same higher-order singular values without being orthogonally equivalent.

Experiments with tensors of order $d = 4$ and larger lead to similar results, but they quickly become computationally expensive as SVDs of large matrices have to be calculated.
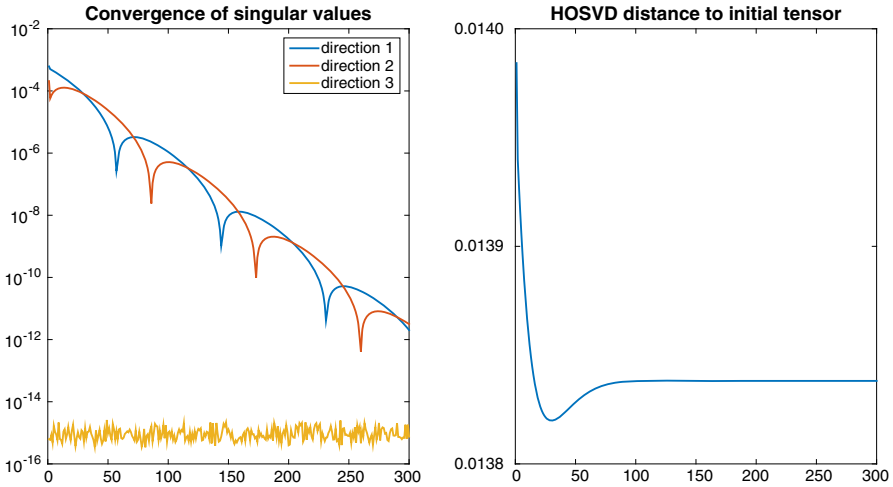
Fig. 2 Recovering a feasible configuration in $\mathbb{R}^{30 \times 30 \times 30}$ via alternating projection. *Left* errors of singular value vectors. *Right* maximum difference between absolute values of entries of HOSVD core tensors of iterates and the generating tensor that provided the feasible configuration. As it does not go to zero, the limiting tensor is not orthogonally equivalent to the generating tensor
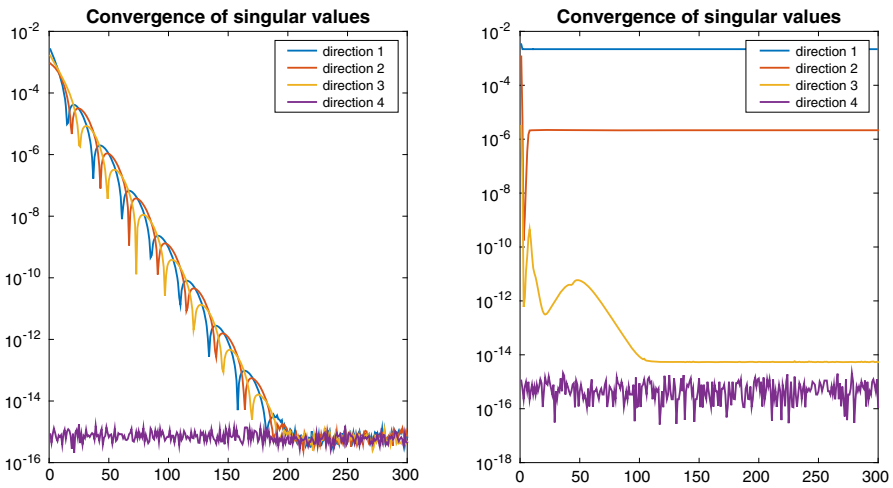

Fig. 3 Experiments in $\mathbb{R}^{10 \times 10 \times 10 \times 10}$. *Left* perturbation of a given feasible configuration $\boldsymbol{\Sigma}$ by $\mathcal{O}(\epsilon)$ with $\epsilon = 10^{-3}$. *Right* infeasible configuration $\sigma^{(j)} = (1, 0, \ldots, 0) + \mathcal{O}(\epsilon_j)$ obtained using $\epsilon_3 = \epsilon_4 = 10^{-6}$, but $\epsilon_1 = \epsilon_2 = 10^{-3}$

Perturbation of a feasible configuration

To support Conjecture 3.5, we now consider random perturbations

$$\boldsymbol{\Sigma}_\epsilon = \left(\boldsymbol{\sigma}_\epsilon^{(1)}, \ldots, \boldsymbol{\sigma}_\epsilon^{(d)} \right) = \boldsymbol{\Sigma} + \mathcal{O}(\epsilon) \in \mathscr{S}_\geq$$

of a known feasible configuration $\boldsymbol{\Sigma} \in \mathfrak{F} \cap \mathfrak{S}_>$ (obtained again from a random tensor $\mathbf{X} \in \mathscr{S}^*$).[3] According to the conjecture, we expect that for small ϵ the configuration $\boldsymbol{\Sigma}_\epsilon$ is also feasible, so a corresponding tensor may be found by the alternating projection method (3.3). This can be verified in numerical experiments. The left plot in Fig. 3 shows the errors $\|\boldsymbol{\sigma}_{\mathbf{X}_{k+1}}^{(j)} - \boldsymbol{\sigma}_\epsilon^{(j)}\|_F$ for one experiment in $\mathbb{R}^{10 \times 10 \times 10 \times 10}$ using $\epsilon = 10^{-3}$.

Infeasible configuration

When conducting our experiments with the alternating projection method, we made the experience that with high probability even a randomly generated configuration will be feasible. Indeed, Fig. 1 supports this in the $2 \times 2 \times 2$ case, as the feasible configurations seem to make up a rather large fraction in $\mathfrak{S}_\geq(2, 2, 2)$.

To construct an infeasible configuration we therefore mimic the proof of Theorem 3.2: we generate $\boldsymbol{\sigma}^{(j)}$ as $(1, 0, \ldots, 0) + \mathscr{O}(\epsilon_j)$ (as described in Footnote 3), where we use very small ϵ_j for $j \geq 3$, e.g., $\epsilon_j = 10^{-6}$. By the arguments presented above this should also enforce $\boldsymbol{\sigma}^{(1)}$ to be close to $\boldsymbol{\sigma}^{(2)}$ to ensure feasibility. To impede this, we use larger ϵ_1 and ϵ_2 instead, e.g., $\epsilon_1 = \epsilon_2 = 10^{-3}$ (an alternative would be to generate $\boldsymbol{\sigma}^{(1)}$ and $\boldsymbol{\sigma}^{(2)}$ completely random). Our results suggest that this indeed results in an infeasible configuration. Accordingly, the alternating projection method fails. The right plot in Fig. 3 shows the outcome of one experiment, again in $\mathbb{R}^{10 \times 10 \times 10 \times 10}$.

References

1. Bachmayr, M., Dahmen, W.: Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. Found. Comput. Math. **15**(4), 839–898 (2015)
2. Bader, B.W., Kolda, T.G., et al.: Matlab tensor toolbox version 2.6 (2015). http://www.sandia.gov/~tgkolda/TensorToolbox/
3. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000, electronic)
4. de Sá, E.M.: Exposed faces and duality for symmetric and unitarily invariant norms. Linear Algebra Appl. **197**(198), 429–450 (1994)
5. de Silva, V., Lim, L.-H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. **30**(3), 1084–1127 (2008)
6. Dieudonné, J.: Treatise on Analysis, vol. III. Academic Press, New York (1972)
7. Grasedyck, L.: Hierarchical singular value decomposition of tensors. SIAM J. Matrix Anal. Appl. **31**(4), 2029–2054 (2009/2010)
8. Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus. Springer, Heidelberg (2012)
9. Hackbusch, W., Kühn, S.: A new scheme for the tensor representation. J. Fourier Anal. Appl. **15**(5), 706–722 (2009)
10. Hitchcock, F.L.: The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. **6**, 164–189 (1927)
11. Hitchcock, F.L.: Multiple invariants and generalized rank of a *p*-way matrix or tensor. J. Math. Phys. **7**, 39–79 (1927)

---

[3] Practically, $\boldsymbol{\Sigma}_\epsilon$ was generated by normalizing and sorting the perturbed $\left|\sigma^{(j)} + \mathscr{O}(\epsilon)\right|$.

12. Koch, O., Lubich, C.: Dynamical tensor approximation. SIAM J. Matrix Anal. Appl. **31**(5), 2360–2375 (2010)
13. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
14. Lubich, C., Rohwedder, T., Schneider, R., Vandereycken, B.: Dynamical approximation by hierarchical Tucker and tensor-train tensors. SIAM J. Matrix Anal. Appl. **34**(2), 470–494 (2013)
15. Mirsky, L.: A trace inequality of John von Neumann. Monatsh. Math. **79**(4), 303–306 (1975)
16. Oseledets, I., Tyrtyshnikov, E.: TT-cross approximation for multidimensional arrays. Linear Algebra Appl. **432**(1), 70–88 (2010)
17. Oseledets, I.V.: Tensor-train decomposition. SIAM J. Sci. Comput. **33**(5), 2295–2317 (2011)
18. Oseledets, I.V., Tyrtyshnikov, E.E.: Breaking the curse of dimensionality, or how to use SVD in many dimensions. SIAM J. Sci. Comput. **31**(5), 3744–3759 (2009)
19. Oseledets, I.V., Tyrtyshnikov, E.E.: Recursive decomposition of multidimensional tensors. Dokl. Akad. Nauk **427**(1), 14–16 (2009, in Russian) [English translation in: Dokl. Math. **80**(1), 460–462 (2009)]
20. Rohwedder, T., Uschmajew, A.: On local convergence of alternating schemes for optimization of convex problems in the tensor train format. SIAM J. Numer. Anal. **51**(2), 1134–1162 (2013)
21. Schneider, R., Uschmajew, A.: Approximation rates for the hierarchical tensor format in periodic Sobolev spaces. J. Complex. **30**(2), 56–71 (2014)
22. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3), 279–311 (1966)
23. von Neumann, J.: Some matrix-inequalities and metrization of matrix-space. Tomsk. Univ. Rev. **1**, 286–300 (1937)