



Truncation of tensors in the hierarchical format

Wolfgang Hackbusch¹

Received: 5 December 2018 / Accepted: 19 December 2018 / Published online: 4 January 2019
© The Author(s) 2019

Abstract

Tensors are in general large-scale data which require a special representation. These representations are also called a format. After mentioning the r -term and tensor subspace formats, we describe the hierarchical tensor format which is the most flexible one. Since operations with tensors often produce tensors of larger memory cost, truncation to reduced ranks is of utmost importance. The so-called higher-order singular-value decomposition (HOSVD) provides a save truncation with explicit error control. The paper explains in detail how the HOSVD procedure is performed within the hierarchical tensor format. Finally, we state special favourable properties of the HOSVD truncation.

Keywords Hierarchical tensor representation · HOSVD truncation

Mathematics Subject Classification 15A69 · 15A18 · 15A99 · 65F99 · 65T99

1 Introduction

In the standard case, tensors of order d are quantities \mathbf{v} indexed by d indices, i.e., the entries of \mathbf{v} are $\mathbf{v}[i_1, \dots, i_d]$, where e.g. all indices run from 1 to n . Hence the data size is n^d . This shows that even number n and d of moderate size yield a huge value so that it is impossible to store all entries. Instead one needs some data-sparse tensor representation. In this paper we mention such representations. The optimal one is the hierarchical representation explained in Chapter 4. A slight generalisation is the tree-based format described in Falcó et al. [3].

Section 2 contains an introduction into tensor spaces and the used notation. We mainly restrict ourselves to the finite-dimensional case, in which we do not have to distinguish between the algebraic and topological tensor spaces. The latter tensor spaces are discussed in [3]. Since true tensor spaces—those of order $d \geq 3$ —have less pleasant properties than matrices (which are tensors of order 2), one tries to interpret tensors as matrices. This leads to the technique of matricisation explained in Sect. 2.2. The range of the obtained matrices defines the minimal subspaces introduced in Sect. 2.3. The dimension of the minimal subspaces yields the associated ranks. The singular-value decomposition applied to the matricisations

✉ Wolfgang Hackbusch
wh@mis.mpg.de

¹ Max-Planck-Institut Mathematik in den Naturwissenschaften, Inselstr. 22, 04103 Leipzig, Germany

leads to the so-called higher-order singular-value decomposition (HOSVD) which will be important later. Finally, in Sect. 2.5, we discuss basis transformations.

In Sect. 3 we briefly discuss two classical representations of tensors: the r -term format (also called CP format) and the tensor subspace format (also called Tucker format). For the latter format the HOSVD is explained in Sect. 3.3: instead of applying HOSVD to the full tensor, we can apply it to the smaller core tensor. As a result of HOSVD we can introduce special HOSVD bases. These bases allow a simple truncation to smaller ranks (i.e., the data sparsity is improved; cf. Sect. 3.4).

Section 4 is devoted to the hierarchical tensor format. In principle it is a recursive application of the tensor subspace format. It is connected with a binary tree. The generalisation to a general tree yields the tree-based format in [3]. However, for practical reasons one should use a binary tree. A key point is the indirect coding of the bases discussed in Sect. 4.2. As a result only transfer matrices are stored instead of large-sized basis vectors. Basis transformations can be performed by simple matrix operations (cf. Sect. 4.3). Similarly, the orthonormalisation of the bases are performed by an orthonormalisation of the transfer matrices (cf. Sect. 4.4). The main challenge is the computation of the HOSVD bases. As demonstrated in Sect. 4.5 one can obtain these bases by singular-value decompositions only involving the transfer matrices. The corresponding truncation can be performed as in the previous chapter (cf. Sect. 4.6).

The SVD truncation to lower ranks can be regarded as a projection onto smaller subspaces. However, different from general projections, the SVD projection has particular properties which are discussed in the final Sect. 5. In Sect. 5.1 we consider the case of the tensor subspace representation of Sect. 3.2. It turns out that certain properties of the given tensor (e.g., side conditions or smoothness properties) are inherited by the projected (truncated) approximation. As proved in Sect. 5.2, the same statement holds for the best approximation in the format of lower ranks. Finally, this statement is generalised to the hierarchical tensor representation.

2 Tensor spaces

2.1 Definitions, notation

Let V_j ($1 \leq j \leq d$) be arbitrary vector spaces over the field \mathbb{K} , where either $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Then the algebraic tensor space $\mathbf{V} := {}_a \otimes_{j=1}^d V_j$ consists of all (finite) linear combinations of *elementary tensors* $\otimes_{j=1}^d v^{(j)}$ ($v^{(j)} \in V_j$). The algebraic definition of \mathbf{V} and of the tensor product $\otimes : V_1 \times \cdots \times V_d \rightarrow \mathbf{V}$ reads as follows (cf. Greub [4, Chap. I, §2]): Let U be any vector space over \mathbb{K} . Then, for any multilinear mapping $\varphi : V_1 \times \cdots \times V_d \rightarrow U$, there exists a unique linear mapping $\Phi : \mathbf{V} \rightarrow U$ such that $\varphi(v^{(1)}, v^{(2)}, \dots, v^{(d)}) = \Phi(\otimes_{j=1}^d v^{(j)})$ for all $v^{(j)} \in V_j$.

In the case of infinite-dimensional tensor spaces, one can equip the tensor space with a norm. The completion with respect to the norm $\|\cdot\|$ yields the topological tensor space $\|\cdot\| \otimes_{j=1}^d V_j$ (cf. Hackbusch [5, § 4]). In this article, we restrict ourselves to the finite-dimensional case. Then the algebraic tensor space introduced above is already complete with respect to any norm and therefore it coincides with the topological tensor space. This fact allows us to avoid the affix ‘a’ in $\mathbf{V} = {}_a \otimes_{j=1}^d V_j$. Instead, $\mathbf{V} = \otimes_{j=1}^d V_j$ is sufficient.

The simplest example of a tensor space is based on the vector spaces $V_j = \mathbb{K}^{n_j}$, where the vectors $v \in \mathbb{K}^{n_j}$ are indexed by $i \in I_j := \{1, \dots, n_j\}$. Instead of \mathbb{K}^{n_j} we also write \mathbb{K}^{I_j} .

Then the elementary product $\mathbf{v} := \bigotimes_{j=1}^d v^{(j)}$ is indexed¹ by d -tuples $\mathbf{i} \in \mathbf{I} := I_1 \times \dots \times I_d$:

$$\mathbf{v}[\mathbf{i}] = \mathbf{v}[i_1, \dots, i_d] = \prod_{j=1}^d v^{(j)}[i_j] \quad \text{for } \mathbf{i} = (i_1, \dots, i_d) \in \mathbf{I}. \tag{1}$$

Therefore the tensor space \mathbf{V} is isomorphic to $\mathbb{K}^{\mathbf{I}}$.

The second example is based on the matrix spaces $U_j = \mathbb{K}^{m_j \times n_j}$. Then the tensor space $\mathbf{U} := \bigotimes_{j=1}^d U_j$ can be interpreted as follows. Set $\mathbf{V} = \bigotimes_{j=1}^d V_j$ with $V_j = \mathbb{K}^{n_j}$ as above, while $\mathbf{W} = \bigotimes_{j=1}^d W_j$ is generated by $W_j = \mathbb{K}^{m_j}$. Matrices $M_j \in U_j$ define a linear map belonging to the vector space $L(V_j, W_j)$. Now the elementary tensor $\mathbf{M} := \bigotimes_{j=1}^d M_j \in \mathbf{U}$ can be regarded as a linear map of $L(\mathbf{V}, \mathbf{W})$ defined by²

$$\mathbf{M} \left(\bigotimes_{j=1}^d v^{(j)} \right) = \bigotimes_{j=1}^d (M_j v^{(j)}) \quad \text{for all } v^{(j)} \in V_j. \tag{2}$$

The tensor product $\bigotimes_{j=1}^d M_j$ of matrices is also called the *Kronecker product*. In the finite-dimensional case, \mathbf{U} coincides with $L(\mathbf{V}, \mathbf{W})$.

The definition of $\mathbf{V} = \bigotimes_{j=1}^d V_j$ by all linear combinations of elementary tensors ensures that any $\mathbf{v} \in \mathbf{V}$ has a representation

$$\mathbf{v} = \sum_{i=1}^r \bigotimes_{j=1}^d v_i^{(j)} \quad (v_i^{(j)} \in V_j). \tag{3}$$

The *tensor rank* of \mathbf{v} is the smallest possible integer r in (3). It is denoted by $\text{rank}(\mathbf{v})$.

If $d = 2$ and $V_j = \mathbb{K}^{n_j}$, the tensor space $\mathbf{V} = V_1 \otimes V_2$ is isomorphic to the matrix space $\mathbb{K}^{n_1 \times n_2}$. The elementary tensor $v \otimes w$ corresponds to the rank-1 matrix vw^T . In this case the tensor rank coincides with the usual matrix rank.

$d = 1$ is the trivial case where $\mathbf{V} = \bigotimes_{j=1}^d V_j$ coincides with V_1 . For $d = 0$, the empty product $\mathbf{V} = \bigotimes_{j=1}^d V_j$ is defined by the underlying field \mathbb{K} .

Remark 1 The dimension of $\mathbf{V} = \bigotimes_{j=1}^d V_j$ is $\dim(\mathbf{V}) = \prod_{j=1}^d \dim(V_j)$. This fact implies that, e.g., $U \otimes V \otimes W$, $(U \otimes V) \otimes W$, $U \otimes (V \otimes W)$, and $(U \otimes W) \otimes V$ are *isomorphic as vector spaces*. Here $(U \otimes V) \otimes W$ is the tensor space of order 2 based on the vector spaces $X := U \otimes V$ and W . However, these spaces are not isomorphic as tensor spaces. For instance they have different elementary tensors. $\mathbf{V} = \bigotimes_{j=1}^d V_j$ and $\mathbf{W} = \bigotimes_{j=1}^d W_j$ are *isomorphic as tensor spaces* if, for all $1 \leq j \leq d$, the vector spaces V_j and W_j are isomorphic.

2.2 Matricisation

Within the theory of tensor spaces, the matrix case corresponding to $d = 2$ is an exceptional case. This means that most of the properties of matrices do not generalise to tensors of order $d \geq 3$. An example is the tensor rank for $d \geq 3$. In general its determination is NP hard (cf. Håstad [9]). Tensors in $\bigotimes_{j=1}^d \mathbb{R}^{n_j}$ can also be regarded as elements in $\bigotimes_{j=1}^d \mathbb{C}^{n_j}$, but the

¹ The indices are written in square brackets since we want to avoid secondary indices.

² Here we use that a linear map is uniquely defined by its action on elementary tensors.

corresponding tensor ranks may be different. Matrix decompositions like the Jordan normal form or the singular-value decomposition do not have an equivalent for $d \geq 3$.

To overcome these difficulties one may try to interpret tensors as matrices. According to Remark 1, for all $1 \leq j \leq d$, the tensor space $\mathbf{V} = \bigotimes_{k=1}^d V_k$ is isomorphic to

$$V_j \otimes V_{[j]}, \quad \text{where } V_{[j]} := \bigotimes_{k \neq j} V_k.$$

Here $k \neq j$ means $k \in \{1, \dots, d\} \setminus \{j\}$. The vector space isomorphism $\mathcal{M}_j : \mathbf{V} = \bigotimes_{k=1}^d V_k \rightarrow V_j \otimes V_{[j]}$ is defined by $\mathcal{M}_j(\bigotimes_{k=1}^d v^{(k)}) \mapsto v^{(j)} \otimes v^{[j]}$ with $v^{[j]} = \bigotimes_{k \neq j} v^{(k)}$ and called the j -th *matricisation*. In the case of $V_k = \mathbb{K}^{I_k}$, the image $\mathcal{M}_j(\mathbf{v})$ of $\mathbf{v} \in \mathbf{V}$ is a matrix $M \in \mathbb{K}^{I_j \times \mathbf{I}_{[j]}}$ with $\mathbf{I}_{[j]} = \times_{k \neq j} I_k$ and the entries

$$M[i_j, \mathbf{i}_{[j]}] = \mathbf{v}[i_1, \dots, i_d]$$

for all $i_j \in I_j$ and $\mathbf{i}_{[j]} = (i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_d) \in \mathbf{I}_{[j]}$. An obvious generalisation reads as follows. Set $D := \{1, \dots, d\}$ and choose a subset $\alpha \subset D$ with $\emptyset \neq \alpha \neq D$. The complement is $\alpha^c := D \setminus \alpha$. Define $\mathbf{I}_\alpha := \times_{j \in \alpha} I_j$ and $\mathbf{V}_\alpha = \bigotimes_{j \in \alpha} V_j$. The matricisation with respect to α uses the isomorphism

$$\mathcal{M}_\alpha : \mathbf{V} \rightarrow \mathbf{V}_\alpha \otimes \mathbf{V}_{\alpha^c}, \quad \mathcal{M}_\alpha(\mathbf{v})[\mathbf{i}_\alpha, \mathbf{i}_{\alpha^c}] = \mathbf{v}[\mathbf{i}] \quad \text{for } \mathbf{i} = (\mathbf{i}_\alpha, \mathbf{i}_{\alpha^c}) \in \mathbf{I}_D. \tag{4}$$

Here \mathbf{i}_α is the tuple $(i_j)_{j \in \alpha} \in \mathbf{I}_\alpha$. $\mathcal{M}_\alpha(\mathbf{v})$ can be regarded as a matrix in $\mathbb{K}^{\mathbf{I}_\alpha \times \mathbf{I}_{\alpha^c}}$. For $\alpha = \{j\}$ we obtain the j -th matricisation \mathcal{M}_j from above. For $\alpha = D$, the set α^c is empty. The formal definition $\bigotimes_{j \in \emptyset} V_j = \mathbb{K}$ explains that $\mathcal{M}_D : \mathbf{V} \rightarrow \mathbf{V} \otimes \mathbb{K}$. Regarding $\mathcal{M}_D(\mathbf{v})$ as a matrix means that there is only one column containing the vectorised tensor \mathbf{v} . Analogously, $\mathcal{M}_\emptyset(\mathbf{v}) = \mathcal{M}_D(\mathbf{v})^\top$ contains \mathbf{v} as row vector.

The α -rank of a tensor \mathbf{v} is already defined by Hitchcock [10, p. 170] via the matrix rank of $\mathcal{M}_\alpha(\mathbf{v})$:

$$\text{rank}_\alpha(\mathbf{v}) := \text{rank}(\mathcal{M}_\alpha(\mathbf{v})).$$

For different α the α -ranks are different. The only relations are

$$\begin{aligned} \text{rank}_\alpha(\mathbf{v}) &= \text{rank}_{\alpha^c}(\mathbf{v}), \\ \text{rank}_\alpha(\mathbf{v}) &\leq \text{rank}_{\alpha'}(\mathbf{v}) \text{rank}_{\alpha''}(\mathbf{v}) \quad \text{for } \alpha = \alpha' \dot{\cup} \alpha'' \end{aligned}$$

(cf. Hackbusch [5, Lemma 6.19], $\dot{\cup}$ is the disjoint union). The connection with the tensor rank introduced in Sect. 2.1 is

$$\text{rank}_\alpha(\mathbf{v}) \leq \text{rank}(\mathbf{v})$$

(cf. Hackbusch [5, Remark 6.21]). For $\alpha = \{j\}$ we write $\text{rank}_j(\mathbf{v})$. The tuple $(\text{rank}_1(\mathbf{v}), \dots, \text{rank}_d(\mathbf{v}))$ is also called the *multilinear rank* of \mathbf{v} .

Let $\mathbf{M} := \bigotimes_{j=1}^d M^{(j)}$ be an elementary Kronecker product of matrices $M^{(j)}$. The tensor $\mathbf{M}\mathbf{v}$ is defined by (2) and satisfies

$$\mathcal{M}_\alpha(\mathbf{M}\mathbf{v}) = (\mathbf{M}_\alpha \otimes \mathbf{M}_{\alpha^c}) \mathcal{M}_\alpha(\mathbf{v}), \tag{5a}$$

where $\mathbf{M}_\alpha = \bigotimes_{j \in \alpha} M^{(j)}$ and $\mathbf{M}_{\alpha^c} = \bigotimes_{j \in \alpha^c} M^{(j)}$ are partial Kronecker products. Interpreting $\mathcal{M}_\alpha(\mathbf{v})$ and $\mathcal{M}_\alpha(\mathbf{M}\mathbf{v})$ as matrices, the equivalent statement is

$$\mathcal{M}_\alpha(\mathbf{M}\mathbf{v}) = \mathbf{M}_\alpha \mathcal{M}_\alpha(\mathbf{v}) \mathbf{M}_{\alpha^c}^\top. \tag{5b}$$

2.3 Minimal subspaces

Given $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^d V_j$, there may be smaller subspaces $U_j \subset V_j$ such that $\mathbf{v} \in \mathbf{U} = \bigotimes_{j=1}^d U_j$. The subspaces of minimal dimension are called the *minimal subspaces* and denoted by $U_j^{\min}(\mathbf{v})$. They satisfy

$$\mathbf{v} \in \bigotimes_{j=1}^d U_j^{\min}(\mathbf{v}) \tag{6}$$

while $\mathbf{v} \in \bigotimes_{j=1}^d U_j$ implies $U_j^{\min}(\mathbf{v}) \subset U_j$.

The generalisation to subsets $\alpha \subset D = \{1, \dots, d\}$ uses the isomorphism $\mathbf{V} \simeq \mathbf{V}_\alpha \otimes \mathbf{V}_{\alpha^c}$. $\mathbf{U}_\alpha^{\min}(\mathbf{v}) \subset \mathbf{V}_\alpha$ is defined as the subspace of minimal dimension such that $\mathbf{v} \in \mathbf{U}_\alpha^{\min}(\mathbf{v}) \otimes \mathbf{V}_{\alpha^c}$. The minimal subspaces can be characterised by

$$\mathbf{U}_\alpha^{\min}(\mathbf{v}) := \text{range}(\mathcal{M}_\alpha(\mathbf{v})). \tag{7}$$

This includes the case $\alpha = \{j\}$, for which $\mathbf{U}_\alpha^{\min}(\mathbf{v})$ is written as $U_j^{\min}(\mathbf{v})$.

In the infinite-dimensional case one cannot interpret $\mathcal{M}_\alpha(\mathbf{v})$ as a matrix. Then the definition (6) must be replaced by

$$\mathbf{U}_\alpha^{\min}(\mathbf{v}) := \left\{ \phi_{\alpha^c}(\mathbf{v}) : \phi_{\alpha^c} \in {}_a \bigotimes_{j \in \alpha^c} V_j' \right\}, \tag{8}$$

where V_j' is the dual space³ of V_j . The application of $\phi_{\alpha^c} = \bigotimes_{j \in \alpha^c} \varphi^{(j)}$ to \mathbf{v} is defined by

$$\phi_{\alpha^c} \left(\bigotimes_{j=1}^d v^{(j)} \right) := \left(\prod_{j \in \alpha^c} \varphi^{(j)}(v^{(j)}) \right) \bigotimes_{j \in \alpha} v^{(j)} \in \mathbf{V}_\alpha.$$

In the general case the α -rank is defined by $\text{rank}_\alpha(\mathbf{v}) = \dim(\mathbf{U}_\alpha^{\min}(\mathbf{v}))$.

An important property is that under natural conditions weak convergence $\mathbf{v}_n \rightarrow \mathbf{v}$ implies

$$\text{rank}_\alpha(\mathbf{v}) \leq \liminf_{n \rightarrow \infty} \text{rank}_\alpha(\mathbf{v}_n) \tag{9}$$

(cf. Hackbusch [5, Theorem 6.24], Falcó–Hackbusch [2]).

2.4 Higher-order singular-value decomposition (HOSVD)

In the following we assume that all V_j are pre-Hilbert spaces equipped with the Euclidean scalar product $\langle \cdot, \cdot \rangle$. The Euclidean scalar product in $\mathbf{V}_\alpha = \bigotimes_{j \in \alpha} V_j$ satisfies

$$\left\langle \bigotimes_{j \in \alpha} v^{(j)}, \bigotimes_{j \in \alpha} w^{(j)} \right\rangle = \prod_{j \in \alpha} \langle v^{(j)}, w^{(j)} \rangle.$$

Interpreting $\mathcal{M}_\alpha(\mathbf{v})$ as a matrix, one may determine its singular-value decomposition (SVD) $\sum_i \sigma_i^{(\alpha)} \mathbf{b}_i^{(\alpha)} (\mathbf{b}_i^{(\alpha^c)})^\top$. The tensor representation is

$$\mathcal{M}_\alpha(\mathbf{v}) = \sum_{i=1}^{r_\alpha} \sigma_i^{(\alpha)} \mathbf{b}_i^{(\alpha)} \otimes \mathbf{b}_i^{(\alpha^c)}, \tag{10}$$

³ In the case of infinite-dimensional normed spaces one has to distinguish the algebraic dual space V_j' from the space V_j^* of *continuous* linear functionals. Definition (8) holds with V_j' as well as with V_j^* .

where $r_\alpha = \text{rank}_\alpha(\mathbf{v})$ is the α -rank and $\sigma_1^{(\alpha)} \geq \sigma_2^{(\alpha)} \geq \dots \geq \sigma_{r_\alpha}^{(\alpha)} > 0$ are the singular values, while $\{\mathbf{b}_i^{(\alpha)} : 1 \leq i \leq r_\alpha\} \subset \mathbf{V}_\alpha$ and $\{\mathbf{b}_i^{(\alpha^c)} : 1 \leq i \leq r_\alpha\} \subset \mathbf{V}_{\alpha^c}$ are the orthonormal systems of the left and right singular vectors. De Lathauwer–De Moor–Vandewalle [1] introduced the name HOSVD for the simultaneous SVD of the matricisations $\mathcal{M}_j(\mathbf{v})$, $1 \leq j \leq d$. Note that in general the SVD spectra $(\sigma_i^{(j)})_{1 \leq i \leq r_j}$ as well as $r_j = \text{rank}_j(\mathbf{v})$ do not coincide. Compare also Hackbusch–Uchmajew [8].

It will turn out that the important quantities in (10) are the singular values $\sigma_i^{(\alpha)}$ and the left singular vectors $\mathbf{b}_i^{(\alpha)}$. These quantities are also characterised by the diagonalisation of the matrix $\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H$:

$$\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H = \sum_{i=1}^{r_\alpha} (\sigma_i^{(\alpha)})^2 \mathbf{b}_i^{(\alpha)} (\mathbf{b}_i^{(\alpha)})^H.$$

In the case of $\alpha = \{j\}$, $\mathcal{M}_j(\mathbf{v})$ is a matrix with $n_j = \dim(V_j)$ rows and $n_{[j]} = \prod_{k \neq j} \dim(V_k)$ columns. Note that $n_{[j]}$ may be a huge quantity. However, $\mathcal{M}_j(\mathbf{v})\mathcal{M}_j(\mathbf{v})^H$ is only of the size $n_j \times n_j$.

2.5 Basis representations, transformations

The notation (1) refers to the unit vectors $e_i^{(j)}$ ($1 \leq i \leq n_j$) of $V_j = \mathbb{K}^{n_j}$, i.e., the tensor is $\mathbf{v} = \sum_{i_1=1}^{n_1} \dots \sum_{i_d=1}^{n_d} \mathbf{v}[i_1, \dots, i_d] \otimes_{j=1}^d e_{i_j}^{(j)}$. We may choose another basis $b_i^{(j)}$ ($1 \leq i \leq n_j$) of $V_j = \mathbb{K}^{n_j}$ and obtain

$$\mathbf{v} = \sum_{i_1=1}^{n_1} \dots \sum_{i_d=1}^{n_d} \mathbf{c}[i_1, \dots, i_d] \otimes_{j=1}^d b_{i_j}^{(j)} \tag{11}$$

with another coefficient tensor $\mathbf{c} \in \mathbf{V} := \otimes_{j=1}^d \mathbb{K}^{n_j}$. The basis $b_i^{(j)}$ ($1 \leq i \leq n_j$) yields the regular matrix $B_j = [b_1^{(j)}, \dots, b_{n_j}^{(j)}]$. Forming the Kronecker product $\mathbf{B} := \otimes_{j=1}^d B_j \in L(\mathbf{V}, \mathbf{V})$, Eq. (11) becomes

$$\mathbf{v} = \mathbf{B}\mathbf{c}.$$

If B_j and B'_j are two bases of V_j , there are transformations $T^{(j)}$ and $S^{(j)} = (T^{(j)})^{-1}$ with

$$B_j = B'_j T^{(j)} \text{ and } B'_j = B_j S^{(j)}, \text{ i.e., } b_i^{(j)} = \sum_{k=1}^{r_j} T_{ki}^{(j)} b'_k{}^{(j)}.$$

Form $\mathbf{T} := \otimes_{j=1}^d T^{(j)}$ and $\mathbf{S} := \otimes_{j=1}^d S^{(j)} = \mathbf{T}^{-1}$. Then

$$\mathbf{v} = \mathbf{B}\mathbf{c} = \mathbf{B}'\mathbf{c}' \text{ holds with } \mathbf{B} = \mathbf{B}'\mathbf{T}, \mathbf{B}' = \mathbf{B}\mathbf{S}, \mathbf{c} = \mathbf{S}\mathbf{c}', \mathbf{c}' = \mathbf{T}\mathbf{c}. \tag{12}$$

Remark 2 According to (5b), the matricisations of \mathbf{v} and its core tensor \mathbf{c} are related by $\mathcal{M}_\alpha(\mathbf{v}) = \mathbf{B}_\alpha \mathcal{M}_\alpha(\mathbf{c}) \mathbf{B}_{\alpha^c}^\top$ with $\mathbf{B}_\alpha := \otimes_{j \in \alpha} B_j$.

3 Tensor representations

3.1 *r*-Term format

Often, the dimension $\prod_{j=1}^d n_j$ of $\otimes_{j=1}^d \mathbb{K}^{n_j}$ is much larger than the available computer memory. Therefore a naive representation of a tensor via its entries (1) is impossible. A classical tensor representation is the *r-term format* (also called the canonical or CP format) related to (3). Let $\mathbf{V} = \otimes_{j=1}^d V_j$. We fix an integer $r \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and define the set

$$\mathcal{R}_r := \left\{ \mathbf{v} \in \mathbf{V} : \mathbf{v} = \sum_{i=1}^r \bigotimes_{j=1}^d v_i^{(j)}, v_i^{(j)} \in V_j \right\},$$

i.e., \mathbf{v} is represented by r elementary tensors with the factors $v_i^{(j)}$. Assuming $\dim(V_j) = n_j \leq n$, the memory cost of $\mathbf{v} \in \mathcal{R}_r$ is rnd (unit: numbers in \mathbb{K}).

One checks that $\mathcal{R}_r = \{ \mathbf{v} : \text{rank}(\mathbf{v}) \leq r \}$. As long as $\text{rank}(\mathbf{v}) \leq r$ holds with r of moderate size, this format yields a suitable representation. If $\text{rank}(\mathbf{v})$ is too large, one may try to find an approximating tensor \mathbf{v}' of smaller rank. Another question is the implementation of tensor operations within this format. Adding $\mathbf{u} \in \mathcal{R}_r$ and $\mathbf{v} \in \mathcal{R}_s$, one obtains the representation of the sum $\mathbf{w} := \mathbf{u} + \mathbf{v}$ in \mathcal{R}_{r+s} . Other operations let the representation rank increase even more. An example is the multiplication of the Kronecker matrix $\mathbf{M} := \sum_{i=1}^{\rho} \otimes_{j=1}^d M_i^{(j)}$ by a tensor $v \in \mathcal{R}_r$. The product belongs to $\mathcal{R}_{r \cdot s}$. Therefore one needs a truncation procedure which approximates a tensor from \mathcal{R}_t (t too large) by an approximation in \mathcal{R}_r for a suitable $r < t$. Unfortunately, this task is rather difficult within the r -term format (cf. Hackbusch [5, §7, §9]).

3.2 Tensor-subspace format

A remedy is the Tucker format or *tensor-subspace format*, which is related to (6) and (11). Let $n_j = \dim(V_j)$. Assume that we know that $\mathbf{v} \in \otimes_{j=1}^d U_j$ holds for subspaces $U_j \subset V_j$ of (hopefully much) smaller dimension than n_j . Choose any basis (or even only a generating system) $b_i^{(j)}$ ($1 \leq i \leq r_j$) of U_j , i.e.,

$$U_j = \text{span}\{b_i^{(j)} : 1 \leq i \leq r_j\}. \tag{13a}$$

Then there is a tensor $\mathbf{c} \in \otimes_{j=1}^d \mathbb{K}^{r_j}$ – the so-called *core tensor*—such that

$$\mathbf{v} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{c}[i_1, \dots, i_d] \bigotimes_{j=1}^d b_{i_j}^{(j)}. \tag{13b}$$

Note the difference to (11). The sums in (11) have n_j terms, whereas (13b) only uses $r_j < n_j$ as upper bound.

Definition 1 We denote the set of all tensors in \mathbf{V} with a representation (13b) by $\mathcal{T}_{\mathbf{r}}$, where $\mathbf{r} = (r_1, \dots, r_d)$ is a multi-index.

Remark 3 The optimal choice of U_j is given by $U_j = U_j^{\min}(\mathbf{v})$ (cf. Sect. 2.3), since then $r_j = \text{rank}_j(\mathbf{v})$ is minimal. The memory cost for the core tensor is $\prod_{j=1}^d r_j$. Therefore this representation is unfavourable for large d .

Build the (rectangular) matrices $B_j = [b_1^{(j)}, \dots, b_{r_j}^{(j)}] \in \mathbb{K}^{n_j \times r_j}$ and the Kronecker product $\mathbf{B} := \bigotimes_{j=1}^d B_j \in L(\bigotimes_{j=1}^d \mathbb{K}^{r_j}, \mathbf{V})$ as in Sect. 2.5. Then (13b) is equivalent to

$$\mathbf{v} = \mathbf{B}\mathbf{c}. \tag{13c}$$

3.3 HOSVD

The first step are transformations into *orthonormal* bases B'_j with $B_j = B'_j T_j$ (e.g., by a QR decomposition of B_j yielding $B'_j = Q$ and $T_j = R$). According to (12), we have $\mathbf{v} = \mathbf{B}'\mathbf{c}'$ with $\mathbf{c}' := \mathbf{T}\mathbf{c}$. Denoting \mathbf{B}' and \mathbf{c}' again by \mathbf{B} and \mathbf{c} , we obtain the representation (13b) with orthonormal bases $(b_i^{(j)})_{1 \leq i \leq r_j}$.

The second step is the HOSVD applied to the core tensor $\mathbf{c} \in \bigotimes_{j=1}^d \mathbb{K}^{r_j}$. Assume that the matricisation $C_\alpha := \mathcal{M}_\alpha(\mathbf{c})$ has the singular-value decomposition $C_\alpha = X_\alpha \Sigma_\alpha Y_\alpha^T$ with diagonal Σ_α and unitary matrices X_α, Y_α . By Remark 2 the matricisation of \mathbf{v} is $\mathcal{M}_\alpha(\mathbf{v}) = \mathbf{B}_\alpha \mathcal{M}_\alpha(\mathbf{c}) \mathbf{B}_{\alpha^c}^T = \mathbf{B}_\alpha X_\alpha \Sigma_\alpha Y_\alpha^T \mathbf{B}_{\alpha^c}^T$. This is the singular-value decomposition of $\mathcal{M}_\alpha(\mathbf{v})$ with the unitary matrices $\mathbf{B}_\alpha X_\alpha$ and $\mathbf{B}_{\alpha^c} Y_\alpha$. Taking $\alpha = \{j\}$, we obtain a new basis transform by $B'_j := B_j X_j$ ($1 \leq j \leq d$). The new basis $(b_i^{(j)})_{1 \leq i \leq r_j}$ is called the j -th HOSVD basis. The core tensor has to be transformed into \mathbf{c}' as above. Again denoting \mathbf{B}' and \mathbf{c}' by \mathbf{B} and \mathbf{c} , we obtain the representation (13b) with respect to the HOSVD bases.

Since we do not need the right singular vectors in Y_α , the practical computation first forms the product $P_j := \mathcal{M}_j(\mathbf{c})\mathcal{M}_j(\mathbf{c})^H \in \mathbb{K}^{r_j \times r_j}$. This is the most expensive step with an arithmetic cost of $\mathcal{O}((\sum_{j=1}^d r_j) \prod_{j=1}^d r_j)$. The second step is the singular-value decomposition of P_j (cost: $\mathcal{O}(\sum_{j=1}^d r_j^3)$).

The representation of $\mathbf{v} \in \mathcal{T}_r$ by the HOSVD bases allows two types of truncations. The number $r_j = \dim(U_j)$ may be larger than necessary, i.e., larger than $\text{rank}_j(\mathbf{v}) = \dim(U_j^{\min}(\mathbf{v}))$. This is detected by vanishing singular values. Assume that $\sigma_{s_j}^{(j)} > 0$, whereas $\sigma_i^{(j)} = 0$ for $s_j < i \leq r_j$. Then the sums in (13b) can be shortened (replace r_j by s_j). After this step, $\mathbf{v} \in \bigotimes_{j=1}^d U_j \subset \mathcal{T}_s$ holds with $U_j = U_j^{\min}(\mathbf{v})$ and $s_j = \text{rank}_j(\mathbf{v})$. Note that the described procedure yields a shorter representation while the tensor is unchanged.

A truncation changing the tensor is described next.

3.4 HOSVD truncation

Assume again that the representation (13b) of $\mathbf{v} \in \mathcal{T}_r$ uses the HOSVD bases. We are looking for an approximation $\mathbf{u} \in \mathcal{T}_s$ with smaller dimensions $s_j < r_j$ of the corresponding subspaces U_j . This problem has two answers. First there is a (not necessarily unique) best approximation $\mathbf{u}_{\text{best}} \in \mathcal{T}_s$ with

$$\|\mathbf{v} - \mathbf{u}_{\text{best}}\| = \inf\{\|\mathbf{v} - \mathbf{u}\| : \mathbf{u} \in \mathcal{T}_s\}$$

($\|\cdot\|$ is the Euclidean norm). The computation must be done iteratively. It is hard to ensure that the corresponding minimisation method converges to the *global* minimum, since there may be many local minima.

A much easier approach is the HOSVD truncation: Given $\mathbf{v} \in \mathcal{T}_r$ with HOSVD bases in (13b), omit all terms involving indices $i_j > s_j$. The other terms are unchanged. Obviously, the resulting tensor $\mathbf{u}_{\text{HOSVD}}$ belongs to \mathcal{T}_s and its computation requires no arithmetical operation.

In the case of matrices one knows that $\mathbf{u}_{\text{HOSVD}} = \mathbf{u}_{\text{best}}$. However, for $d \geq 3$, $\mathbf{u}_{\text{HOSVD}}$ is not necessarily the best, but the quasi-optimal approximation:

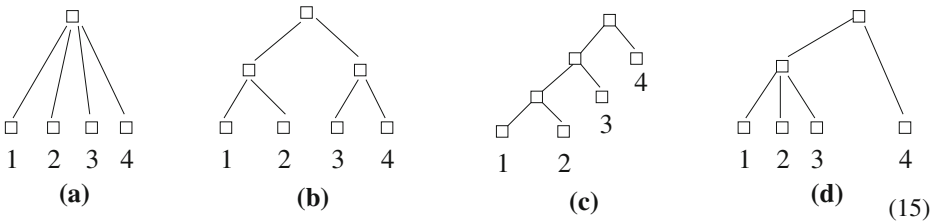
$$\|\mathbf{v} - \mathbf{u}_{\text{HOSVD}}\| \leq \sqrt{\sum_{j=1}^d \sum_{i=s_j+1}^{r_j} (\sigma_i^{(j)})^2} \leq \sqrt{d} \|\mathbf{v} - \mathbf{u}_{\text{best}}\| \tag{14}$$

(cf. [5, Theorem 10.3]). Since the singular values $\sigma_i^{(j)}$ are known, the first inequality in (14) yields a precise error estimate. Given a tolerance ε , one can choose r_j such that the error is below ε . The second inequality proves quasi-optimality.

4 The hierarchical tensor format

4.1 Definition, notation

The *tree-based tensor formats* use a so-called dimension partition tree T_D . The root of the tree is $D = \{1, \dots, d\}$, while the leaves are $\{1\}, \dots, \{d\}$. The tree describes how D is divided recursively. The vertices of the tree are subsets of D . Either a vertex α is a singleton (and therefore a leaf) or it has sons α_i with the property that α is the disjoint union of the α_i . Examples for $d = 4$ are given below:



The first interpretation is that the tree (a) corresponds to $V_1 \otimes V_2 \otimes V_3 \otimes V_4$, (b) to the isomorphic space $(V_1 \otimes V_2) \otimes (V_3 \otimes V_4)$, (c) to $((V_1 \otimes V_2) \otimes V_3) \otimes V_4$, and (d) to $(V_1 \otimes V_2 \otimes V_3) \otimes V_4$.

The second interpretation involves the associated subspaces. The tree (a) corresponds to the Tucker format in Sect. 3.2: All subspaces U_1, \dots, U_d are joined into $U_1 \otimes U_2 \otimes U_3 \otimes U_4$. In the case of tree (b) one first forms the subspaces $U_1 \otimes U_2$ and $U_3 \otimes U_4$ and determines subspaces $U_{\{1,2\}} \subset U_1 \otimes U_2$ and $U_{\{3,4\}} \subset U_3 \otimes U_4$. Finally $U_{\{1,2\}} \otimes U_{\{3,4\}}$ is defined. The trees (c) and (d) lead to analogous constructions. The final subspace \mathbf{U}_D must be such that $\mathbf{v} \in \mathbf{U}_D$ holds for the tensor \mathbf{v} which we want to represent. Obviously, the one-dimensional subspace $\mathbf{U}_D = \text{span}\{\mathbf{v}\}$ is sufficient.

Restricting ourselves to binary trees T_D , we obtain the *hierarchical tensor format* (cases (b), (c) in (15); cf. Hackbusch–Kühn [7]). The practical advantage of a binary tree is the fact that the quantities appearing in the later computations are matrices. The further restriction to linear trees as in case (c) of (15) leads to the so-called *TT format* or *matrix product format* (cf. Verstraete–Cirac [14], Oseledets–Tyrtyshnikov [11,12]).

Consider a vertex $\alpha \subset D$ of the binary tree T_D together with its sons α_1 and α_2 :

$$\begin{array}{c} \alpha \rightarrow \mathbf{U}_\alpha \subset \mathbf{U}_{\alpha_1} \otimes \mathbf{U}_{\alpha_2} \\ \swarrow \quad \searrow \\ \mathbf{U}_{\alpha_1} \leftarrow \alpha_1 \quad \alpha_2 \rightarrow \mathbf{U}_{\alpha_2} \end{array} \tag{16}$$

The sons are associated with subspaces $U_{\alpha_i} \subset V_{\alpha_i} = \bigotimes_{j \in \alpha_i} V_j$ ($i = 1, 2$), while $U_\alpha \subset U_{\alpha_1} \otimes U_{\alpha_2} \subset V_\alpha$ is the characteristic property of U_α . If $\alpha = D$, $\mathbf{v} \in U_D$ is required. Therefore we can choose

$$U_D = \text{span}\{\mathbf{v}\}. \tag{17}$$

The minimal subspaces $U_\alpha^{\min}(\mathbf{v})$ introduced in (7) and (8) satisfy⁴ the inclusion

$$U_\alpha^{\min}(\mathbf{v}) \subset U_{\alpha_1}^{\min}(\mathbf{v}) \otimes U_{\alpha_2}^{\min}(\mathbf{v}).$$

This proves the following remark.

Remark 4 The existence of subspaces U_α ($\alpha \in T_D$) with the required properties is ensured by the optimal choice $U_\alpha = U_\alpha^{\min}(\mathbf{v})$. Vice versa, $U_\alpha \supset U_\alpha^{\min}(\mathbf{v})$ holds for all subspaces U_α satisfying $\mathbf{v} \in U_D$ und $U_\alpha \subset U_{\alpha_1} \otimes U_{\alpha_2}$.

4.2 Implementation of the subspaces

In principle, all subspaces U_α ($\alpha \in T_D$) are described by basis⁵ vectors:

$$U_\alpha = \text{span}\{\mathbf{b}_\ell^{(\alpha)} : 1 \leq \ell \leq r_\alpha\}.$$

However, $\mathbf{b}_\ell^{(\alpha)}$ are already tensors of order $\#\alpha$ which should not be stored explicitly. Therefore we distinguish two cases.

Case A. $\alpha = \{j\}$ is a leaf. Then the basis vectors $b_i^{(j)}$ of $U_\alpha = U_j$ are stored explicitly.

Case B. α is a non-leaf vertex with sons α_1 and α_2 . Note that $\{\mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_j^{(\alpha_2)} : 1 \leq i \leq r_{\alpha_1}, 1 \leq j \leq r_{\alpha_2}\}$ is a basis of $U_{\alpha_1} \otimes U_{\alpha_2}$. The inclusion $U_\alpha \subset U_{\alpha_1} \otimes U_{\alpha_2}$ implies that the basis vector $\mathbf{b}_\ell^{(\alpha)} \in U_\alpha$ must have a representation⁶

$$\mathbf{b}_\ell^{(\alpha)} = \sum_{i,j} c_{ij}^{(\alpha,\ell)} \mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_j^{(\alpha_2)} \tag{18}$$

with coefficients $c_{ij}^{(\alpha,\ell)}$ forming an $r_{\alpha_1} \times r_{\alpha_2}$ matrix

$$C^{(\alpha,\ell)} = (c_{ij}^{(\alpha,\ell)}) \quad (1 \leq \ell \leq r_\alpha). \tag{19}$$

The tuple $(C^{(\alpha,\ell)})_{1 \leq \ell \leq r_\alpha}$ of matrices can be regarded as a tensor \mathbf{C}_α of order 3 with entries $\mathbf{C}_\alpha[i, j, \ell] = C_{ij}^{(\alpha,\ell)}$. In the case of $\alpha = D$, $r_D = 1$ holds. The desired representation of the tensor \mathbf{v} is

$$\mathbf{v} = c_1^{(D)} \mathbf{b}_1^{(D)}. \tag{20}$$

Remark 5 The representation of a tensor \mathbf{v} by the hierarchical format uses the data $b_i^{(j)}$ ($1 \leq j \leq d, 1 \leq i \leq r_j$), $C^{(\alpha,\ell)}$ ($1 \leq \ell \leq r_\alpha, \alpha$ non-leaf vertex of T_D), and $c_1^{(D)}$. The memory cost of the hierarchical format is bounded by $dnr + (d - 1)r^3 + 1$, where $n := \max_j \dim(V_j)$ and $r := \max_{\alpha \in T_D} r_\alpha$.

⁴ This statement also holds for more than two sons. Hence, the construction (16) and Remark 4 can be generalised to any tree-based format.

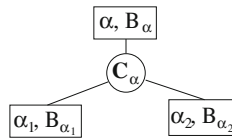
⁵ More generally, the basis vectors may be replaced by any set of spanning vectors.

⁶ In the case of a general tree-based format with δ_α sons α_i ($1 \leq i \leq \delta_\alpha$) of α , Eq. (18) becomes $\mathbf{b}_\ell^{(\alpha)} = \sum_{i_1, \dots, i_{\delta_\alpha}} c^{(\alpha,\ell)}[i_1, \dots, i_{\delta_\alpha}] \otimes_{j=1}^{\delta_\alpha} \mathbf{b}_{i_j}^{(\alpha_j)}$ (cf. (13b)).

Although the representation of \mathbf{v} by the quantities $(b_i^{(j)}, \mathbf{C}_\alpha, c_1^{(D)})$ is rather indirect, all tensor operations can be performed by a recursion in the tree T_D (either from the leaves to the root or in the opposite direction). Below we describe transformations, the orthonormalisation of the bases, and the HOSVD computation. Concerning other operations we refer to [5, § 13].

4.3 Transformations

We recall that the bases $\{\mathbf{b}_\ell^{(\alpha)} : 1 \leq \ell \leq r_\alpha\}$ are well-defined by (18), but they are not directly accessible except for leaves $\alpha \in T_D$. Transformations of the bases are described by the corresponding modifications of the matrices $C^{(\alpha,\ell)}$. As in Sect. 3.2 we form matrices $\mathbf{B}_\alpha = [\mathbf{b}_1^{(\alpha)} \dots \mathbf{b}_{r_\alpha}^{(\alpha)}]$ related to a linear map in $L(\mathbb{K}^{r_\alpha}, \mathbf{V}_\alpha)$. For simplicity we will call \mathbf{B}_α the basis (of the spanned subspace).



The left figure illustrates the connection of the basis \mathbf{B}_α with \mathbf{B}_{α_1} and \mathbf{B}_{α_2} at the son vertices via the data \mathbf{C}_α . Whenever one of these bases changes, also \mathbf{C}_α must be updated. Eq. (21) describes the update caused by a transformation of \mathbf{B}_α , while (22) considers the transformations of \mathbf{B}_{α_1} and \mathbf{B}_{α_2} .

Basis transformation in α . Assume that α is not a leaf and that \mathbf{B}_α and \mathbf{B}'_α are two bases related by $\mathbf{B}'_\alpha = \mathbf{B}_\alpha S^{(\alpha)}$, i.e., $\mathbf{b}'_k^{(\alpha)} = \sum_{j=1}^{r_\alpha} S_{jk}^{(\alpha)} \mathbf{b}_j^{(\alpha)}$ ($1 \leq k \leq r'_\alpha$). The corresponding coefficient matrices $C^{(\alpha,\ell)}$ and $C'^{(\alpha,\ell)}$ satisfy

$$C'^{(\alpha,k)} = \sum_{j=1}^{r_\alpha} S_{jk}^{(\alpha)} C^{(\alpha,j)} \quad (1 \leq k \leq r'_\alpha). \tag{21}$$

Using the tensor \mathbf{C}_α , this transformation becomes $\mathbf{C}'_\alpha = (I \otimes I \otimes (S^{(\alpha)})^\top) \mathbf{C}_\alpha$.

Basis transformation in the son vertices α_i . Let α_1, α_2 be the sons of α . Let \mathbf{B}_{α_i} and \mathbf{B}'_{α_i} be two bases related by $\mathbf{B}'_{\alpha_i} T^{(\alpha_i)} = \mathbf{B}_{\alpha_i}$ ($i = 1, 2$). The corresponding coefficient matrices $C^{(\alpha,\ell)}$ and $C'^{(\alpha,\ell)}$ are related by

$$C'^{(\alpha,\ell)} = T^{(\alpha_1)} C^{(\alpha,\ell)} (T^{(\alpha_2)})^\top \quad \text{for } 1 \leq \ell \leq r_\alpha. \tag{22}$$

This is equivalent to $\mathbf{C}'_\alpha = (T^{(\alpha_1)} \otimes T^{(\alpha_2)} \otimes I) \mathbf{C}_\alpha$.

4.4 Orthonormalisation

Orthonormality of the (non-accessible) bases $\{\mathbf{b}_\ell^{(\alpha)}\}$ can be checked by corresponding properties of the coefficient matrices $C^{(\alpha,\ell)}$. The following sufficient condition is easy to prove.

Remark 6 Let α be a non-leaf vertex. The basis $\{\mathbf{b}_\ell^{(\alpha)}\}$ is orthonormal, if (a) the bases $\{\mathbf{b}_i^{(\alpha_1)}\}$ and $\{\mathbf{b}_j^{(\alpha_2)}\}$ of the sons α_1, α_2 are orthonormal and (b) the matrices $C^{(\alpha,\ell)}$ in (19) are orthonormal with respect to the Frobenius scalar product: $\langle C^{(\alpha,\ell)}, C^{(\alpha,m)} \rangle_F = \sum_{ij} c_{ij}^{(\alpha,\ell)} \overline{c_{ij}^{(\alpha,m)}} = \delta_{\ell m}$.

The bases can be orthonormalised as follows. Orthonormalise the explicitly given bases at the leaves (e.g., by QR). As soon as $\{\mathbf{b}_i^{(\alpha_1)}\}$ and $\{\mathbf{b}_j^{(\alpha_2)}\}$ are orthonormal, orthonormalise the matrices $C^{(\alpha,\ell)}$. The new matrices $C_{\text{new}}^{(\alpha,\ell)}$ define a new orthonormal basis $\{\mathbf{b}_{\ell,\text{new}}^{(\alpha)}\}$ via (21).

The above mentioned calculations require basis transformations. Here the following has to be taken into account (cf. Sect. 4.3 and [5, §11.3.1.4]).

- Case A1. Let α_1 be the first son of α . Assume that the basis $\{\mathbf{b}_i^{(\alpha_1)}\}$ is transformed into a new basis $\{\mathbf{b}_{i,\text{new}}^{(\alpha_1)}\}$ so that $\mathbf{b}_i^{(\alpha_1)} = \sum_k T_{ki} \mathbf{b}_{k,\text{new}}^{(\alpha_1)}$. Changing $C^{(\alpha,\ell)}$ into $C_{\text{new}}^{(\alpha,\ell)} := T C^{(\alpha,\ell)}$, the basis $\{\mathbf{b}_\ell^{(\alpha)}\}$ remains unchanged.
- Case A2. If $\mathbf{b}_i^{(\alpha_2)} = \sum_k T_{ki} \mathbf{b}_{k,\text{new}}^{(\alpha_2)}$ is a transformation of the second son of α , $C^{(\alpha,\ell)}$ must be changed into $C^{(\alpha,\ell)} T^\top$.
- Case B. Consider a non-leaf vertex α . If the basis $\{\mathbf{b}_\ell^{(\alpha)}\}$ should be transformed into $\mathbf{b}_{\ell,\text{new}}^{(\alpha)} := \sum_i T_{\ell i} \mathbf{b}_i^{(\alpha)}$, one has to change the coefficient matrices $C^{(\alpha,\ell)}$ by $C_{\text{new}}^{(\alpha,\ell)} := \sum_i T_{\ell i} C^{(\alpha,i)}$. (In addition, this transformation causes changes at the father vertex according to Case A1 or Case A2).

As in Sect. 3.3, the bases have to be orthonormalised before the HOSVD bases are computed.

4.5 HOSVD bases

The challenge is the computation of the HOSVD, more precisely of the singular values $\sigma_i^{(\alpha)}$ and the left singular vectors (tensors) $\mathbf{b}_i^{(\alpha)}$ of $\mathcal{M}_\alpha(\mathbf{v})$. We recall that these data require the diagonalisation of the square matrix⁷ $\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H$. In the case of the tensor subspace representation of Sect. 3.3 it was possible to reduce $\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H$ to $\mathcal{M}_\alpha(\mathbf{c})\mathcal{M}_\alpha(\mathbf{c})^H$ involving the (smaller) core tensor. Now we reduce the computation of $\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H$ to matrix operations only involving the data \mathbf{C}_α .

The basis $\mathbf{B}_\alpha = \{\mathbf{b}_i^{(\alpha)} : 1 \leq i \leq r_\alpha\}$ spans the subspace $\mathbf{U}_\alpha \subset \mathbf{V}_\alpha = \bigotimes_{j \in \alpha} V_j$. The requirement $\mathbf{v} \in \mathbf{U}_D$ implies that $\mathbf{U}_\alpha^{\min}(\mathbf{v}) \subset \mathbf{U}_\alpha$ (cf. Remark 4). Together with $\mathbf{U}_\alpha^{\min}(\mathbf{v}) = \text{range}(\mathcal{M}_\alpha(\mathbf{v}))$ (cf. (7)) we conclude that $\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H$ must be of the form

$$X_\alpha := \mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} \mathbf{b}_i^{(\alpha)} (\mathbf{b}_j^{(\alpha)})^H = \mathbf{B}_\alpha E_\alpha \mathbf{B}_\alpha^H \tag{23}$$

with some coefficients $e_{ij}^{(\alpha)}$ which form an $r_\alpha \times r_\alpha$ matrix

$$E_\alpha = (e_{ij}^{(\alpha)})_{i,j=1}^{r_\alpha}.$$

To simplify matters we assume that the bases are already orthonormal⁸ (cf. Sect. 4.4). We start with the root $\alpha = D$ of the tree T_D . Since $r_D = 1$, $E_D = e_{11}^{(D)}$ is a scalar. The definition of $\mathcal{M}_D(\mathbf{v})$ in Sect. 2.2 shows that $X_\alpha = \mathbf{v}\mathbf{v}^H$. On the other hand, the equality $\mathbf{v} = c_1^{(D)} \mathbf{b}_1^{(D)}$ in (20) implies

$$E_D = e_{11}^{(D)} = |c_1^{(D)}|^2 \quad \text{and} \quad \sigma_1^{(D)} = |c_1^{(D)}|, \tag{24}$$

⁷ Here we interpret $\mathcal{M}_\alpha(\mathbf{v})$ as a matrix, not as a tensor.

⁸ The general case is treated in [5, Theorem 5.14]. There the proof is based on the tensor interpretation of the quantities, whereas here we use the isomorphic matrix interpretation.

where $\sigma_1^{(D)}$ is the only singular value of $\mathcal{M}_D(\mathbf{v})$. Its left singular vector is \mathbf{v} . The following recursion starts with $\alpha = D$.

We assume that for some non-leaf vertex $\alpha \in T_D$ the singular values $\sigma_i^{(\alpha)}$ and the matrix E_α are known. Now we want to determine E_{α_1} and E_{α_2} for the sons α_1 and α_2 of α . Concerning X_α and X_{α_1} , we recall the definition of $\mathcal{M}_\alpha(\mathbf{v})$ by (4). The entries of X_α are

$$\begin{aligned} X_\alpha[\mathbf{i}_\alpha, \mathbf{j}_\alpha] &= \sum_{\mathbf{k}_{\alpha^c} \in \mathbf{I}_{\alpha^c}} \mathcal{M}_\alpha(\mathbf{v})[\mathbf{i}_\alpha, \mathbf{k}_{\alpha^c}] \overline{\mathcal{M}_\alpha(\mathbf{v})[\mathbf{j}_\alpha, \mathbf{k}_{\alpha^c}]} \\ &= \sum_{\mathbf{k}_{\alpha^c} \in \mathbf{I}_{\alpha^c}} \mathbf{v}[(\mathbf{i}_\alpha, \mathbf{k}_{\alpha^c})] \overline{\mathbf{v}[(\mathbf{j}_\alpha, \mathbf{k}_{\alpha^c})]}. \end{aligned} \tag{25}$$

On the left-hand side, e.g., $\mathbf{i}_\alpha \in \mathbf{I}_\alpha = \times_{j \in \alpha} I_j$ and \mathbf{k}_{α^c} form the pair of matrix indices, while in the second line $(\mathbf{i}_\alpha, \mathbf{k}_{\alpha^c}) \in \mathbf{I}_D$ is the index of \mathbf{v} . Analogously we have

$$X_{\alpha_1}[\mathbf{i}_{\alpha_1}, \mathbf{j}_{\alpha_1}] = \sum_{\mathbf{k}_{\alpha_1^c} \in \mathbf{I}_{\alpha_1^c}} \mathbf{v}[(\mathbf{i}_{\alpha_1}, \mathbf{k}_{\alpha_1^c})] \overline{\mathbf{v}[(\mathbf{j}_{\alpha_1}, \mathbf{k}_{\alpha_1^c})]}.$$

The complement of α_1 is $\alpha_1^c = \alpha^c \dot{\cup} \alpha_2$ so that $\mathbf{I}_{\alpha_1^c} = \mathbf{I}_{\alpha^c} \times \mathbf{I}_{\alpha_2}$. Hence the summation over $\mathbf{k}_{\alpha_1^c} \in \mathbf{I}_{\alpha_1^c}$ becomes a double sum over $\mathbf{k}_{\alpha^c} \in \mathbf{I}_{\alpha^c}$ and $\mathbf{k}_{\alpha_2} \in \mathbf{I}_{\alpha_2}$. The sum over $\mathbf{k}_{\alpha^c} \in \mathbf{I}_{\alpha^c}$ already appears in (25) so that

$$\begin{aligned} X_{\alpha_1}[\mathbf{i}_{\alpha_1}, \mathbf{j}_{\alpha_1}] &= \sum_{\mathbf{k}_{\alpha_2} \in \mathbf{I}_{\alpha_2}} \sum_{\mathbf{k}_{\alpha^c} \in \mathbf{I}_{\alpha^c}} \mathbf{v}[(\mathbf{i}_{\alpha_1}, \mathbf{k}_{\alpha_2}, \mathbf{k}_{\alpha^c})] \overline{\mathbf{v}[(\mathbf{j}_{\alpha_1}, \mathbf{k}_{\alpha_2}, \mathbf{k}_{\alpha^c})]} \\ &\stackrel{(25)}{=} \sum_{\mathbf{k}_{\alpha_2} \in \mathbf{I}_{\alpha_2}} X_\alpha[(\mathbf{i}_{\alpha_1}, \mathbf{k}_{\alpha_2}), (\mathbf{j}_{\alpha_1}, \mathbf{k}_{\alpha_2})]. \end{aligned}$$

Returning to the matrices $\mathcal{M}_\alpha(\mathbf{v})$ and $\mathcal{M}_{\alpha_1}(\mathbf{v})$, the latter sum can be regarded as a matrix multiplication when we interpret $\mathbf{b}_i^{(\alpha)}$ in (23) as the $r_{\alpha_1} \times r_{\alpha_2}$ matrix $\sum_{v,\mu} C_{v,\mu}^{(\alpha,i)} b_v^{(\alpha_1)} (b_\mu^{(\alpha_2)})^\top$:

$$\begin{aligned} X_{\alpha_1} &= \sum_{\mathbf{k}_{\alpha_2} \in \mathbf{I}_{\alpha_2}} X_\alpha[(\cdot, \mathbf{k}_{\alpha_2}), (\cdot, \mathbf{k}_{\alpha_2})] \\ &= \sum_{\mathbf{k}_{\alpha_2} \in \mathbf{I}_{\alpha_2}} \left(\sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} \mathbf{b}_i^{(\alpha)} (\mathbf{b}_j^{(\alpha)})^\mathbf{H} \right) [(\cdot, \mathbf{k}_{\alpha_2}), (\cdot, \mathbf{k}_{\alpha_2})] \\ &= \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} \mathbf{b}_i^{(\alpha)} \overline{\mathbf{b}_j^{(\alpha)}[\cdot, \mathbf{k}_{\alpha_2}]} \\ &\stackrel{(18)}{=} \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} \left(\sum_{v,\mu} c_{v,\mu}^{(\alpha,i)} \mathbf{b}_v^{(\alpha_1)} (\mathbf{b}_\mu^{(\alpha_2)})^\top \right) \left(\sum_{\lambda,\varkappa} c_{\lambda,\varkappa}^{(\alpha,j)} \mathbf{b}_\lambda^{(\alpha_1)} (\mathbf{b}_\varkappa^{(\alpha_2)})^\top \right)^\mathbf{H}. \end{aligned}$$

Since the basis is orthonormal, we obtain that $\mathbf{b}_v^{(\alpha_1)} (\mathbf{b}_\mu^{(\alpha_2)})^\top (\mathbf{b}_\lambda^{(\alpha_1)} (\mathbf{b}_\varkappa^{(\alpha_2)})^\top)^\mathbf{H} = \mathbf{b}_v^{(\alpha_1)} (\mathbf{b}_\mu^{(\alpha_2)})^\top \overline{\mathbf{b}_\varkappa^{(\alpha_2)} (\mathbf{b}_\lambda^{(\alpha_1)})^\mathbf{H}} = \langle \mathbf{b}_\mu^{(\alpha_2)}, \mathbf{b}_\varkappa^{(\alpha_2)} \rangle \mathbf{b}_v^{(\alpha_1)} (\mathbf{b}_\lambda^{(\alpha_1)})^\mathbf{H} = \delta_{\mu\varkappa} \mathbf{b}_v^{(\alpha_1)} (\mathbf{b}_\lambda^{(\alpha_1)})^\mathbf{H}$ ($\delta_{\mu\varkappa}$: Kronecker delta). Hence

$$X_{\alpha_1} = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} \sum_{v,\mu,\lambda} c_{v\mu}^{(\alpha,i)} \overline{c_{\lambda\mu}^{(\alpha,j)}} \mathbf{b}_v^{(\alpha_1)} (\mathbf{b}_\lambda^{(\alpha_1)})^H.$$

This proves that (23) holds for α_1 instead of α with coefficients $e_{v\lambda}^{(\alpha_1)} = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} \sum_{\lambda} c_{v\mu}^{(\alpha,i)} \overline{c_{\lambda\mu}^{(\alpha,j)}}$ forming the matrix $E_{\alpha_1} = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} C^{(\alpha,i)} (C^{(\alpha,j)})^H$ (cf. (19)). A similar treatment of X_{α_2} proves the following theorem.

Theorem 1 *The matrices $E_\alpha, E_{\alpha_1}, E_{\alpha_2}$ are connected by*

$$E_{\alpha_1} = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} C^{(\alpha,i)} (C^{(\alpha,j)})^H, \quad E_{\alpha_2} = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)} (C^{(\alpha,i)})^T \overline{C^{(\alpha,j)}}.$$

As in Sect. 3.3 the HOSVD bases $\{\mathbf{b}_{i,\text{HOSVD}}^{(\alpha)}\}$ is defined by the diagonalisation of $X_\alpha := \mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^H = \sum_i (\sigma_i^{(\alpha)})^2 \mathbf{b}_{i,\text{HOSVD}}^{(\alpha)} (\mathbf{b}_{i,\text{HOSVD}}^{(\alpha)})^H$. A comparison with (23) shows that X_α is diagonalised if and only if

$$E_\alpha = \text{diag}\{(\sigma_1^{(\alpha)})^2, (\sigma_2^{(\alpha)})^2, \dots\}. \tag{26}$$

Since $r_D = 1$ at the root $\alpha = D$, we have $\sigma_1^{(D)} = |c_1^{(D)}|$ (cf. (24)) and $\mathbf{b}_1^{(D)} = \mathbf{b}_{1,\text{HOSVD}}^{(D)}$. Assume that the HOSVD basis $\{\mathbf{b}_{i,\text{HOSVD}}^{(\alpha)}\}$ is already chosen for the representation (we recall that the definition of $\mathbf{b}_{i,\text{HOSVD}}^{(\alpha)}$ is implicitly given by the coefficient matrices $C^{(\alpha,i)}$). Combining (26) with Theorem 1 we obtain

$$E_{\alpha_1} = \sum_{i=1}^{r_\alpha} (\sigma_i^{(\alpha)})^2 C^{(\alpha,i)} (C^{(\alpha,i)})^H, \quad E_{\alpha_2} = \sum_{i=1}^{r_\alpha} (\sigma_i^{(\alpha)})^2 (C^{(\alpha,i)})^T \overline{C^{(\alpha,i)}}.$$

Diagonalisation of the explicitly given matrices E_{α_1} and E_{α_2} yields

$$E_{\alpha_1} = U \Sigma_{\alpha_1}^2 U^H, \quad E_{\alpha_2} = V \Sigma_{\alpha_2}^2 V^H$$

with orthogonal matrices U, V and diagonal matrices $\Sigma_{\alpha_i} = \text{diag}\{\sigma_1^{(\alpha_i)}, \dots\}$. Since also \mathbf{B}_{α_i} is orthogonal (i.e., $\mathbf{B}_{\alpha_i}^H \mathbf{B}_{\alpha_i} = I$), the diagonalisation is given by $X_{\alpha_1} = (\mathbf{B}_{\alpha_1} U) \Sigma_{\alpha_1}^2 (\mathbf{B}_{\alpha_1} U)^H$, $X_{\alpha_2} = (\mathbf{B}_{\alpha_2} V) \Sigma_{\alpha_2}^2 (\mathbf{B}_{\alpha_2} V)^H$. Hence $\mathbf{B}_{\alpha_1}^{\text{HOSVD}} = \mathbf{B}_{\alpha_1} U$ and $\mathbf{B}_{\alpha_2}^{\text{HOSVD}} = \mathbf{B}_{\alpha_2} V$ are the desired HOSVD bases at the vertices α_1 and α_2 . If α_i is a leaf, this transformation is performed explicitly. Otherwise the coefficient matrices are modified according to Sect. 4.3. The procedure is repeated for the sons of α_1, α_2 until we reach the leaves. Then at all vertices HOSVD bases are introduced together with singular values $\sigma_v^{(\alpha)}$.

If there are vanishing singular values $\sigma_i^{(\alpha)}$, the corresponding contributions can be omitted. This reduces the associated subspace \mathbf{U}_α (cf. (16)) to the minimal subspace $\mathbf{U}_\alpha^{\min}(\mathbf{v})$. Correspondingly the value of r_α becomes $\text{rank}_\alpha(\mathbf{v})$.

4.6 HOSVD truncation

We assume that following the procedure described above the hierarchical representation uses the HOSVD bases. The format \mathcal{H}_τ with $\tau = (r_\alpha)_{\alpha \in T_D}$ consists of all tensors $\mathbf{v} \in \mathbf{V}$ with $\text{rank}_\alpha(\mathbf{v}) \leq r_\alpha$. Given $\mathbf{v} \in \mathcal{H}_\tau$ we ask for an approximation $\mathbf{u} \in \mathcal{H}_\mathfrak{s}$ for a smaller tuple \mathfrak{s} with $\mathfrak{s} \leq \tau$.

The HOSVD truncation is similar to the procedure in Sect. 3.4. In terms of the (implicitly defined) bases the approximation $\mathbf{u}_{\text{HOSVD}}$ is obtained by omitting all contributions involving the HOSVD basis vectors $\mathbf{b}_i^{(\alpha)}$ for $s_\alpha < i \leq r_\alpha$. In practice this means that the coefficient matrices $C^{(\alpha,i)}$ are omitted for $s_\alpha < i \leq r_\alpha$, while the remaining $r_{\alpha_1} \times r_{\alpha_2}$ matrices $C^{(\alpha,i)}$ are reduced to size $s_{\alpha_1} \times s_{\alpha_2}$ by deleting the last $r_{\alpha_1} - s_{\alpha_1}$ rows and $r_{\alpha_2} - s_{\alpha_2}$ columns. If $\alpha = \{j\}$ is a leaf, the explicitly given basis $\{b_{i,\text{HOSVD}}^{(j)} : 1 \leq i \leq r_j\}$ is replaced by $\{b_{i,\text{HOSVD}}^{(j)} : 1 \leq i \leq s_j\}$.

The approximation error $\mathbf{v} - \mathbf{u}_{\text{HOSVD}}$ satisfies (cf. [5, Theorem 11.58]), where $\mathbf{v}_{\text{HOSVD}}$ is the truncated value of \mathbf{v} :

$$\|\mathbf{v} - \mathbf{v}_{\text{HOSVD}}\| \leq \sqrt{\sum_{\alpha} \sum_{v \geq s_{\alpha}+1} (\sigma_v^{(\alpha)})^2} \leq \sqrt{2d - 3} \|\mathbf{v} - \mathbf{v}_{\text{best}}\|. \tag{27}$$

The first inequality allows us to explicitly control the error with respect to the Euclidean norm by the choice of the omitted singular values. The second inequality proves quasi-optimality of this truncation. $\mathbf{u}_{\text{best}} \in \mathcal{H}_s$ is the best approximation. The parameter d is the order of the tensor.

The number $2d - 3$ on the right-hand side becomes smaller if $s_\alpha = r_\alpha$ holds for some vertices α . For instance, the TT format as described in [11] uses the maximal value $s_j = r_j = \dim(V_j)$ for the leaves. Then (27) holds with $\sqrt{d - 1}$ instead of $\sqrt{2d - 3}$.

5 Properties of the SVD projection

5.1 Case of the tensor-subspace format

The HOSVD truncation of the tensor-subspace format in Sect. 3.4 is the Kronecker product

$$\Pi := \bigotimes_{j=1}^d P_j,$$

where $P_j : V_j \rightarrow \text{span}\{b_i^{(j)} : 1 \leq i \leq s_j\}$ with $s_j < r_j$ is the orthogonal projection. Π is again an orthogonal projection onto $\bigotimes_{j=1}^d \text{span}\{b_i^{(j)} : 1 \leq i \leq s_j\}$.

The tensor product Π of the single projections P_j can also be written as a usual product $\Pi = \prod_{j=1}^d \mathbf{P}_j$ of

$$\mathbf{P}_j := \mathbf{I}_{[j]} \otimes P_j \quad \text{with } \mathbf{I}_{[j]} := \bigotimes_{k \neq j} I_k, \tag{28}$$

where I_j is the identity map on V_j . Since the projections \mathbf{P}_j commute, the order of the factors in $\prod_{j=1}^d \mathbf{P}_j$ does not matter. We recall the singular-value decomposition of the matricisation $\mathcal{M}_j(\mathbf{v})$ (cf. Sect. 2.4):

$$\mathcal{M}_j(\mathbf{v}) = \sum_{i=1}^{r_j} \sigma_i^{(j)} b_i^{(j)} (\mathbf{b}_i^{[j]})^\top,$$

where the superscript $[j] = \{j\}^c$ denotes the complement of the leaf $\alpha = \{j\}$. Using (5b), we get

$$\mathcal{M}_j(\mathbf{P}_j \mathbf{v}) = P_j \mathcal{M}_j(\mathbf{v}) = \sum_{i=1}^{r_j} \sigma_i^{(j)} \left(P_j b_i^{(j)} \right) (\mathbf{b}_i^{[j]})^\top = \sum_{i=1}^{s_j} \sigma_i^{(j)} b_i^{(j)} (\mathbf{b}_i^{[j]})^\top.$$

However, we may also define

$$\hat{\mathbf{P}}_j := \mathbf{P}_{[j]} \otimes I_j, \tag{29}$$

where $\mathbf{P}_{[j]}$ is the orthogonal projection of $\mathbf{V}_{[j]} = \bigotimes_{k \neq j} V_k$ onto $\text{span}\{\mathbf{b}_i^{[j]} : 1 \leq i \leq s_j\}$. As (5b) again shows

$$\mathcal{M}_j(\hat{\mathbf{P}}_j \mathbf{v}) = \mathcal{M}_j(\mathbf{v}) \mathbf{P}_{[j]}^\top = \sum_{i=1}^{r_j} \sigma_i^{(j)} b_i^{(j)} (\mathbf{P}_{[j]} \mathbf{b}_i^{[j]})^\top = \sum_{i=1}^{s_j} \sigma_i^{(j)} b_i^{(j)} (\mathbf{b}_i^{[j]})^\top,$$

we obtain the identical value $\mathbf{P}_j \mathbf{v} = \hat{\mathbf{P}}_j \mathbf{v}$ although the projections are different. This property has interesting consequences. We introduce

$$\Pi_j := \left(\prod_{k \neq j} \mathbf{P}_k \right) \hat{\mathbf{P}}_j$$

and observe that $\Pi_j \mathbf{v} = \left(\prod_{k \neq j} \mathbf{P}_k \right) \hat{\mathbf{P}}_j \mathbf{v} = \left(\prod_{k \neq j} \mathbf{P}_k \right) \mathbf{P}_j \mathbf{v} = \Pi \mathbf{v}$ holds for the special tensor \mathbf{v} although $\Pi_j \neq \Pi$. Note that all maps \mathbf{P}_k and $\hat{\mathbf{P}}_j$ are elementary tensors containing the identity $I_j : V_j \rightarrow V_j$ with respect to the j -th direction. This proves the next lemma for which we introduce the following notation. Let $\varphi_j : V_j \rightarrow W_j$ be a linear map. It gives rise to the elementary Kronecker product $\phi_j := I_1 \otimes \dots \otimes I_{j-1} \otimes \varphi_j \otimes I_{j+1} \otimes \dots \otimes I_d$.

Lemma 1 *Let $\varphi_j : V_j \rightarrow W_j$ and ϕ_j as above. Then $\phi_j \Pi_j = \Pi_j \phi_j$ holds (the latter Π_j contains the identity $I_j : W_j \rightarrow W_j$ instead of $I_j : V_j \rightarrow V_j$).*

This allows the following estimate with respect to the Euclidean norm.

Conclusion 2 *Given $\mathbf{v} \in \mathbf{V}$, let $\mathbf{u}_{\text{HOSVD}} \in \mathcal{T}_s$ be the HOSVD approximation defined in Sect. 3.4. With ϕ_j from above we have*

$$\|\phi_j \mathbf{u}_{\text{HOSVD}}\| \leq \|\phi_j \mathbf{v}\|. \tag{30}$$

Proof $\mathbf{u}_{\text{HOSVD}} = \Pi_j \mathbf{v}$ shows that $\phi_j \mathbf{u}_{\text{HOSVD}} = \phi_j \Pi_j \mathbf{v} = \Pi_j \phi_j \mathbf{v}$. Since Π_j is a product of orthogonal projection, $\|\Pi_j \phi_j \mathbf{v}\| \leq \|\phi_j \mathbf{v}\|$ follows. \square

In the case of infinite-dimensional Hilbert spaces \mathbf{V} we may consider unbounded linear maps ϕ_j . The subspace of elements \mathbf{v} for which $\phi_j \mathbf{v}$ is defined, is called the *domain* of ϕ_j .

Conclusion 3 *If $\mathbf{v} \in \mathbf{V}$ belongs to the domain of ϕ_j , then also $\mathbf{u}_{\text{HOSVD}}$ belongs to the domain and satisfies (30).*

An important example is the topological tensor space $\mathbf{V} = L^2(\Omega) = \bigotimes_{k=1}^d L^2(\Omega_j)$, where Ω is the Cartesian product of the Ω_j . Set $\phi_j = \partial^k / \partial x_j^k$. If the function $\mathbf{v} \in \mathbf{V}$ possesses a k -th derivative with respect to x_j , then by Conclusion 3 also $\mathbf{u}_{\text{HOSVD}}$ is k -times differentiable in the L^2 sense and satisfies $\|\partial^k \mathbf{u}_{\text{HOSVD}} / \partial x_j^k\|_{L^2} \leq \|\partial^k \mathbf{v} / \partial x_j^k\|_{L^2}$. Assuming sufficient smoothness of \mathbf{v} and using the Gagliardo–Nirenberg inequality, we proved in [6] estimates of $\|\mathbf{v} - \mathbf{u}_{\text{HOSVD}}\|_\infty$ with respect to the maximum norm by means of the L^2 norm of $\mathbf{v} - \mathbf{u}_{\text{HOSVD}}$. This is important for the pointwise evaluation of the truncated function.

Another trivial conclusion from (30) is that $\phi_j \mathbf{v} = 0$ implies $\phi_j \mathbf{u}_{\text{HOSVD}} = 0$. For instance, let $\varphi_j \in V'_j$ be a functional on V_j (i.e., $W_j = \mathbb{K}$). Examples of φ_j are the mean value $\varphi_j(u) = \mathbf{1}^\top u$ or a zero at a certain index i^* : $\varphi_j(u) = u_{i^*} = 0$. We say that \mathbf{v} satisfies the side condition φ_j if $\phi_j \mathbf{v} = 0$. We conclude that $\mathbf{u}_{\text{HOSVD}}$ satisfies the same side condition. In the case of $\varphi_j(u) = \mathbf{1}^\top u$, also $\mathbf{u}_{\text{HOSVD}}$ has a vanishing mean. If $\varphi_j(u) = u_{i^*}$, $\mathbf{u}_{\text{HOSVD}}[\mathbf{i}] = 0$ holds for \mathbf{i} with $i_j = i^*$.

In the case of matrix spaces V_j , structural properties like symmetry or sparsity can be described by functionals. One concludes that the HOSVD approximations lead to matrices of the same structure.

5.2 Best approximation \mathbf{u}_{best}

We recall that the HOSVD approximation $\mathbf{u}_{\text{HOSVD}} \in \mathcal{T}_r$ of $\mathbf{v} \in \mathbf{V}$ is not (necessarily) the best approximation defined by

$$\mathbf{u}_{\text{best}} \in \mathcal{T}_r \quad \text{with} \quad \|\mathbf{v} - \mathbf{u}_{\text{best}}\| = \inf\{\|\mathbf{v} - \mathbf{u}\| : \mathbf{u} \in \mathcal{T}_r\} \tag{31}$$

(cf. Definition 1). Nevertheless, \mathbf{u}_{best} has similar properties as $\mathbf{u}_{\text{HOSVD}}$.

Define

$$U_k := U_k^{\min}(\mathbf{u}_{\text{best}}) \quad \text{for } 1 \leq k \leq d.$$

Let $P_k : V_k \rightarrow U_k$ be the orthogonal projection onto U_k . Based on these projections we define \mathbf{P}_k and Π as in Sect. 5.1. Now we fix one index j and define $\Pi_{[j]} := \prod_{k \neq j} \mathbf{P}_k$. Set $\mathbf{v}_j := \Pi_{[j]} \mathbf{v} \in U_1 \otimes \dots \otimes U_{j-1} \otimes V_j \otimes U_{j+1} \otimes \dots \otimes U_d$ and note that $\mathbf{P}_j \mathbf{v}_j = \mathbf{u}_{\text{best}}$. Based on the SVD of $\mathcal{M}_j(\mathbf{v}_j)$ we can determine its HOSVD approximation $\mathbf{u}_{\text{HOSVD}} \in \mathcal{T}_r$. Since it is the minimiser of $\min_{\mathbf{u} \in \mathcal{T}_r} \|\mathbf{v}_j - \mathbf{u}\|$, we have $\|\mathbf{v}_j - \mathbf{u}_{\text{HOSVD}}\| \leq \|\mathbf{v}_j - \mathbf{u}_{\text{best}}\|$. For an indirect proof assume that $\|\mathbf{v}_j - \mathbf{u}_{\text{HOSVD}}\| < \|\mathbf{v}_j - \mathbf{u}_{\text{best}}\|$. Both $\mathbf{u}_{\text{HOSVD}}$ and \mathbf{u}_{best} are in the range of $\Pi_{[j]}$, i.e.,

$$(I - \Pi_{[j]}) \mathbf{u}_{\text{HOSVD}} = (I - \Pi_{[j]}) \mathbf{u}_{\text{best}} = 0.$$

Pythagoras' equality yields

$$\begin{aligned} \|\mathbf{v} - \mathbf{u}_{\text{HOSVD}}\|^2 &= \|\Pi_{[j]}(\mathbf{v} - \mathbf{u}_{\text{HOSVD}})\|^2 + \|(I - \Pi_{[j]})(\mathbf{v} - \mathbf{u}_{\text{HOSVD}})\|^2 \\ &= \|\mathbf{v}_j - \mathbf{u}_{\text{HOSVD}}\|^2 + \|(I - \Pi_{[j]}) \mathbf{v}\|^2 \\ &< \|\mathbf{v}_j - \mathbf{u}_{\text{best}}\|^2 + \|(I - \Pi_{[j]}) \mathbf{v}\|^2 \\ &= \|\Pi_{[j]}(\mathbf{v} - \mathbf{u}_{\text{best}})\|^2 + \|(I - \Pi_{[j]})(\mathbf{v} - \mathbf{u}_{\text{best}})\|^2 \\ &= \|\mathbf{v} - \mathbf{u}_{\text{best}}\|^2 \end{aligned}$$

in contradiction to the optimality of \mathbf{u}_{best} . Hence, $\|\mathbf{v} - \mathbf{u}_{\text{HOSVD}}\| = \|\mathbf{v} - \mathbf{u}_{\text{best}}\|$ must hold. Depending on the multiplicity of certain singular values, the SVD approximation may be unique. In this case $\mathbf{u}_{\text{HOSVD}} = \mathbf{u}_{\text{best}}$ holds. If the SVD approximation is not unique, we may choose \mathbf{u}_{best} as $\mathbf{u}_{\text{HOSVD}} = \mathbf{P}_j \mathbf{v}_j$. Knowing that \mathbf{P}_j is a SVD projection, we may replace \mathbf{P}_j by $\hat{\mathbf{P}}_j$ as defined in (29). The projection $\Pi_j := \Pi_{[j]} \hat{\mathbf{P}}_j$ has the same properties as Π_j in §5.1. This proves the following (cf. Uschmajew [13]).

Theorem 4 *The statements of Lemma 1 and the Conclusions 2 and 3 also hold for the best approximation \mathbf{u}_{best} in (31) and the related mapping Π_j .*

5.3 Case of the hierarchical format

The HOSVD truncation within the hierarchical format (cf. Sect. 4.6) can be expressed by orthogonal projections \mathbf{P}_α for all vertices α of the tree T_D . However, different from Sect. 5.1, projections \mathbf{P}_α and \mathbf{P}_β commute if and only if $\alpha \cap \beta = \emptyset$. The truncation is described by the product

$$\Pi := \prod_{\alpha \in T_D} \mathbf{P}_\alpha,$$

where the factors are ordered in such a way that \mathbf{P}_α is applied before \mathbf{P}_{α_1} and \mathbf{P}_{α_2} (α_1, α_2 sons of α) follow. Because of these restrictions, the analysis is more involved. We refer the reader to Hackbusch [6, § 4]. As a result the statements in Sect. 5.1 also hold for the hierarchical format.

Acknowledgements Open access funding provided by Max Planck Society. The author likes to thank Antonio Falcó for fruitful discussions during a visit at the Universidad Cardenal Herrera-CEU.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. De Lathauwer, L., De Moor, B.L.R., Vandewalle, J.: A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21**, 1253–1278 (2000)
2. Falcó, A., Hackbusch, W.: On minimal subspaces in tensor representations. *Found. Comput. Math.* **12**, 765–803 (2012)
3. Falcó, A., Hackbusch, W., Nouy, A.: Tree-based tensor formats (**To appear in SEMA**)
4. Greub, W.H.: *Multilinear Algebra*, 2nd edn. Springer, New York (1978)
5. Hackbusch, W.: *Tensor Spaces and Numerical Tensor Calculus, SSCM*, vol. 42. Springer, Berlin (2012)
6. Hackbusch, W.: L^∞ estimation of tensor truncations. *Numer. Math.* **125**, 419–440 (2013)
7. Hackbusch, W., Kühn, S.: A new scheme for the tensor representation. *J. Fourier Anal. Appl.* **15**, 706–722 (2009)
8. Hackbusch, W., Uschmajew, A.: On the interconnection between the higher-order singular values of real tensors. *Numer. Math.* **135**, 875–894 (2017)
9. Håstad, J.: Tensor rank is NP-complete. *J. Algorithms* **11**, 644–654 (1990)
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. Oseledets, I.V.: Tensor-train decomposition. *SIAM J. Sci. Comput.* **33**, 2295–2317 (2011)
12. Oseledets, I.V., Tyrtshnikov, E.E.: TT-cross approximation for multidimensional arrays. *Linear Algebra Appl.* **432**, 70–88 (2010)
13. Uschmajew, A.: Regularity of tensor product approximations to square integrable functions. *Constr. Approx.* **34**, 371–391 (2011)
14. Verstraete, F., Cirac, J.I.: Matrix product states represent ground states faithfully. *Phys. Rev. B* **73**, 094,423 (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.