*Acta Numerica* (2014), pp. 651–742
doi:10.1017/S0962492914000087


# Numerical tensor calculus*

Wolfgang Hackbusch
*Max-Planck-Institut für Mathematik in den Naturwissenschaften,*
*Inselstr. 22, D-04103 Leipzig, Germany*
*E-mail:* wh@mis.mpg.de

The usual large-scale discretizations are applied to two or three spatial dimensions. The standard methods fail for higher dimensions because the data size increases exponentially with the dimension. In the case of a regular grid with $n$ grid points per direction, a spatial dimension $d$ yields $n^d$ grid points. A grid function defined on such a grid is an example of a tensor of order $d$. Here, suitable tensor formats help, since they try to approximate these huge objects by a much smaller number of parameters, which increases only linearly in $d$. In this way, data of size $n^d = 1000^{1000}$ can also be treated.

This paper introduces the algebraic and analytical aspects of tensor spaces. The main part concerns the numerical representation of tensors and the numerical performance of tensor operations.

## CONTENTS



* Colour online for monochrome figures available at journals.cambridge.org/anu.

## 1. Introduction

Even now, the numerical treatment of fully populated $n \times n$ matrices for large $n$ can lead to storage problems. In this case, various methods are known to reduce the storage to $O(n \log^* n)$ (see, *e.g.*, Hackbusch 2009), provided that the matrix satisfies certain conditions that enable data compression. This is only one example for which polynomial complexity can be improved to linear complexity $O(n)$ or almost linear complexity $O(n \log^* n)$. In this paper we shall consider problems where even linear complexity $O(n)$ is far too large. Under optimal conditions, it will be possible to replace $O(n)$ by logarithmic complexity $O(\log n)$.

Except for very particular examples, it is impossible to compress the data without compression error. The accuracy $\varepsilon > 0$ will influence the constant hidden in $O(\log n)$. Again under optimal conditions, the complexity as a function of $n$ and $\varepsilon$ will behave like $O(\log \frac{1}{\varepsilon} \log n)$.

While there is a tradition of tensor *decompositions* we aim for efficient *representations* (see Section 1.4 for the distinction). The aspect of numerical algorithms for tensors is also emphasized by Beylkin and Mohlenkamp (2005).

### *1.1. Examples*

In the following, we describe some examples where tensors appear in numerical analysis. Furthermore, we give some *ad hoc* definitions. The precise definitions will follow later.

#### *1.1.1. Grid functions as tensors*

A matrix $M = (M_{ij})_{1 \le i,j \le n}$ is not necessarily derived from a linear mapping. If we solve a partial differential equation – say the Poisson equation with zero boundary conditions – in the square $\Omega = (0,1)^2$ by a difference method with step size $h = 1/(n+1)$, the solution $u_h$ is a grid function with values at $(\nu h, \mu h)$ for $1 \le \nu, \mu \le n$. Hence, the 'vector' $u_h$ has the shape of a matrix, $(u_{\nu\mu})_{1 \le \nu,\mu \le n}$. Replacing the domain $\Omega = (0,1)^2$ by the $d$-dimensional cube $\Omega = (0,1)^d$ and using an analogous discretization, we obtain a system of equations for a grid function $\mathbf{u}_h$ with entries

$$u_{i_1 i_2 \cdots i_d} \quad \text{for } 1 \le i_1, i_2, \ldots, i_d \le n. \tag{1.1}$$

Here, the data size is

$$N := n^d.$$

Obviously, linear complexity $O(N)$ for $n = 10^6$, $d = 3$ ($N = 10^{18}$), or $n = 10^4$, $d = 100$ ($N = 10^{400}$), or $n = 10$, $d = 10^6$ ($N = 10^{10^6}$), for example, far exceeds the available storage of any computer.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Considering the grid function $u_h \in \mathbb{R}^N$ as an element of the tensor space $\mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n$ of order $d$, we can make use of the tensor product of $\mathbb{R}^n$ vectors defined as follows:

$$\mathbf{v} := v^{(1)} \otimes \cdots \otimes v^{(d)} \text{ has the entries } \mathbf{v}[i_1 i_2 \cdots i_d] = v_{i_1}^{(1)} \cdot v_{i_2}^{(2)} \cdot \ldots \cdot v_{i_d}^{(d)}.$$

For instance, a monomial $\mathbf{x}^{\boldsymbol{\nu}}$ ($\mathbf{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{\nu} \in \mathbb{N}_0^d$ multi-index) restricted to the grid can be represented by $v^{(1)} \otimes \cdots \otimes v^{(d)}$, where $v^{(j)} \in \mathbb{R}^n$ has the entries $v_i^{(j)} = (ih)^{\nu_j}$. Obviously, the required data size is $dn$ (with $d$ as factor, not exponent!) Any product of the form $v^{(1)} \otimes \cdots \otimes v^{(d)}$ is called an *elementary tensor*. Algebraic tensors are sums of elementary tensors since the algebraic tensor space is defined as the span of all elementary tensors.

### *1.1.2. Kronecker matrices*

We continue with the example of the discrete Poisson equation in $\Omega = (0,1)^d$. The grid function discussed above is the solution of a linear system $\mathbf{A}_h \mathbf{u}_h = \mathbf{f}_h$. While $\mathbf{f}_h$ has a similar shape to $\mathbf{u}_h$, the $N \times N$ matrix $\mathbf{A}_h$ requires a closer look. The Laplace operator is the sum of the terms $\partial^2/\partial x_j^2$. Discretize the second derivative by the second divided difference. In the one-dimensional case, this leads to a tridiagonal $n \times n$ matrix $L$. The application of second differences to the $j$th direction can be expressed by the Kronecker product

$$\mathbf{L}_j := \underbrace{I \otimes \cdots \otimes I}_{j-1 \text{ factors}} \otimes L \otimes \underbrace{I \otimes \cdots \otimes I}_{d-j \text{ factors}},$$

where $I$ is the $n \times n$ unit matrix. The Kronecker product is defined by[1]

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{bmatrix}. \tag{1.2}$$

Note that the matrices $A$ and $B$ of the respective sizes $n_1 \times m_1$ and $n_2 \times m_2$ lead to an $n_1 n_2 \times m_1 m_2$ matrix.

Hence, $\mathbf{A}_h = \sum_{j=1}^d \mathbf{L}_j$ can be represented by means of a small number of much simpler building blocks. For the solution of $\mathbf{A}_h \mathbf{u}_h = \mathbf{f}_h$ it would be helpful to represent the inverse matrix $\mathbf{A}_h^{-1}$ too. Although there is no efficient *exact* representation of $\mathbf{A}_h^{-1}$, we shall describe a very accurate approximation in Section 5.10, which requires only a few Kronecker products of $n \times n$ matrices.

[1] This definition requires a suitable ordering of the indices. However, the ordering is only needed for the visualization in (1.2). The later, more mathematical definition in Remark 2.4 works for general index sets without any ordering.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*1.1.3. Eigenvectors*

Now we consider the eigenvalue problem $\mathbf{A}_h\mathbf{u}_h = \lambda\mathbf{u}_h$, where $\mathbf{A}_h = \sum_{j=1}^{d} \mathbf{L}_j$ contains terms of the form $\mathbf{L}_j := I \otimes \cdots \otimes I \otimes L^{(j)} \otimes I \otimes \cdots \otimes I$. Such matrices arise from the discretization of so-called *separable* differential operators in a Cartesian domain $\Omega$. In this case one easily proves that the complete set of eigenvectors $\mathbf{u}_h$ can be represented in the form

$$\mathbf{u}_h = v^{(1)} \otimes \cdots \otimes v^{(d)}, \quad \text{where } v^{(j)} \text{ is an eigenvector of } L^{(j)}.$$

Hence, $\mathbf{u}_h$ is an elementary tensor requiring only $dn$ data. This is one of the rare examples where a simple tensor representation applies without any compression error.

*1.1.4. Multivariate functions*

Considering the previous partial differential equation without discretization, we have to deal with multivariate functions $u(x_1, \ldots, x_d)$, $f(x_1, \ldots, x_d)$ and the differential operator. If the underlying domain is a Cartesian product, we can apply the tensor notation. Let $u^{(j)}$ be a univariate function whose variable is denoted by $x_j$. Then the elementary tensor $\mathbf{u} := u^{(1)} \otimes \cdots \otimes u^{(d)}$ is the multivariate function defined by

$$\mathbf{u}(x_1, \ldots, x_d) = u^{(1)}(x_1) \cdot u^{(2)}(x_2) \cdot \ldots \cdot u^{(d)}(x_d).$$

An example of an elementary tensor has already been mentioned. If $\mathbf{u}(\mathbf{x}) = \mathbf{x}^{\boldsymbol{\nu}}$ is a monomial in $d$ variables, we may write $\mathbf{u} = u^{(1)} \otimes \cdots \otimes u^{(d)}$ with $u^{(j)}(x_j) = (x_j)^{\nu_j}$. Correspondingly, a polynomial $\mathbf{u}$ becomes a sum of such elementary tensors. For an analytic function it can make sense to consider even infinite sums of elementary tensors.

The Kronecker product can easily be generalized to operators (*e.g.*, to differential operators). The precise definition will be given in Remark 2.4.

*1.1.5. Operations*

So far, the main subject of interest has been how to store certain tensors efficiently. This is only one aspect. Another important aspect is the efficient computation of *tensor operations*. We recall the example of the system $\mathbf{A}_h\mathbf{u}_h = \mathbf{f}_h$ from above. In principle, such a linear system can be solved iteratively by

$$\mathbf{u}_h^{m+1} = \mathbf{u}_h^m - \mathbf{B}_h(\mathbf{A}_h\mathbf{u}_h^m - \mathbf{f}_h),$$

where $\mathbf{B}_h$ is a suitable (preconditioning) matrix. As mentioned above, good approximations $\mathbf{B}_h$ of $\mathbf{A}_h^{-1}$ are available. To compute $\mathbf{u}_h^{m+1}$, we have to perform the matrix–vector multiplication $\mathbf{A}_h\mathbf{u}_h^m$, the subtraction $\mathbf{d}_h := \mathbf{A}_h\mathbf{u}_h^m - \mathbf{f}_h$, another matrix–vector multiplication $\delta_h := \mathbf{B}_h\mathbf{d}_h$, and again a subtraction $\mathbf{u}_h^m - \delta_h$. It will turn out that such an iteration is

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

not feasible without some kind of truncation (approximation), which therefore becomes an important operation. For other purposes we need further operations, such as scalar products, Hadamard products, or convolutions. The aim is that the computational cost (measured in elementary arithmetic operations) should be of a similar size to the storage size of the involved operands.

### *1.2. Notations*

*Numbers.* $\mathbb{Z}$ and $\mathbb{N}$ denote the integers and natural numbers, respectively, while $\mathbb{N}_0 := \{0, 1, \ldots\} = \mathbb{N} \cup \{0\}$.

*Field.* $\mathbb{K}$ denotes a general field. For numerical purposes, for which we need approximations, only $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ are of interest. Since tensor spaces over $\mathbb{R}$ may have properties different from those over $\mathbb{C}$, the flexible choice of $\mathbb{K}$ makes sense.

*Vector spaces.* $V_1$, $V_2, \ldots, V_d$ are the vector spaces generating the tensor space $\mathbf{V}$ (see below). If other $d$-tuples of vector spaces are needed, we use $W_j$, $X_j$, or $Y_j$ $(1 \le j \le d)$. Tacitly, we assume that all spaces involved are defined over the *same* field $\mathbb{K}$. There are no restrictions of the vector space concerning its dimension (any finite or infinite cardinality is admitted).

Vectors from $V_j$ are denoted by $v^{(j)}$. A family of such vectors is indexed by a further subscript: $\{v_\mu^{(j)} : 1 \le \mu \le r\}$.

The model space $\mathbb{K}^n$ is usually denoted in a different way. Let $I$ be an index set of cardinality $\#I = n$. Then $\mathbb{K}^I$ contains the vectors $(v_i)_{i \in I}$, $v_i \in \mathbb{K}$. Here, $I$ may be a set without ordering. This fact is helpful if $I$ is a Cartesian product $I_1 \times I_2$, *etc.* Even when $\#I_1 = \#I_2 = \cdots = \#I_d$, the notation $V_j = \mathbb{K}^{I_j}$ indicates the connection to $j$ (the position of $V_j$ is the tensor product).

*Tensor spaces.* The standard notation is $\mathbf{V} = \bigotimes_{j=1}^d V_j$. Even without mentioning this explicit definition, quantities $\mathbf{V}$ and $V_j$ appearing in a statement are tacitly related by $\mathbf{V} = \bigotimes_{j=1}^d V_j$. If further tensor spaces are needed, the analogous statement holds for

$$\mathbf{W} = \bigotimes_{j=1}^d W_j, \quad \mathbf{X} = \bigotimes_{j=1}^d X_j, \quad \mathbf{Y} = \bigotimes_{j=1}^d Y_j.$$

Bold capitals such as $\mathbf{V}$ are used for tensor spaces, while bold lower-case characters denote tensors.

In the model case of $V_j = \mathbb{K}^{I_j}$, the tensor space becomes $\mathbf{V} = \bigotimes_{j=1}^d \mathbb{K}^{I_j} = \mathbb{K}^{\mathbf{I}}$ with the index set

$$\mathbf{I} := I_1 \times \cdots \times I_d\,.$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

In this case, the entries of a tensor $\mathbf{v} \in \mathbf{V}$ are denoted by $\mathbf{v}_\mathbf{i} = \mathbf{v}[\mathbf{i}] = \mathbf{v}[i_1, \ldots, i_d]$ with $\mathbf{i} = (i_1, \ldots, i_d) \in \mathbf{I}$. Secondary subscripts, as in (1.1), are avoided for the sake of readability.

*Banach and Hilbert spaces.* For the notation of the related norms $\|\cdot\|$ and $\|\cdot\|_j$ and their corresponding dual norms, see Section 3. Hilbert spaces and the related scalar products are explained in Section 3.4. We let

$$L(X, Y) \quad \text{and} \quad \mathcal{L}(X, Y) \tag{1.3}$$

denote the space of mappings from $X$ into $Y$, which are linear and, respectively, linear and continuous. For $Y = \mathbb{K}$, we obtain the algebraic dual or, respectively, continuous dual space

$$X' \quad \text{and} \quad X^*.$$

*List of symbols.* The most important symbols not mentioned above are as follows:

| | |
|---|---|
| $\mathcal{H}_\mathfrak{t}$ | hierarchical format (equation (7.2e)) |
| HOSVD | higher-order singular value decomposition (Section 6.3) |
| $\mathcal{M}_j(\mathbf{v}), \mathcal{M}_\alpha(\mathbf{v})$ | matricization (Section 2.6) |
| $\operatorname{rank}(\mathbf{v})$ | tensor rank (Section 2.4) |
| $\underline{\operatorname{rank}}(\mathbf{v})$ | border rank (Definition 5.1) |
| $\operatorname{rank}_j, \operatorname{rank}_\alpha$ | $j$th ($\alpha$th) rank (Section 2.6) |
| $\mathcal{R}_r$ | $r$-term format (equation (2.4)) |
| $\mathcal{T}_\mathbf{r}$ | tensor subspace format (equation (6.2)) |
| $\mathbb{T}_\rho$ | matrix product or TT format (equation (9.5)) |
| $U_j^{\min}(\mathbf{v})$ | minimal subspace (Section 6.4) |
| $\mathbf{V}_\alpha$ | tensor space $\bigotimes_{j\in\alpha} V_j$ (Section 7.3) |
| $r, r_j, r_\alpha, \rho_j$ | ranks of formats $\mathcal{R}_r, \mathcal{T}_\mathbf{r}, \mathcal{H}_\mathfrak{t}, \mathbb{T}_\rho$ (equations (2.5), (6.2), (7.2c)) |
| $\#S$ | cardinality of a set $S$ |

In the past few decades, tensor applications have occurred mainly outside the fields of (applied) mathematics, so a terminology has been developed which is not the standard mathematical one (*e.g.*, tensors have become 'multi-way arrays'). Here, the author has deliberately tried to introduce a mathematical terminology suitable for (multi)linear algebra and analysis.

### *1.3. Matrix case*

Let $I_1$ and $I_2$ be finite index sets. The tensor space $\mathbb{K}^{I_1} \otimes \mathbb{K}^{I_2}$ of order two is equivalent to the set $\mathbb{K}^{I_1 \times I_2}$ of matrices. Using matrix notation, the tensor product of $v \in \mathbb{K}^{I_1}$ and $w \in \mathbb{K}^{I_2}$ is equal to

$$v \otimes w \cong v\, w^{\mathsf{T}}. \tag{1.4}$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Remark 1.1.** If $V_1$ and $V_2$ are two general but finite-dimensional vector spaces, the tensor space $V_1 \otimes V_2$ is isomorphic to the space of linear maps $L(V_2, V_1)$ from $V_2$ into $V_1$.

*Proof.* This follows, for example, from the isomorphism $V_j \cong \mathbb{K}^{n_j}$ with $n_j := \dim(V_j)$ and $L(V_2, V_1) \cong \mathbb{K}^{n_1 \times n_2}$. Another proof uses the fact that, for finite dimensions, $V_2$ and the dual space $V_2'$ are isomorphic (let $\phi : V_2 \to V_2'$ the isomorphism). Then $v^{(1)} \otimes v^{(2)}$ can be identified with the mapping

$$(v^{(1)} \otimes v^{(2)})(w^{(2)}) := (\phi(v^{(2)}))(w^{(2)}) \cdot v^{(1)}$$

(note that $\phi(v^{(2)}) \in V_2'$ and $(\phi(v^{(2)}))(w^{(2)}) \in \mathbb{K}$). □

Many of the pleasant properties of matrix spaces (*i.e.*, properties for $d = 2$) do not extend to $d \geq 3$, which characterizes the *true tensor case*. In order to recover features from matrix theory, we shall describe techniques mapping general tensors of order $d \geq 3$ to matrices (see Section 2.6).

### *1.4. Decomposition versus representation*

In this article we consider tensor *representations* and their use for tensor operations. Most of the papers on tensor applications in the past have discussed tensor *decompositions*. First, we try to explain the difference for the example of matrices.

If an $n \times n$ matrix $M$ is of rank $k \ll n$, it is advantageous to represent $M$ by the product $AB^{\mathsf{T}}$, where $A, B \in \mathbb{K}^{n \times k}$. Note that the choice of factor is not unique. Any regular $k \times k$ matrix $T$ generates factors $A' := A\,T^{-1}$ and $B' := B\,T^{\mathsf{T}}$ with the property $M = A'\,B'^{\mathsf{T}}$. For the purpose of representing an $n \times n$ matrix by two $n \times k$ matrices, all factorizations are equally welcome.

Another issue is the diagonalization of $M$ by $T^{-1}D\,T$. Here, the purpose is not to replace the data $M$ by $T$ and $D$. Instead, we are interested in the particular eigenvalues contained in $D$ and the eigenvectors collected in $T$. Here, $T^{-1}D\,T$ is a *decomposition* of $M$. Except for special cases and modulo permutations, this decomposition is unique. The singular value decomposition $M = UDV^{\mathsf{T}}$ ($U, V$ orthogonal matrices, $D$ diagonal) is of a similar nature.

Generally speaking, a *representation* of a tensor $\mathbf{v}$ is given by a mapping $\rho$ and a parameter tuple $p$ such that

$$\mathbf{v} = \rho(p).$$

The most important requirement is that the data size of $p$ is much smaller than $\dim(\mathbf{V})$. Injectivity of $\rho$ is of minor interest. It is more important that the mapping $\rho$ should be well conditioned. At least in the finite-dimensional case, surjectivity should hold in the sense that for all $\mathbf{v} \in \mathbf{V}$ there is a parameter tuple $p$ with $\mathbf{v} = \rho(p)$. The previous example of the

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

low-rank representation of a matrix corresponds to $M = \rho(A, B)$. If we allow parameters $A, B \in \mathbb{K}^{n\times k}$ for any $k$, the mapping is surjective.

The *decomposition* involves the *inverse* mapping $\rho^{-1}$. Given a tensor $\mathbf{v} \in \mathbf{V}$, we are interested in the parameter $p$ satisfying $\mathbf{v} = \rho(p)$. In the case of a singular value decomposition, the computation performs $\rho^{-1} : M \mapsto (U, D, V)$. This explanation shows that the decomposition is an inverse problem. It may be ill-posed (*e.g.*, the singular value decomposition in the case of approximately equal singular values).

In Sections 5–7 we shall discuss different types of tensor formats. We call them 'representations' since we are only interested in the direction $\rho : p \mapsto \mathbf{v}$. In many fields of application they are called 'decompositions'. In that case, we are interested in the parameter $p$, since it reveals certain properties of the object that has produced the tensor-valued measurement $\mathbf{v}$. For applications and analysis in this direction we refer to Cattell (1944), Tucker (1966), Harshman (1970), Appellof and Davidson (1981), Henrion (1994), De Lathauwer, De Moor and Vandewalle (2000*b*), Comon (2002), Smilde, Bro and Geladi (2004), Kroonenberg (2008); see also the review by Kolda and Bader (2009).

### *1.5. Software*

In various places software has been developed:

http://anchp.epfl.ch/htucker (Kressner and Tobler 2012)
http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox (Bader and Kolda 2007)
http://esat.kuleuven.be/sista/cot/ and http://esat.kuleuven.be/sista/tensorlab/ (by L. Sorber, M. Van Barel and L. De Lathauwer)
http://gitorious.org/tensorcalculus/pages/Home (by M. Espig *et al.*)
http://spring.inm.ras.ru/osel (TT TOOLBOX by I. Oseledets)
http://www.eigenvector.com (by B. M. Wise and N. B. Gallagher)
http://www.models.life.ku.dk/nwaytoolbox (details by Andersson and Bro 2000)

### *1.6. Preview*

We start with the definition of algebraic tensors in Section 2 and of topological tensors in Section 3, before discussing numerical issues. Suitable tensor formats are the key to efficient treatment, as explained in Section 4. The concrete formats are the $r$-term representation in Section 5, the tensor subspace format in Section 6, the hierarchical format in Section 7, and the TT format in Section 9. Besides the efficient representation of a tensor, we need algorithms that perform the standard tensor operations by means of the underlying format parameters. This is explained in Section 5.2 for

the $r$-term format. Operations in the important hierarchical format are discussed in Section 8. If the problem to be solved can be described by an optimization formulation, optimization with respect to the format parameters is discussed in Section 10. The multivariate cross approximation from Section 11 determines the tensor parameters from a small number of evaluations of the tensor. The tensorization described in Section 12 maps vectors into tensors and profits from the tensor compression. In the optimal case, this technique reduces the data size $n^d$ to $O(\log(n^d)) = O(d \log n)$. Finally, in Section 13, we briefly mention three application fields.

For a survey of literature in this field we refer to Grasedyck, Kressner and Tobler (2013).

## 2. Algebraic tensors

Let $V_j$ $(1 \le j \le d)$ be $d$ vector spaces (over the same field $\mathbb{K}$). In Section 2.1 we recall the definition of the algebraic tensor space $\mathbf{V} = {}_a\bigotimes_{j=1}^d V_j$, where the suffix $a$ indicates the *algebraic* nature. Later, in Section 3, we shall consider the *topological* tensor space ${}_{\|\cdot\|}\bigotimes_{j=1}^d V_j$ associated with a norm $\|\cdot\|$. Note that the vector spaces may be of any finite or infinite dimension. The choice of underlying field $\mathbb{K}$ is arbitrary.

### *2.1. Basic definitions and properties of tensors*

Let vector spaces $V_j$ $(1 \le j \le d)$ be given. The notation of the tensor space $\mathbf{V} = {}_a\bigotimes_{j=1}^d V_j$ means that there is a multilinear map $\otimes : V_1 \times \cdots \times V_d \to \mathbf{V}$ such that

$$\mathbf{V} = \operatorname{span}\left\{ \bigotimes_{j=1}^{d} v^{(j)} : v^{(j)} \in V_j \right\}. \tag{2.1}$$

The multilinear mapping $\otimes$ is universal in the sense that, for any multilinear mapping $\varphi : V_1 \times \cdots \times V_d \to U$ into some vector space $U$, there is a unique linear (not multilinear!) mapping $\Phi : \mathbf{V} \to U$ such that $\varphi = \Phi \circ \otimes$, that is, the following diagram is commutative:

$$\begin{array}{ccc} V_1 \times \cdots \times V_d & \xrightarrow{\ \varphi\ } & U \\ {\scriptstyle\otimes}\big\downarrow & \nearrow_{\Phi} & \\ {}_a\bigotimes_{j=1}^d V_j & & \end{array}$$

This characterization fixes the tensor space $\mathbf{V}$ up to (tensor space) isomorphisms; see Greub (1975, Chap. I, Sec. 2). Any product of the form $\bigotimes_{j=1}^d v^{(j)}$ is called an *elementary tensor*.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

An equivalent definition may use the following properties.

**Proposition 2.1.**

(a) The algebraic tensor space $\mathbf{V}$ is characterized by the following properties.

(i) The span property:[2]

$$\mathbf{V} = \operatorname{span}\left\{ \bigotimes_{j=1}^{d} v^{(j)} : v^{(j)} \in V_j \right\}.$$

(ii) Multilinearity: for all $\lambda \in \mathbb{K}$, $v^{(j)}, w^{(j)} \in V_j$, and $j \in \{1, \dots, d\}$,

$$\begin{aligned} v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes (\lambda v^{(j)} + w^{(j)}) \otimes \cdots \otimes v^{(d)} \\ = \lambda v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(j)} \otimes \cdots \otimes v^{(d)} \\ + v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes w^{(j)} \otimes \cdots \otimes v^{(d)}. \end{aligned}$$

(iii) Linearly independent vectors $\{v_i^{(j)} : i \in B_j\} \subset V_j$ $(1 \le j \le d)$ produce linearly independent vectors $\{\bigotimes_{j=1}^{d} v_{i_j}^{(j)} : i_j \in B_j\}$ in $\mathbf{V}$.

(b) Properties (i) and (iii) yield the equivalent statement that any bases $\{v_i^{(j)} : i \in B_j\}$ of $V_j$ define a basis $\{\bigotimes_{j=1}^{d} v_{i_j}^{(j)} : i_j \in B_j\}$ of $\mathbf{V}$.

(c) A simple consequence is that

$$\dim(\mathbf{V}) = \prod_{j=1}^{d} \dim(V_j) \tag{2.2}$$

(note that $\dim(V_j) = \#B_j$ may be infinite cardinalities).

For completeness we remark that for $d = 0$ the empty tensor product is defined by the field $\bigotimes_{j=1}^{0} V_j = \mathbb{K}$. For $d = 1$ the tensor space $\bigotimes_{j=1}^{1} V_j = V_1$ coincides with the generating vector space.

**Remark 2.2.** We have the following two degenerate cases.

(a) If $\dim(V_j) = 0$ for some $j$, $\mathbf{V} = \{0\}$ follows from (2.2).

(b) If $\dim(V_j) = 1$ for some $j$, $\mathbf{V} = {}_a\bigotimes_{j=1}^{d} V_j$ is isomorphic to the tensor space

$$\mathbf{V}_{[j]} = {}_a\bigotimes_{k \neq j} V_k$$

of order $d - 1$, where $\bigotimes_{k \neq j} = \bigotimes_{k \in \{1,\dots,d\} \setminus \{j\}}$.

[2] Note that $\operatorname{span}\{v_i : i \in I\}$ is the set of all *finite* linear combinations, even when the index set $I$ is infinite.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Consequently, the non-degenerate case is characterized by $\dim(V_j) \geq 2$ for all $1 \leq j \leq d$.

In the case of identical vector spaces $V := V_j$ for $1 \leq j \leq d$, *symmetric tensors* are defined by the property $\bigotimes_{j=1}^{d} v^{(j)} = \bigotimes_{j=1}^{d} v^{(\pi(j))}$ for any permutation $\pi$ of the set $\{1, \ldots, d\}$. Similarly, *antisymmetric tensors* are defined by the property $\bigotimes_{j=1}^{d} v^{(j)} = \operatorname{sign}(\pi) \bigotimes_{j=1}^{d} v^{(\pi(j))}$. The latter type is very important for quantum chemistry, since Pauli's principle requires antisymmetric functions for electrons. We shall not go into details concerning the numerical treatment of (anti)symmetric tensor, and refer to Mohlenkamp (2010) and Beylkin, Mohlenkamp and Pérez (2008).

### *2.2. Spaces of linear maps*

By definition (2.1), elementary tensors form a spanning system. This yields the following remark, which simplifies the definition of linear maps.

**Remark 2.3.** Let $\Phi$ be any linear mapping from the tensor space into some vector space $U$. Then $\Phi$ is uniquely defined by its action on elementary tensors.

The notation $L(X, Y)$ denotes the set of linear mappings from $X$ to $Y$ (see (1.3)). The vector space $V_j$ may, in particular, coincide with $L(X_j, Y_j)$. In the case of $X_j = \mathbb{K}^{n_j}$ and $Y_j = \mathbb{K}^{m_j}$, we get the matrix space $L(X_j, Y_j) = \mathbb{K}^{n_j \times m_j}$. The last statement of the next remark is proved in Hackbusch (2012, Example 3.53).

**Remark 2.4.** Let $X_j, Y_j$ be any vector spaces and set $W_j = L(X_j, Y_j)$. Define the corresponding tensor spaces by

$$\mathbf{W} = {}_a\bigotimes_{j=1}^{d} W_j, \quad \mathbf{X} = {}_a\bigotimes_{j=1}^{d} X_j, \quad \mathbf{Y} = {}_a\bigotimes_{j=1}^{d} Y_j.$$

Then $\mathbf{W}$ can be interpreted as a subspace of $L(\mathbf{X}, \mathbf{Y})$ via[3]

$$\left(\bigotimes_{j=1}^{d} A^{(j)}\right)\left(\bigotimes_{j=1}^{d} x^{(j)}\right) = \bigotimes_{j=1}^{d} \left[A^{(j)} x^{(j)}\right] \quad \begin{cases} A^{(j)} \in W_j = L(X_j, Y_j), \\ x^{(j)} \in X_j. \end{cases}$$

For finite dimensions, equality $\mathbf{W} = L(\mathbf{X}, \mathbf{Y})$ holds, whereas $\mathbf{W} \subsetneq L(\mathbf{X}, \mathbf{Y})$ is a proper subspace in the infinite-dimensional case.

In the case of matrices, the tensor product $\bigotimes_{j=1}^{d} \mathbb{K}^{n_j \times m_j}$ is also called the *Kronecker product* (see Section 1.1.2).

[3] Here we make use of Remark 2.3. This equation defines the linear map $\bigotimes_{j=1}^{d} A^{(j)}$. Any element of $\mathbf{W}$ is a linear combination of such elementary tensors $\bigotimes_{j=1}^{d} A^{(j)}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The (algebraic) dual of a vector space is $V' = L(V, \mathbb{K})$ containing the functionals on $V$. Setting $X_j = V_j$ and $Y_j = \mathbb{K}$, we can apply the previous remark. Here, we have to exploit that $\bigotimes_{j=1}^d \mathbb{K} = \mathbb{K}$ (see Remark 2.2). Together, we obtain the following result (Hackbusch 2012, § 3.3.2.2).

**Remark 2.5.** Let $\mathbf{V} = {}_a\bigotimes_{j=1}^d V_j$. The tensor space $\mathbf{W} = {}_a\bigotimes_{j=1}^d V_j'$ can be interpreted as a subspace of $\mathbf{V}'$ via

$$\left(\bigotimes_{j=1}^d \underbrace{\varphi^{(j)}}_{\in V_j'}\right)\left(\bigotimes_{j=1}^d \underbrace{v^{(j)}}_{\in V_j}\right) = \prod_{j=1}^d \underbrace{\varphi^{(j)}(v^{(j)})}_{\in\mathbb{K}}.$$

Equality $\mathbf{W} = \mathbf{V}'$ holds only in the finite-dimensional case.

Consider a fixed index $j \in \{1, \dots, d\}$ and let $\varphi^{(j)} \in V_j'$. We can extend the action of $\varphi^{(j)}$ to $\mathbf{V}$ by

$$\hat\varphi^{(j)} := \mathrm{id}_1 \otimes \cdots \otimes \mathrm{id}_{j-1} \otimes \varphi^{(j)} \otimes \mathrm{id}_{j+1} \otimes \cdots \otimes \mathrm{id}_d,$$

where $\mathrm{id}_k \in L(V_k, V_k)$ is the identity map. Remark 2.4 can be applied with $Y_j = \mathbb{K}$ and $Y_k = V_k$ for $k \neq j$. Note that $\bigotimes_{k=1}^d Y_k = \bigotimes_{k\neq j} V_k$ because of $Y_j = \mathbb{K}$ (see Remark 2.2). Therefore,

$$\hat\varphi^{(j)} \in L\left({}_a\bigotimes_{k=1}^d V_k,\ {}_a\bigotimes_{k\neq j} V_j\right) \quad \text{with } \hat\varphi^{(j)}\left(\bigotimes_{k=1}^d v^{(k)}\right) = \varphi^{(j)}(v^{(j)}) \cdot \bigotimes_{k\neq j} v^{(k)}. \tag{2.3}$$

### *2.3. Isomorphisms and transformations*

We have to distinguish two types of isomorphisms. In the sense of a *vector space isomorphism*, we have $V_1 \otimes V_2 \otimes V_3 \cong (V_1 \otimes V_2) \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3)$, since all spaces have the same dimension. A stronger equivalence relation is the following isomorphism. Two tensor spaces are *isomorphic in the sense of tensor spaces* if they are of the same order, $\mathbf{V} = {}_a\bigotimes_{j=1}^d V_j$ and $\mathbf{W} = {}_a\bigotimes_{j=1}^d W_j$ with identical $d$, and if $V_j \cong W_j$ are isomorphic for all $1 \le j \le d$.

Let $\{b_\nu'^{(j)} : \nu \in I_j\}$ and $\{b_\nu''^{(j)} : \nu \in I_j\}$ be two bases of $V_j$. By $W_j := \ell_0(I_j)$ we denote the set of sequences $(a_i)_{i\in I_j}$ with only finitely many non-zero entries. Let $\phi', \phi'' : W_j \to V_j$ be the isomorphisms defined via

$$\phi'(w^{(j)}) = \sum_{i\in I_j} w_i^{(j)} b_i'^{(j)} \quad \text{and} \quad \phi''(w^{(j)}) = \sum_{i\in I_j} w_i^{(j)} b_i''^{(j)}.$$

The change from basis $\{b_\nu'^{(j)}\}$ to basis $\{b_\nu''^{(j)}\}$ corresponds to the transformation $T_j := \phi''^{-1}\phi' : W_j \to W_j$. The tensor space $\mathbf{V} = {}_a\bigotimes_{j=1}^d V_j$ with

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

the bases

$$\mathbf{b}'_{\boldsymbol{\nu}} := \bigotimes_{j=1}^{d} b'^{(j)}_{\nu_j} \quad \text{or} \quad \mathbf{b}''_{\boldsymbol{\nu}} := \bigotimes_{j=1}^{d} b''^{(j)}_{\nu_j}$$

(see Proposition 2.1(b)) is isomorphic to $\mathbf{W} = {}_a\bigotimes_{j=1}^{d} W_j$. The basis transformation from $\{\mathbf{b}'_{\boldsymbol{\nu}}\}$ to $\{\mathbf{b}''_{\boldsymbol{\nu}}\}$ corresponds to

$$\mathbf{T} := \bigotimes_{j=1}^{d} T_j : \mathbf{W} \to \mathbf{W}$$

with $T_j$ from above.

### *2.4. Tensor rank*

The rank of a matrix is a very useful quantity in linear algebra. If one tries to generalize the rank to the tensor case $d \geq 3$, different – but equivalent – characterizations of the matrix rank lead to different and non-equivalent rank concepts for tensors. Below we introduce the *tensor rank* as a non-negative integer from $\mathbb{N}_0$. Another approach, used in Section 2.6, leads to a tuple $\mathbf{r} = (r_1, \ldots, r_d)$ from $\mathbb{N}_0^d$. One may even introduce ranks $r_\alpha$ for any subset $\alpha \subset \{1, \ldots, d\}$ (see (2.10)).

A possible characterization of the rank of a matrix $M \in \mathbb{K}^{n\times m}$ is as follows. If $r = \operatorname{rank}(M)$, then there exist $r$ vectors $a_\nu \in \mathbb{K}^n$ and $b_\nu \in \mathbb{K}^m$ such that $M = \sum_{\nu=1}^{r} a_\nu b_\nu^{\mathsf{T}}$ (see (1.4)) and there is no such representation with fewer terms than $r$.

In the tensor case we follow the definition (2.1), which states that any algebraic tensor has a representation

$$\mathbf{v} = \sum_{\nu=1}^{r} v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$$

with some $r \in \mathbb{N}_0$ and $v_\nu^{(j)} \in V_j$ ($v_\nu^{(j)} = 0$ is not excluded). Define the set of all linear combinations of $r$ elementary tensors by

$$\mathcal{R}_r := \mathcal{R}_r(\mathbf{V}) := \left\{ \sum_{\nu=1}^{r} v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)} : \; v_\nu^{(j)} \in V_j \right\} \quad (r \in \mathbb{N}_0). \tag{2.4}$$

The span property (2.1) implies that $\mathbf{V} = \bigcup_{r\in\mathbb{N}_0} \mathcal{R}_r$. The sets $\mathcal{R}_r$ are nested:

$$\{0\} = \mathcal{R}_0 \subset \mathcal{R}_1 \subset \cdots \subset \mathcal{R}_{r-1} \subset \mathcal{R}_r \subset \cdots \subset \mathbf{V} \quad \text{for all } r \in \mathbb{N}.$$

The previous characterization of the matrix rank leads us to the following generalization.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Definition 2.6 (tensor rank).** The *tensor rank* of $\mathbf{v} \in {}_a\bigotimes_{j=1}^{d} V_j$ is defined by

$$\operatorname{rank}(\mathbf{v}) := \min\{r : \mathbf{v} \in \mathcal{R}_r\} \in \mathbb{N}_0. \tag{2.5}$$

We may redefine $\mathcal{R}_r$ by $\mathcal{R}_r = \{\mathbf{v} \in \mathbf{V} : \operatorname{rank}(\mathbf{v}) \le r\}$. We shall use the shorter term 'rank' instead of 'tensor rank' (the terms 'canonical rank' or 'CP rank' are also used). Note that there is an ambiguity if $\mathbf{v}$ is a matrix as well as a tensor (*e.g.*, in (1.2)). If necessary, we use the explicit terms 'matrix rank' and 'tensor rank'.

**Remark 2.7.** The tensor rank is invariant under (tensor space) isomorphisms of $V_j$. More generally, application of an operator $\mathbf{A} = \bigotimes_{j=1}^{d} A^{(j)}$ with $A^{(j)} \in L(V_j, W_j)$ yields $\operatorname{rank}(A\mathbf{v}) \le \operatorname{rank}(\mathbf{v})$.

*Proof.* By definition, there is a representation $\mathbf{v} = \sum_{\nu=1}^{r} \bigotimes_{j=1}^{d} v_\nu^{(j)}$ with $r := \operatorname{rank}(\mathbf{v})$. Hence,

$$\mathbf{w} = \mathbf{A}\mathbf{v} = \sum_{\nu=1}^{r} \bigotimes_{j=1}^{d} [A^{(j)} v_\nu^{(j)}]$$

requires at most $r$ terms, that is, $\operatorname{rank}(\mathbf{w}) \le r$. In the case of an isomorphism, apply the same argument to $\mathbf{A}^{-1}\mathbf{w} = \mathbf{v}$, where

$$\mathbf{A}^{-1} = \bigotimes_{j=1}^{d} (A^{(j)})^{-1}. \qquad \square$$

**Remark 2.8.** Unlike matrix rank, the tensor rank has rather obscure properties.

(a) In general, the determination of $\operatorname{rank}(\mathbf{v})$ is NP hard (Håstad 1990).

(b) For finite-dimensional $V_j$, there is a *maximal rank* $r_{\max}$ (*i.e.*, $r_{\max}$ is the smallest integer with $\mathcal{R}_{r_{\max}} = \mathbf{V}$), but in general $r_{\max}$ is not explicitly known. For equal dimensions $\dim(V_j) = n$, the inequalities

$$\frac{d}{2(d-1)} n^{d-1} + O(n^{d-2}) \ge r_{\max} \ge n^{d-1}/d \tag{2.6}$$

follow from Howell (1978).

(c) While random matrices have maximal rank with probability one, there may be more than one tensor rank with positive probability. These ranks are called *typical*.

(d) If $\mathbf{v} \in \mathbf{V}_{\mathbb{R}}$ holds for the tensor space over the field $\mathbb{R}$, it also belongs to the larger space $\mathbf{V}_{\mathbb{C}}$ considered over $\mathbb{C}$. Unlike the matrix case, $\mathbf{v}$ may possess different ranks in $\mathbf{V}_{\mathbb{R}}$ and $\mathbf{V}_{\mathbb{C}}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

(e) While the limit of matrices of rank $\le r$ again has a rank bounded by $r$, the tensor rank may increase in the limit. This means that, in general, $\mathcal{R}_r$ is not closed (see Section 5.4).

### *2.5. Application: Strassen's algorithm*

It might be surprising that the rank of a tensor is connected with the complexity of the matrix–matrix multiplication. As an illustration, we recall the algorithm of Strassen.

The standard matrix–matrix multiplication of $n \times n$ matrices costs $2n^3$ operations. A reduction to $4.7n^{\log_2 7} = 4.7n^{2.8074}$ proposed by Strassen (1969) is based on the fact that two $2 \times 2$ block matrices can be multiplied as follows:

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}, \quad a_i, b_i, c_i \text{ submatrices with} \tag{2.7}$$

$$c_1 = m_1 + m_4 - m_5 + m_7, \quad c_2 = m_2 + m_4,$$
$$c_3 = m_3 + m_5, \quad c_4 = m_1 + m_3 - m_2 + m_6,$$
$$m_1 = (a_1 + a_4)(b_1 + b_4), \quad m_2 = (a_3 + a_4)b_1, \quad m_3 = a_1(b_2 - b_4),$$
$$m_4 = a_4(b_3 - b_1), \quad m_5 = (a_1 + a_2)b_4,$$
$$m_6 = (a_3 - a_1)(b_1 + b_2), \quad m_7 = (a_2 - a_4)(b_3 + b_4),$$

where only seven multiplications of block matrices are needed.

We introduce a tensor $\mathbf{v} \in \mathbb{K}^{4\times4\times4}$, whose entries are defined by

$$c_\nu = \sum_{\mu,\lambda=1}^{4} \mathbf{v}_{\nu\mu\lambda}\, a_\mu\, b_\lambda \quad (1 \le \nu \le 4). \tag{2.8a}$$

For instance, for $\nu = 1$ the identity $c_1 = a_1b_1 + a_2b_3$ shows that $\mathbf{v}_{111} = \mathbf{v}_{123} = 1$, and $\mathbf{v}_{1\mu\lambda} = 0$ otherwise. Assume a representation of $\mathbf{v}$ in $r$ terms:

$$\mathbf{v} = \sum_{i=1}^{r} \bigotimes_{j=1}^{3} v_i^{(j)}. \tag{2.8b}$$

Insertion into (2.8a) yields

$$\begin{aligned} c_\nu &= \sum_{i=1}^{r} \sum_{\mu,\lambda=1}^{4} v_i^{(1)}[\nu]\, v_i^{(2)}[\mu]\, v_i^{(3)}[\lambda]\, a_\mu\, b_\lambda \\ &= \sum_{i=1}^{r} v_i^{(1)}[\nu] \left( \sum_{\mu=1}^{4} v_i^{(2)}[\mu]\, a_\mu \right) \left( \sum_{\lambda=1}^{4} v_i^{(3)}[\lambda]\, b_\lambda \right), \end{aligned}$$

that is, only $r$ multiplications are needed. Algorithm (2.7) corresponds to a representation (2.8b) with $r = 7$, implying $\operatorname{rank}(\mathbf{v}) \le 7$. In fact, Winograd

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

(1971) proves that $\operatorname{rank}(\mathbf{v}) = 7$. For a detailed discussion of the complexity of matrix multiplication we refer to Landsberg (2012, Chap. 11).

### *2.6. Matricization and Tucker ranks*

For the time being assume that $V_j = \mathbb{K}^{n_j}$. Elements $\mathbf{v} \in \mathbf{V}$ have the entries $\mathbf{v}[i_1, \ldots, i_d]$ with $1 \le i_j \le n_j$. We regroup the indices into a single index $i_j$ and a $(d-1)$-tuple $\mathbf{i}_{[j]} := (i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d)$. Define the matrix $M_j := \mathcal{M}_j(\mathbf{v}) \in \mathbb{K}^{n_j \times n_{[j]}}$ with $n_{[j]} := \prod_{k \neq j} n_k$ by the entries

$$M_j[i_j, \mathbf{i}_{[j]}] := \mathbf{v}[i_1, \ldots, i_d]$$

(the ordering of the indices will be irrelevant). Note that $\mathcal{M}_j : \mathbf{V} \to \mathbb{K}^{n_j \times n_{[j]}}$ is a vector space isomorphism. $\mathcal{M}_j$ transfers tensors into matrices and is called *matricization* or *unfolding*. In particular, we can define the matrix rank (the so-called Tucker rank or $j$th rank)

$$r_j = \operatorname{rank}_j(\mathbf{v}) := \operatorname{rank}(\mathcal{M}_j(\mathbf{v})) \quad \text{for } 1 \le j \le d. \tag{2.9}$$

Note that for matrices $M$ we have $\mathcal{M}_1(M) = M$ and $\mathcal{M}_2(M) = M^{\mathsf{T}}$. Therefore $r_1 = r_2$ is the standard matrix rank. For $d \ge 3$, however, all $r_j$ may be different. We can form the tuple

$$\mathbf{r} := (r_1, \ldots, r_d) \quad \text{with } r_j \text{ from (2.9)},$$

which is sometimes called the *multilinear rank* of $\mathbf{v}$.

**Example 2.9.** Consider a tensor $\mathbf{v}$ from the 16-dimensional tensor space $\mathbf{V} := \mathbb{K}^2 \otimes \mathbb{K}^2 \otimes \mathbb{K}^2 \otimes \mathbb{K}^2$. Then, for example, $\mathcal{M}_2(\mathbf{v})$ belongs to $\mathbb{K}^{2\times 8}$:

$$\mathcal{M}_2(\mathbf{v}) = \begin{pmatrix} \mathbf{v}_{1\mathbf{1}11} & \mathbf{v}_{1\mathbf{1}12} & \mathbf{v}_{1\mathbf{1}21} & \mathbf{v}_{1\mathbf{1}22} & \mathbf{v}_{2\mathbf{1}11} & \mathbf{v}_{2\mathbf{1}12} & \mathbf{v}_{2\mathbf{1}21} & \mathbf{v}_{2\mathbf{1}22} \\ \mathbf{v}_{1\mathbf{2}11} & \mathbf{v}_{1\mathbf{2}12} & \mathbf{v}_{1\mathbf{2}21} & \mathbf{v}_{1\mathbf{2}22} & \mathbf{v}_{2\mathbf{2}11} & \mathbf{v}_{2\mathbf{2}12} & \mathbf{v}_{2\mathbf{2}21} & \mathbf{v}_{2\mathbf{2}22} \end{pmatrix}.$$

Next, we consider $\alpha = \{1, 3\}$. $\mathcal{M}_{\{1,3\}}(\mathbf{v})$ belongs to $\mathbb{K}^{4\times 4}$:

$$\mathcal{M}_{\{1,3\}}(\mathbf{v}) = \begin{pmatrix} \mathbf{v}_{1\mathbf{1}1\mathbf{1}} & \mathbf{v}_{1\mathbf{1}1\mathbf{2}} & \mathbf{v}_{1\mathbf{2}1\mathbf{1}} & \mathbf{v}_{1\mathbf{2}1\mathbf{2}} \\ \mathbf{v}_{1\mathbf{1}2\mathbf{1}} & \mathbf{v}_{1\mathbf{1}2\mathbf{2}} & \mathbf{v}_{1\mathbf{2}2\mathbf{1}} & \mathbf{v}_{1\mathbf{2}2\mathbf{2}} \\ \mathbf{v}_{2\mathbf{1}1\mathbf{1}} & \mathbf{v}_{2\mathbf{1}1\mathbf{2}} & \mathbf{v}_{2\mathbf{2}1\mathbf{1}} & \mathbf{v}_{2\mathbf{2}1\mathbf{2}} \\ \mathbf{v}_{2\mathbf{1}2\mathbf{1}} & \mathbf{v}_{2\mathbf{1}2\mathbf{2}} & \mathbf{v}_{2\mathbf{2}2\mathbf{1}} & \mathbf{v}_{2\mathbf{2}2\mathbf{2}} \end{pmatrix}.$$

The boldface subscripts characterize the rows, while the other ones indicate the columns.

Given a subset $\alpha \subset \{1, \ldots, d\}$ and its complement $\alpha^c := \{1, \ldots, d\} \backslash \alpha$ with[4] $\alpha \neq \emptyset \neq \alpha^c$, the $\alpha$th rank can be defined by

$$\operatorname{rank}_\alpha(\mathbf{v}) := \operatorname{rank}(\mathcal{M}_\alpha(\mathbf{v})) \quad \text{with} \quad \mathcal{M}_\alpha(\mathbf{v})[\mathbf{i}_\alpha, \mathbf{i}_{\alpha^c}] := \mathbf{v}[i_1, \ldots, i_d], \tag{2.10}$$

[4] Formally, we may set $r_\emptyset = r_{\{1,\ldots,d\}} := 1$ for $\mathbf{v} \neq 0$ and $r_\emptyset = r_{\{1,\ldots,d\}} := 0$ for $\mathbf{v} = 0$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

where $\mathbf{i}_\alpha = (i_j)_{j\in\alpha}$ and $\mathbf{i}_{\alpha^c} = (i_j)_{j\in\alpha^c}$. The definition of such ranks can already be found in Hitchcock (1927, p. 170).

Considering a matrix $M \in \mathbb{K}^{n\times m}$ as a mapping $M : \mathbb{K}^m \to \mathbb{K}^n$, the rank can be defined by $\operatorname{rank}(M) := \dim(\operatorname{range}(M))$. The latter definition can be exploited for general finite-dimensional vector spaces $V_j$. The tensor space $\mathbf{V} = \bigotimes_{j=1}^d V_j$ is isomorphic (in the sense of vector spaces) to

$$\hat{\mathbf{V}} := V_j \otimes \mathbf{V}_{[j]} \quad \text{with} \quad \mathbf{V}_{[j]} := \bigotimes_{k\neq j} V_k.$$

In the finite-dimensional case, $\hat{\mathbf{V}}$ can be identified with the space $L(\mathbf{V}_{[j]}, V_j)$ (see Remark 1.1). Therefore, a tensor $\mathbf{v} \in \mathbf{V}$ corresponds to a mapping $\hat{\mathbf{v}} \in L(\mathbf{V}_{[j]}, V_j)$. Its range defines a subspace $U_j \subset V_j$. The definition $\operatorname{rank}_j(\mathbf{v}) := \dim(U_j)$ coincides with (2.9) for $V_j = \mathbb{K}^{n_j}$. A similar definition holds for the infinite-dimensional case (see (6.1) and Section 6.4).

The invariance property of Remark 2.7 also holds for the newly introduced ranks.

**Remark 2.10.** $\operatorname{rank}_j(\cdot)$ and $\operatorname{rank}_\alpha(\cdot)$ are invariant under tensor space isomorphisms.

The relation to the previously introduced tensor rank is given by[5]

$$\operatorname{rank}_j(\mathbf{v}) \le \operatorname{rank}(\mathbf{v}), \quad \operatorname{rank}_\alpha(\mathbf{v}) \le \operatorname{rank}(\mathbf{v})$$

for all $1 \le j \le d$ and $\alpha \subset \{1, \dots, d\}$.

## 3. Topological tensors

So far, the spaces have not been endowed with any topology. Now we assume that $V_j$ are normed spaces, or even Banach spaces (*i.e.*, complete normed spaces), where the norm is denoted by $\|\cdot\|_j$. Further, the tensor space $\mathbf{V}$ is equipped with a norm denoted by $\|\cdot\|$. In the case of the algebraic tensor construction, $\mathbf{V}$ is uniquely defined by $V_j$ (up to isomorphism). This is different for the topological case: the norms in $(V_j, \|\cdot\|_j)$ do not determine the norm of $(\mathbf{V}, \|\cdot\|)$ in a unique way. Next, we shall discuss various properties of $\|\cdot\|$ and define the projective and the injective crossnorms. The case of Hilbert spaces will be investigated in Section 3.4.

For statements in this section without proof we refer to Hackbusch (2012, § 4), where complete proofs are given.

[5] In the estimates, $\operatorname{rank}(\mathbf{v})$ can be replaced by the possibly smaller border rank defined in (5.8).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*3.1. Continuity of the tensor product*

As mentioned above, we are free to choose some norm on the algebraic tensor space $\mathbf{V}_{\rm alg} = {}_a\bigotimes_{j=1}^d V_j$ and to define $\mathbf{V}_{\rm top}$ as the closure of $\mathbf{V}_{\rm alg}$ with respect to the chosen norm. The notation for $\mathbf{V}_{\rm top}$ is ${}_{\|\cdot\|}\bigotimes_{j=1}^d V_j$. Since $\mathbf{V}_{\rm top} = \mathbf{V}_{\rm alg}$ for the finite-dimensional case, the following discussion is of particular interest for the infinite-dimensional case.

Even when the norm $\|\cdot\|$ of $\mathbf{V}_{\rm top}$ is not fixed by the norms $\|\cdot\|_j$ of $V_j$, there is an important restriction to the choice of $\|\cdot\|$. We must always ensure continuity of the tensor product, that is, of the mapping

$$\bigotimes : V_1 \times \cdots \times V_d \to \mathbf{V} \quad \text{via} \quad \left(v^{(1)}, \ldots, v^{(d)}\right) \mapsto \bigotimes_{j=1}^d v^{(j)}.$$

For linear and multilinear mappings, continuity and boundedness are equivalent, so the continuity of the tensor product can be expressed by

$$\left\| \bigotimes_{j=1}^d v^{(j)} \right\| \le C \prod_{j=1}^d \|v^{(j)}\|_j \quad \text{for all } v^{(j)} \in V_j \quad (1 \le j \le d), \tag{3.1}$$

where $C$ is a fixed constant. This inequality states that the norm $\|\cdot\|$ cannot be too strong. In fact, we shall characterize the strongest possible norm in (3.3).

**General assumption.** In the following, we always assume that the continuity requirement (3.1) is satisfied.

We have not yet specified the completeness, or otherwise, of the normed spaces $V_j$. In fact, completeness is irrelevant for the definition of ${}_{\|\cdot\|}\bigotimes_{j=1}^d V_j$, as can be seen from the following remark.

**Remark 3.1.** Let $V_j$ $(1 \le j \le d)$ be Banach spaces, while $V_{0,j}$ are dense subspaces. Then the topological tensor spaces generated are equal:

$${}_{\|\cdot\|}\bigotimes_{j=1}^d V_j = {}_{\|\cdot\|}\bigotimes_{j=1}^d V_{0,j}.$$

*3.2. Crossnorms and projective norms*

One may attempt to define a norm on $\mathbf{V}$ by means of $\|\cdot\|_j$ by setting

$$\left\| \bigotimes_{j=1}^d v^{(j)} \right\| = \prod_{j=1}^d \|v^{(j)}\|_j \quad \text{for all } v^{(j)} \in V_j \quad (1 \le j \le d). \tag{3.2}$$

However, this definition concerns only the elementary tensors $\bigotimes_{j=1}^d v^{(j)}$. There exist different norms which coincide on the set of elementary tensors, that is, condition (3.2) does not fix the norm $\|\cdot\|$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Definition 3.2.** A norm on $\mathbf{V}$ satisfying (3.2) is called a crossnorm.

Note that a crossnorm satisfies the continuity requirement (3.1) with the constant $C = 1$.

Condition (3.1) describes an upper bound for norms on $\mathbf{V}$. Therefore one may ask for the strongest norm satisfying (3.1).

For $\mathbf{v} \in {}_a\bigotimes_{j=1}^d V_j$, define $\|\cdot\|_{\wedge(V_1,\dots V_d)}$ by

$$\|\mathbf{v}\|_{\wedge(V_1,\dots V_d)} := \|\mathbf{v}\|_\wedge := \inf\left\{\sum_{i=1}^n \prod_{j=1}^d \|v_i^{(j)}\|_j : \mathbf{v} = \sum_{i=1}^n \bigotimes_{j=1}^d v_i^{(j)}\right\}. \tag{3.3}$$

This defines the *projective norm* on $\mathbf{V}_{\mathrm{alg}}$, and $\mathbf{V}_{\mathrm{top}} := {}_\wedge\bigotimes_{j=1}^d V_j$ is the corresponding Banach tensor space.

**Theorem 3.3.** $\|\cdot\|_\wedge$ is a crossnorm satisfying the particular property (3.2). Moreover, it is the strongest norm satisfying (3.1).

*Proof.* We give the proof of the last statement in the case of $d = 2$. Let $\|\cdot\|$ be a norm on $V \otimes_a W$ such that $(v, w) \mapsto v \otimes w$ is continuous. Then (3.1) holds, that is, $\|v \otimes w\| \le C\|v\|_V\|w\|_W$. Let

$$\mathbf{x} = \sum_{i=1}^n v_i \otimes w_i \in V \otimes_a W.$$

The triangle inequality yields

$$\|\mathbf{x}\| \le \sum_{i=1}^n \|v_i \otimes w_i\|.$$

Together with (3.1),

$$\|\mathbf{x}\| \le C \sum_{i=1}^n \|v_i\|_V\|w_i\|_W$$

follows. Taking the infimum over all representations $\mathbf{x} = \sum_{i=1}^n v_i \otimes w_i$, the inequality $\|\mathbf{x}\| \le C\|\mathbf{x}\|_\wedge$ follows, that is, $\|\cdot\|_\wedge$ is stronger than $\|\cdot\|$. Since $\|\cdot\|$ is arbitrary under the continuity side condition, $\|\cdot\|_\wedge$ is the strongest norm. □

### *3.3. Tensor spaces of functionals and injective norms*

Another crossnorm is the injective norm, which we shall need in Section 6.4, and for which the tensor space of the dual spaces must be defined.

Let $V_j^* = \mathcal{L}(V_j, \mathbb{K}) \subset V_j' = L(V_j, \mathbb{K})$ be the Banach space of the continuous linear functionals on $V_j$ with the dual norm

$$\|\varphi\|_j^* = \|\varphi\|_{V_j^*} := \sup\{|\varphi(v)| : v \in V_j, \|v_i\|_j \le 1\}.$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The algebraic tensor space ${}_a\bigotimes_{j=1}^d V_j^*$ is a subspace of ${}_a\bigotimes_{j=1}^d V_j'$. Therefore it can be interpreted as a subspace of the algebraic dual $\mathbf{V}'$, as explained in Remark 2.5. To ensure that ${}_a\bigotimes_{j=1}^d V_j^* \subset \mathbf{V}^*$ holds too, we have to require that

$$\left\| \bigotimes_{j=1}^d \varphi^{(j)} \right\|^* \le C \prod_{j=1}^d \|\varphi^{(j)}\|_j^* \quad \text{for all } \varphi^{(j)} \in V_j^* \quad (1 \le j \le d), \tag{3.4}$$

that is, the tensor product is continuous with respect to the dual norms.

Again we can ask for the strongest dual norm $\|\cdot\|^*$ so that (3.4) is valid. Since a weaker norm leads to a stronger dual norm, we can equivalently seek for the weakest norm of $\mathbf{V}$ fulfilling (3.4).

The injective norm $\|\cdot\|_\vee$ is defined by

$$\|\mathbf{v}\|_{\vee(V_1,\dots V_d)} := \|\mathbf{v}\|_\vee := \sup\left\{ \left| \left( \bigotimes_{j=1}^d \varphi^{(j)} \right)(\mathbf{v}) \right| : \varphi^{(j)} \in V_j^*, \|\varphi^{(j)}\|_j^* = 1 \right\}.$$

**Theorem 3.4.** $\|\cdot\|_{\vee(V_1,\dots V_d)}$ is a norm satisfying the particular crossnorm property

$$\left\| \bigotimes_{j=1}^d \varphi^{(j)} \right\|^* = \prod_{j=1}^d \|\varphi^{(j)}\|_j^*. \tag{3.5}$$

Moreover, it is the weakest norm satisfying (3.4).

We conclude that all norms with $\|\cdot\|_{\vee(V_1,\dots V_d)} \lesssim \|\cdot\| \lesssim \|\cdot\|_{\wedge(V_1,\dots V_d)}$ satisfy both requirements (3.1) and (3.4).

**Definition 3.5 (reasonable crossnorm).** A norm satisfying both identities (3.2) and (3.5) is called a 'reasonable' crossnorm.

Examples of reasonable crossnorms are $\|\cdot\|_\vee$ and $\|\cdot\|_\wedge$.

### *3.4. Hilbert spaces*

#### *3.4.1. Canonical scalar product*

Now we assume that $V_j$ are (pre-)Hilbert spaces associated with scalar products $\langle\cdot,\cdot\rangle_j$ and corresponding norms $\|\cdot\|_j$. Further, the algebraic tensor space $\mathbf{V}_{\rm alg} = {}_a\bigotimes_{j=1}^d V_j$ is equipped with a scalar product $\langle\cdot,\cdot\rangle$ and corresponding norm $\|\cdot\|$ (hence it is a pre-Hilbert space). The closure with respect to $\|\cdot\|$ yields the topological Hilbert tensor space $\mathbf{V} = {}_{\|\cdot\|}\bigotimes_{j=1}^d V_j$.

In contrast to the Banach case, there is a *canonical scalar product* on $\mathbf{V}_{\rm alg}$ uniquely generated by the scalar products $\langle\cdot,\cdot\rangle_j$. For elementary tensors

$$\bigotimes_{j=1}^d v^{(j)}, \quad \bigotimes_{j=1}^d w^{(j)} \in \mathbf{V}_{\rm alg}$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

we define

$$\left\langle \bigotimes_{j=1}^{d} v^{(j)}, \bigotimes_{j=1}^{d} w^{(j)} \right\rangle := \prod_{j=1}^{d} \langle v^{(j)}, w^{(j)} \rangle_j \quad \text{for all } v^{(j)},\, w^{(j)} \in V_j. \tag{3.6}$$

The statement of Remark 2.3 can easily be extended to bilinear[6] maps: a bilinear mapping is uniquely defined by its action on elementary tensors. Since the scalar product is bilinear,[6] definition (3.6) yields a bilinear[6] map $\langle \cdot, \cdot \rangle$ on $\mathbf{V}_{\mathrm{alg}}$. We check that the other properties of the scalar product (symmetry and positivity) also hold (Hackbusch 2012, Lemma 4.124).

The norm $\|\cdot\|$ derived from the scalar product has pleasant properties; in particular, the general requirement (3.1) is met.

**Remark 3.6.** The norm $\|\cdot\|$ corresponding to the canonical scalar product is a reasonable crossnorm.

*Proof.* It is a crossnorm, since

$$\left\| \bigotimes_{j=1}^{d} v^{(j)} \right\|^2 = \left\langle \bigotimes_{j=1}^{d} v^{(j)}, \bigotimes_{j=1}^{d} v^{(j)} \right\rangle := \prod_{j=1}^{d} \langle v^{(j)}, v^{(j)} \rangle_j = \prod_{j=1}^{d} \|v^{(j)}\|_j^2.$$

Here $V_j$ and $\mathbf{V}$ can be identified with their duals. Therefore, the equality above also implies (3.5). □

The standard examples of a Hilbert tensor space with canonical scalar product are the $L^2$ function spaces.

**Example 3.7.** Let $I_j \subset \mathbb{R}$ be intervals and let $V_j := L^2(I_j)$ be Hilbert spaces with the usual $L^2$ scalar product. Then the canonical scalar product from (3.6) yields the $L^2$ scalar product

$$\langle f, g \rangle_{L^2} = \int_\Omega f(x_1, \ldots, x_d)\, \overline{g(x_1, \ldots, x_d)}\, \mathrm{d}x_1 \cdots \mathrm{d}x_d$$

of

$$L^2(\Omega) = {}_{L^2}\bigotimes_{j=1}^{d} L^2(I_j) \quad \text{with } \Omega := I_1 \times \cdots \times I_d.$$

*3.4.2. Other scalar products and mixed spaces*

The Hilbert tensor space can also be based on a scalar product different from the canonical scalar product. The simplest example is the Sobolev space $H^1(\Omega)$ with the scalar product

$$\langle f, g \rangle_{H^1} = \langle f, g \rangle_{L^2} + \sum_{j=1}^{d} \left\langle \frac{\partial f}{\partial x_j}, \frac{\partial g}{\partial x_j} \right\rangle_{L^2},$$

[6] Or sesquilinear in the case of $\mathbb{K} = \mathbb{C}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

with $\Omega$ and $\langle\cdot,\cdot\rangle_{L^2}$ as above. It is the topological tensor space

$$H^1(\Omega) = {}_{H^1}\bigotimes_{j=1}^{d} H^1(I_j).$$

The canonical scalar product induced by the scalar products of $H^1(I_j)$ is different from the $H^1(\Omega)$ scalar product. It is the so-called mixed $H^1$ scalar product

$$\langle f, g\rangle_{H^1_{\mathrm{mix}}} = \sum_{\boldsymbol{\nu}} \langle \partial^{\boldsymbol{\nu}} f, \partial^{\boldsymbol{\nu}} g\rangle_{L^2} \quad \text{with } \partial^{\boldsymbol{\nu}} = \left(\frac{\partial}{\partial x_1}\right)^{\nu_1} \cdots \left(\frac{\partial}{\partial x_d}\right)^{\nu_d},$$

where the sum is taken over all multi-indices $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)$ with $0 \le \nu_j \le 1$. Note that the mixed derivatives for $\nu_j = 1$ are of order $d$. The Hilbert space

$$H^1_{\mathrm{mix}}(\Omega) = {}_{H^1_{\mathrm{mix}}}\bigotimes_{j=1}^{d} H^1(I_j)$$

is a proper[7] subspace of $H^1(\Omega)$. Equivalently, the norm $\|\cdot\|_{H^1_{\mathrm{mix}}}$ is strictly stronger than $\|\cdot\|_{H^1}$.

By definition (3.6), the canonical scalar product is well-defined on elementary tensors and, therefore, also on any finite linear combination. This yields the following statement.

**Remark 3.8.** The algebraic tensor product ${}_a\bigotimes_{j=1}^{d} H^1(I_j)$ is a subspace of $H^1_{\mathrm{mix}}(\Omega)$. Consequently, a function $f \in H^1(\Omega)\backslash H^1_{\mathrm{mix}}(\Omega)$ can be the limit of a sum of $r$ elementary tensors with arbitrarily slow convergence as $r \to \infty$. Therefore, the boundedness of mixed derivatives is a necessary condition for *fast* convergence.

The existence of mixed derivatives for general $f \in H^1(\Omega)$ is lost by the closure with respect to the $H^1$ norm.

Nevertheless, $H^1(\Omega)$ can also be constructed by means of canonical scalar products. For this purpose, define the Hilbert tensor spaces $H_{1,j}(\Omega)$ for $j = 0, 1, \ldots, d$ as the closure of

$${}_a\bigotimes_{k=1}^{d} V_k \quad \text{with } V_k = \begin{cases} L^2(I_k) & \text{for } k \neq j, \\ H^1(I_j) & \text{for } k = j, \end{cases}$$

with respect to the canonical scalar product

$$\langle f, g\rangle_{(j)} = \int_\Omega \frac{\partial f}{\partial x_j}\,\overline{\frac{\partial g}{\partial x_j}}\,\mathrm{d}x_1 \cdots \mathrm{d}x_d$$

[7] Let $f$ be defined on $[0,2]$ such that $f \in H^1([0,2])$, but $f \notin H^2([0,2])$. Then the bivariate function $F(x,y) := f(x+y)$ belongs to $H^1([0,1]^2)$ but not to $H^1_{\mathrm{mix}}([0,1]^2)$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

for $1 \le j \le d$ and $\langle f, g\rangle_{(0)} = \langle f, g\rangle_{L^2}$, that is, $H_{1,0}(\Omega) = L^2(\Omega)$. Then the intersection yields the Sobolev space

$$H^1(\Omega) = \bigcap_{j=0}^{d} H_{1,j}(\Omega),$$

and the scalar product of $H^1(\Omega)$ is the sum of the canonical scalar product of $H_{1,j}(\Omega)$:

$$\langle f, g\rangle_{H^1} = \sum_{j=0}^{d} \langle f, g\rangle_{(j)}.$$

Note the embeddings

$$L^2(\Omega) \supset H_{1,j}(\Omega) \supset H^1(\Omega) \supset H^1_{\mathrm{mix}}(\Omega).$$

**Remark 3.9.** The fact that $H^1_{\mathrm{mix}}(\Omega) \subsetneqq H^1(\Omega)$ leads to the difficulty that the norm $\|\cdot\|_{H^1}$ of $H^1(\Omega)$ does not satisfy condition (3.4), that is, the algebraic tensor product $\bigotimes_{j=1}^{d}(H^1(I_j))'$ of the dual spaces can be completed in $(H^1_{\mathrm{mix}}(\Omega))'$ but it is not contained in the smaller space $(H^1(\Omega))'$.

## 4. Tensor formats

### *4.1. Formats in general*

Let $S$ be some mathematical set. In general, a representation of $s \in S$ is based on a mapping

$$\rho_S : P_S \to S, \tag{4.1}$$

where the set $P_S$ consists of tuples $p = (p_1, \ldots, p_n)$ of parameters which are assumed to be realizable in computer language implementations, that is, $p_j$ are reals, integers, array of these types, *etc.* The (abstract) mapping $\rho_S$ only defines $p \in P_S$ and cannot be used in implementations.

Since most of the mathematical sets are infinite, surjectivity of $\rho_S$ cannot hold in general. One remedy is the replacement of $S$ by a finite subset $S_0 \subset S$ such that $\rho_S : P_S \to S_0$ becomes surjective. For instance, we may restrict the integers to an interval $S_0 = \mathbb{Z} \cap [-i_{\max}, i_{\max}]$. A consequence is that we have to expect problems when we try to perform the addition $i_{\max} + 1$. Another type of replacement $S_0 \subset S$ is known for the case $S = \mathbb{R}$. Here, the set $S_0$ of machine numbers satisfies a density property: any real number[8] $x \in S = \mathbb{R}$ can be approximated by $x_0 \in S_0$ such that the relative error is bounded by the machine precision.

[8] Here, we ignore the problems of overflow and underflow, which is a difficulty of the same kind as discussed above for integers.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

From now on we assume that the mapping $\rho_S : P_S \to S$ is surjective (possibly because $S$ is replaced by a subset which again is called $S$).

There is no need to require *injectivity* of $\rho_S$. In general, the inverse $\rho_S^{-1}(s)$ of some $s \in S$ is set-valued. Any $p \in \rho_S^{-1}(s)$ may be used equally well to represent $s \in S$. If the elements of $\rho_S^{-1}(s)$ have different data size, we would prefer parameters $p \in \rho_S^{-1}(s)$ with a data size close to the minimum.

In the case of tensors ($S = \mathbf{V}$), the mapping $\rho_{\mathbf{V}}$ fixes a certain *tensor format*. In the case of algebraic tensors, the representation $\rho_{\mathbf{V}}$ is surjective. However, in practice we restrict ourselves to tensors which can be represented by parameters $p$ of acceptable data size. This can be justified if a tensor $\mathbf{v}$, which is hard to represent exactly, is well approximated by $\tilde{\mathbf{v}} = \rho_{\mathbf{V}}(\tilde{p})$ with $\tilde{p}$ of acceptable data size.

Below we introduce the 'full format', the most trivial and most costly one. The next two sections describe two classical tensor formats. The $r$-term format (also called the *canonical format* or *CP format*) is investigated in Section 5. The tensor subspace format (or 'Tucker format') is described in Section 6. More recent formats are the hierarchical format in Section 7 and, as a subset, the 'TT format' (or 'matrix product format') in Section 9.

### *4.2. Full format*

If *all* entries $\mathbf{v}[i_1, \ldots, i_d]$ of $\mathbf{v} \in \bigotimes_{j=1}^{d} \mathbb{R}^{n_j}$ are stored explicitly, $\mathbf{v}$ is said to be represented in *full format*. Obviously, this is only feasible if the tensor space dimension $N := \prod_{j=1}^{d} n_j$ is not too large. Since we exclude the trivial cases $n_j \leq 1$ (see Remark 2.2(b)), the size of $N \geq 2^d$ increases exponentially in $d$. Here, the parameter $p \in P_S$ in (4.1) consists of all entries $\mathbf{v}[i_1, \ldots, i_d]$.

A variation of the full format is the sparse format (like the sparse matrix format), where only the non-zero entries are stored. For instance, this format is used in the sparse grid approach: see Bungartz and Griebel (2004), Garcke (2013), and Section 5.11.

The difficulties arising from the possibly huge data size can be avoided by the *functional representation*, where a tensor $\mathbf{v}$ is given by a function

$$\text{function } v(i_1, \ldots, i_d),$$

which returns the value $\mathbf{v}[i_1, \ldots, i_d]$ for a single multi-index. Such an implementation requires that a description by a subroutine is available. This is particularly true if the tensor is the restriction of a $d$-variate function to a grid. For instance, if $\mathbf{v}[i_1, \ldots, i_d] = f(i_1 h, \ldots, i_d h)$ for $0 \leq i_j \leq n$ and $h = 1/n$, an implementation of the function $f$ yields the desired function $v$.

Usually a functional representation cannot be evaluated for *all* entries. In Section 11 we shall describe an approximation ('multivariate cross approximation') of a tensor based on a functional representation, which will require only a much smaller number of evaluations.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

## 5. $r$-term format

### *5.1. Representation and storage*

We recall that all algebraic[9] tensors are linear combinations of elementary tensors (see property (i) in Proposition 2.1); therefore they can be written in the $r$-term format[10] (also called the *canonical* or *CP* format):[11]

$$\mathbf{v} = \sum_{\nu=1}^{r} v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}. \tag{5.1}$$

We recall the definition (2.4) of $\mathcal{R}_r$. All tensors of the form (5.1) with fixed $r$ yield the set $\mathcal{R}_r$. The representation of $\mathbf{v}$ requires only $rd$ vectors $v_\nu^{(j)}$. Assuming $V_j = \mathbb{K}^{n_j}$ $(1 \le j \le d)$, the storage required by representation (5.1) is

$$\text{storage} \le r \sum_{j=1}^{d} n_j \quad (= rnd \text{ for } n_j = n).$$

Here, it is important that the exponent in $n^d$ has become a factor.

The storage cost is acceptable if $r$ is of moderate size. For general tensors this will not be true. The bound from (2.6) yields $r = O(n^{d-1})$, that is, the storage may become $O(dn^d)$, which is worse than the full format. Fortunately, many tensors of practical interest are of small rank or can be well approximated by low-rank tensors. Examples will be given in Sections 5.6–5.11.

If we use the $r$-term format for practical purposes, we should not aim for the minimal $r = \operatorname{rank}(\mathbf{v})$ because of Remark 2.8(a).

Often, we fix $r$ and try to find a representation or an approximation in $\mathcal{R}_r$. Then we call $r$ the *representation rank*. Note that the representation rank is a property of the format and not of a particular tensor $\mathbf{v} \in \mathcal{R}_r$.

If the tensor describes a $d$-variate grid function with $n_j$ grid points in the $j$th direction, the numbers $n_j$ may be large. The technique in Section 12 will show that the storage size $n_j$ per vector $v_\nu^{(j)}$ can be reduced further (in optimal cases up to $O(\log(n_j))$).

[9] Topological tensors which are not algebraic may have a representation (5.1) with $r = \infty$, but in general this is not ensured.

[10] $r$ is a parameter which may be replaced by other expressions: $3n$-format or 5-format.

[11] We are free to move a factor from one vector in $v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$ to another vector. There are some variants of the formulation (5.1) that try to overcome this non-uniqueness. The vectors $v_\nu^{(j)}$ may be normalized by $\|v_\nu^{(j)}\| = 1$. Then a factor $\alpha_\nu$ must be added: $\mathbf{v} = \sum_{\nu=1}^{r} \alpha_\nu v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$. Alternatively, the normalization may be expressed by the side condition $\|v_\nu^{(1)}\| = \|v_\nu^{(2)}\| = \cdots = \|v_\nu^{(d)}\|$ in (5.1).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

In the case of matrices, the $r$-term format contains Kronecker products:

$$\mathbf{A} = \sum_{\nu=1}^{r} A_\nu^{(1)} \otimes \cdots \otimes A_\nu^{(d)}, \tag{5.2}$$

where $A_\nu^{(j)} \in \mathbb{R}^{n_j \times m_j}$. The corresponding storage requirement depends on the representation of the matrices $A_\nu^{(j)}$. The full representation of $A_\nu^{(j)}$ leads to the bound $r \sum_{j=1}^{d} n_j m_j$. Sparse matrices require only $O(rnd)$ (assuming $n_j = m_j = n$). Moreover, densely populated matrices may allow a hierarchical matrix representation (Hackbusch 2009). The combination of this matrix format with the $r$-term (Kronecker) tensor format is called the *HKT format* in Hackbusch, Khoromskij and Tyrtyshnikov (2005), and requires storage of $O(rdn \log^* n)$.

Under certain conditions, the $r$-term representation (5.1) is unique: see Kruskal (1977), De Lathauwer (2008), Stegeman (2010), Chiantini and Ottaviani (2012), Sørensen, De Lathauwer, Comon, Icart and Deneire (2012) and Landsberg (2012, Chap. 12). This means that up to the ordering of the terms, the elementary tensors $v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$ in (5.1) are unique. The vectors $v_\nu^{(j)}$ are fixed up to a factor (see footnote 11). This property is important if (5.1) is used as a decomposition (see Section 1.4).

### *5.2. Operations with tensors*

As mentioned in the Introduction, the representation of tensors must be such that operations between tensors are also supported. Generally speaking, the cost of an operation should be of a similar size to the data size of the operands.

Below we give a list of operations and show how they perform in the $r$-term format.

#### *5.2.1. Addition*

The addition of two tensors,

$$\mathbf{v} = \sum_{\nu=1}^{r} \bigotimes_{j=1}^{d} v_\nu^{(j)} \quad \text{and} \quad \mathbf{w} = \sum_{\nu=1}^{s} \bigotimes_{j=1}^{d} w_\nu^{(j)}, \tag{5.3}$$

yields

$$\mathbf{x} = \sum_{\nu=1}^{r+s} \bigotimes_{j=1}^{d} x_\nu^{(j)} \quad \text{with } x_\nu^{(j)} := \begin{cases} v_\nu^{(j)} & \text{for } \nu \le r, \\ w_\nu^{(j)} & \text{for } \nu > r. \end{cases}$$

Obviously, this operation requires no arithmetic but increases the representation rank.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*5.2.2. Scalar product*

The canonical scalar product of the tensor $\mathbf{v}, \mathbf{w}$ from (5.3) is performed by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{\nu=1}^{r} \sum_{\mu=1}^{s} \prod_{j=1}^{d} \langle v_\nu^{(j)}, w_\mu^{(j)} \rangle,$$

requiring $rs$ scalar products in $V_j$ and further $rsd$ arithmetic operations. The evaluation of the norm by means of $\langle \mathbf{v}, \mathbf{v} \rangle$ reduces the computational cost by a factor of $1/2$ because of symmetry.

*5.2.3. Matrix–vector multiplication*

Consider $\mathbf{A}$ from (5.2) with $r = r_A$ and $\mathbf{v}$ from (5.1) with $r = r_v$. The matrix-vector multiplication yields

$$\mathbf{A}\mathbf{v} = \sum_{\nu=1}^{r_A} \sum_{\mu=1}^{r_v} \big(A_\nu^{(1)} v_\mu^{(1)}\big) \otimes \cdots \otimes \big(A_\nu^{(d)} v_\mu^{(d)}\big), \tag{5.4}$$

which is an $r_A \cdot r_v$-term representation of the (exact) result. The computation requires $d \cdot r_A \cdot r_v$ matrix–vector multiplication of the form $A_\nu^{(j)} v_\mu^{(j)}$.

*5.2.4. Matrix–matrix multiplication*

If $\mathbf{A}$ and $\mathbf{B}$ are two matrices of the form (5.2), their product is of a similar shape to (5.4):

$$\mathbf{A}\mathbf{B} = \sum_{\nu=1}^{r_A} \sum_{\mu=1}^{r_B} \big(A_\nu^{(1)} B_\mu^{(1)}\big) \otimes \cdots \otimes \big(A_\nu^{(d)} B_\mu^{(d)}\big).$$

*5.2.5. Hadamard product*

The Hadamard product of two vectors[12] $v, w \in \mathbb{K}^I$ is the vector of the entrywise products: $v \odot w = (v_i \cdot w_i)_{i \in I}$. The Hadamard product of $\mathbf{v}, \mathbf{w} \in \bigotimes_{j=1}^{d} \mathbb{K}^{I_j}$ has the entries $(\mathbf{v} \odot \mathbf{w})[i_1, \ldots, i_d] := \mathbf{v}[i_1, \ldots, i_d] \cdot \mathbf{w}[i_1, \ldots, i_d]$. Hence, the Hadamard product of $\mathbf{v} \in \mathcal{R}_r$ and $\mathbf{w} \in \mathcal{R}_s$ yields

$$\mathbf{v} \odot \mathbf{w} = \left( \sum_{\nu=1}^{r} \bigotimes_{j=1}^{d} v_\nu^{(j)} \right) \odot \left( \sum_{\mu=1}^{s} \bigotimes_{j=1}^{d} w_\mu^{(j)} \right) = \sum_{\nu=1}^{r} \sum_{\mu=1}^{s} \bigotimes_{j=1}^{d} \big(v_\nu^{(j)} \odot w_\mu^{(j)}\big), \tag{5.5}$$

which is again an $rs$-term representation. The computational cost amounts to $rs \sum_{j=1}^{d} n_j$.

[12] For functions the Hadamard product is the standard one: $(f \cdot g)(x) = f(x) \cdot g(x)$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*5.2.6. Convolution*

A convolution $v \star w$ between two vectors from $\mathbb{K}^n$ may, for example, be defined by

$$(v \star w)[i] = \sum_{k=1}^{n} v[k]w[i-k] \quad (1 \le i \le n),$$

where the indices are to be understood modulo $n$. Using this definition for all $d$ directions of the tensor, the convolution of tensors is defined by (5.5) with $\odot$ replaced by $\star$. Again the resulting representation rank is $rs$.

### *5.3. Truncation*

All tensor-valued operations mentioned above yield a result with increased representation rank. Without a reduction of the rank, repeated operations would let the rank explode. Therefore, after an operation, one should try to replace the result $\mathbf{v} \in \mathcal{R}_s$ ($s$ large) by an approximation $\tilde{\mathbf{v}} \in \mathcal{R}_r$ with smaller representation rank $r$. We either fix the rank $r$ or choose an accuracy $\varepsilon > 0$ and try to satisfy $\|\mathbf{v} - \tilde{\mathbf{v}}\| \le \varepsilon$ with an $r$ as small as possible. In the first case, we need a mapping $T_{r,s} : \mathcal{R}_s \to \mathcal{R}_r$ called *truncation*. In the second case, the truncation $T_{\varepsilon,s}$ also determines the destination rank $r$.

So far, we have not defined $T_{r,s} : \mathcal{R}_s \to \mathcal{R}_r$ uniquely. The following definition seems to be evident: define $\tilde{\mathbf{v}} = T_{r,s}\mathbf{v}$ by

$$\|\mathbf{v} - \tilde{\mathbf{v}}\| = \inf_{\mathbf{w} \in \mathcal{R}_r} \|\mathbf{v} - \mathbf{w}\| \tag{5.6}$$

for a suitable norm. Unfortunately, this infimum might not be attained, that is, there is no $\tilde{\mathbf{v}}$ satisfying the requirement. The reason is the fact that, in general, $\mathcal{R}_r$ is not closed. In the next section, we study these problems in more detail.

### *5.4. Non-closedness of $\mathcal{R}_r$, border rank, instability*

There is a simple counterexample which shows that $\mathcal{R}_2$ is not closed. The assumptions are $d \ge 3$ (avoiding the matrix case $d = 2$) and $\dim V_j \ge 2$ (see Remark 2.2(b)). Then there are linearly independent vectors $v^{(j)}, w^{(j)} \in V_j$. It can be proved that the tensor

$$\mathbf{v} := v^{(1)} \otimes v^{(2)} \otimes w^{(3)} + v^{(1)} \otimes w^{(2)} \otimes v^{(3)} + w^{(1)} \otimes v^{(2)} \otimes v^{(3)} \tag{5.7a}$$

has tensor rank 3: see Hackbusch (2012, Lemma 3.41) and Buczyński and Landsberg (2014). Next, we define the sequence of tensors

$$\begin{aligned} \mathbf{v}_n := (w^{(1)} + nv^{(1)}) &\otimes (v^{(2)} + \tfrac{1}{n}w^{(2)}) \otimes v^{(3)} \\ + \quad v^{(1)} &\otimes \quad v^{(2)} \quad \otimes (w^{(3)} - nv^{(3)}), \end{aligned} \tag{5.7b}$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

which have $\operatorname{rank}(\mathbf{v}_n) = 2$. The identity $\mathbf{v} - \mathbf{v}_n = -\frac{1}{n} w^{(1)} \otimes w^{(2)} \otimes v^{(3)}$ is easy to see. Hence, independently of the choice of norm, we obtain

$$\lim_{n\to\infty} \mathbf{v}_n = \mathbf{v}.$$

This shows $3 = \operatorname{rank}(\mathbf{v}) = \operatorname{rank}(\lim \mathbf{v}_n) > \operatorname{rank}(\mathbf{v}_n) = 2$; that is, the closure $\overline{\mathcal{R}_2}$ contains an element $\mathbf{v} \in \mathcal{R}_3 \backslash \mathcal{R}_2$ proving the non-closedness of $\mathcal{R}_2$.

Tensor $\mathbf{v}$ from (5.7a) is also the counterexample for the minimization of $\|\mathbf{v} - \mathbf{w}\|$ over $\mathbf{w} \in \mathcal{R}_2$. Because of $\mathbf{v}_n \in \mathcal{R}_2$ and $\mathbf{v}_n \to \mathbf{v}$, the infimum is $\inf_{\mathbf{w}\in\mathcal{R}_2} \|\mathbf{v} - \mathbf{w}\| \geq \lim_{n\to\infty} \|\mathbf{v} - \mathbf{v}_n\| = 0$, but there is no minimizer $\tilde{\mathbf{v}} \in \mathcal{R}_2$ with $\|\mathbf{v} - \tilde{\mathbf{v}}\| = \inf_{\mathbf{w}\in\mathcal{R}_2} \|\mathbf{v} - \mathbf{w}\| = 0$ as $\operatorname{rank}(\mathbf{v}) = 3$. De Silva and Lim (2008) prove that tensors $\mathbf{v}$ without a minimizer of $\inf_{\mathbf{u}\in\mathcal{R}_r} \|\mathbf{v} - \mathbf{u}\|$ are not of measure zero, that is, there is a positive expectation that random tensors $\mathbf{v}$ are of this type.

The observed properties lead to a modification of the tensor rank, where $\mathcal{R}_r$ is replaced by its closure (Bini, Lotti and Romani 1980).

**Definition 5.1.** The *border rank* is defined by

$$\underline{\operatorname{rank}}(\mathbf{v}) := \min\{r : \mathbf{v} \in \overline{\mathcal{R}_r}\} \in \mathbb{N}_0. \tag{5.8}$$

In principle, the non-existence of a minimizer is not a problem from the numerical point of view, since we only determine an approximation to the exact minimizer, in any case. However, in this case the non-existence of a minimizer is inextricably linked to numerical instability.

We introduce a condition number $\varkappa\big((v_i^{(j)})_{1\le i\le r}^{1\le j\le d}\big)$ for the representation of $\mathbf{v} = \sum_{i=1}^r \bigotimes_{j=1}^d v_i^{(j)} \neq 0$ by means of

$$\varkappa\Big((v_i^{(j)})_{1\le i\le r}^{1\le j\le d}\Big) := \left(\sum_{i=1}^r \left\| \bigotimes_{j=1}^d v_i^{(j)} \right\|\right) \Big/ \left\| \sum_{i=1}^r \bigotimes_{j=1}^d v_i^{(j)} \right\|. \tag{5.9}$$

A large condition number $\varkappa = \varkappa\big((v_i^{(j)})\big)$ leads to numerical cancellation. For instance, an entry $\mathbf{v}[\mathbf{i}]$ – say of size $O(1)$ – is a sum of terms of size $O(\varkappa)$. As a result, $\log_2(\varkappa)$ binary digits are lost.

The infimum over all $r$-term representations of $\mathbf{v} \in \mathcal{R}_r$ yields the number

$$\varkappa(\mathbf{v}, r) := \inf\left\{ \varkappa\Big((v_i^{(j)})_{1\le i\le r}^{1\le j\le d}\Big) : \mathbf{v} = \sum_{i=1}^r \bigotimes_{j=1}^d v_i^{(j)} \right\}.$$

A sequence $\mathbf{v}_n \in \mathcal{R}_r$ $(n \in \mathbb{N})$ is called *stable* in $\mathcal{R}_r$ if

$$\varkappa((\mathbf{v}_n)_{n\in\mathbb{N}}, r) := \sup_{n\in\mathbb{N}} \varkappa(\mathbf{v}_n, r) < \infty;$$

otherwise, the sequence is unstable.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Proposition 5.2.** Suppose $\dim(V_j) < \infty$ and $\mathbf{v} \in \bigotimes_{j=1}^{d} V_j$. A stable sequence $\mathbf{v}_n \in \mathcal{R}_r$ with $\lim \mathbf{v}_n = \mathbf{v}$ exists if and only if $\mathbf{v} \in \mathcal{R}_r$.

*Proof.* One direction is trivial: if $\mathbf{v} \in \mathcal{R}_r$, the constant sequence with $\mathbf{v}_n = \mathbf{v}$ is stable.

For the other direction assume that the sequence $\mathbf{v}_n \to \mathbf{v}$ with $\mathbf{v}_n \in \mathcal{R}_r$ is stable. Set $C := 2\varkappa((\mathbf{v}_n), r)$. After choosing a subsequence, we have $\mathbf{v}_n \to \mathbf{v}$ with representations $\mathbf{v}_n = \sum_{i=1}^{r} \bigotimes_{j=1}^{d} v_{n,i}^{(j)}$ such that

$$\sum_{i=1}^{r} \left\| \bigotimes_{j=1}^{d} v_{n,i}^{(j)} \right\| \le C\|\mathbf{v}\|.$$

Scale $v_{n,i}^{(j)}$ equally, that is, $\|v_{n,i}^{(j)}\|$ is independent of $i$. Then all $v_{n,i}^{(j)} \in V_j$, $n \in \mathbb{N}$, are uniformly bounded. Again choosing a subsequence, limits $\hat{v}_i^{(j)} := \lim_{n\to\infty} v_{n,i}^{(j)}$ exist and

$$\mathbf{v} = \lim \mathbf{v}_n = \lim \sum_{i=1}^{r} \bigotimes_{j=1}^{d} v_{n,i}^{(j)} = \sum_{i=1}^{r} \bigotimes_{j=1}^{d} \lim v_{n,i}^{(j)} = \sum_{i=1}^{r} \bigotimes_{j=1}^{d} \hat{v}_i^{(j)} \in \mathcal{R}_r$$

proves the assertion. □

Negation of the statement of Proposition 5.2 yields the next statement.

**Conclusion 5.3.** Any sequence $\mathbf{v}_n \in \mathcal{R}_r$ converging to $\mathbf{v} \notin \mathcal{R}_r$ must be unstable.

In the example (5.7a,b), the condition number $\varkappa(\mathbf{v}_n, r)$ is of size $O(n)$. Hence, $\mathbf{v}_n$ has an approximation error $O(1/n)$ and is polluted by a cancellation error $O(n \cdot \mathrm{eps})$, where eps is the machine precision. Obviously, the best result is obtained for $n \approx \sqrt{\mathrm{eps}}$. A similar error is known from the numerical differentiation. In fact, there is a direct connection. Consider the tensor-valued function

$$\mathbf{v}(t) := \left(v^{(1)} + tw^{(1)}\right) \otimes \left(v^{(2)} + tw^{(2)}\right) \otimes \left(v^{(3)} + tw^{(3)}\right).$$

The derivative $\mathbf{v}'(0)$ at $t = 0$ is equal to $\mathbf{v}$ from (5.7a). Obviously, $\mathbf{v}'(0)$ can be approximated by the divided difference $n[\mathbf{v}(1/n) - \mathbf{v}(0)]$. Since $\mathbf{v}(t)$ is an elementary tensor for all $t$, the difference quotient belongs to $\mathcal{R}_2$.

### *5.5. Truncation (continued)*

Since the minimization problem (5.6) can be ill-posed, one needs a stabilization. Either the result $T_{r,s}\mathbf{v} = \tilde{\mathbf{v}} = \sum_{i=1}^{r} \tilde{\mathbf{v}}_i$ ($\tilde{\mathbf{v}}_i$ being an elementary tensor) has to satisfy a side condition $\sum_{i=1}^{r} \|\tilde{\mathbf{v}}_i\| \le C\|\mathbf{v}\|$ or a penalty term is added

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

to the cost function: find $T_{r,s}\mathbf{v} = \tilde{\mathbf{v}} = \sum_{i=1}^{r} \tilde{\mathbf{v}}_i$ as minimizer of

$$\|\mathbf{v} - \tilde{\mathbf{v}}\|^2 + \lambda \sum_{i=1}^{r} \|\tilde{\mathbf{v}}_i\|^2 \quad (\lambda > 0). \tag{5.10}$$

A method based on these modifications is described by Espig and Hackbusch (2012). The computational cost of the method is

$$O\left( r(r+s)d^2 + dr^3 + r(r+s+d) \sum\nolimits_{j=1}^{d} \operatorname{rank}_j(\mathbf{v}) \right).$$

Another well-known method is the alternating least-squares method, where one minimizes (5.10) first with respect to the vectors $u_i^{(1)}$ in $\tilde{\mathbf{v}}_i = \bigotimes_{j=1}^{d} u_i^{(j)}$, then with respect to $u_i^{(2)}$, *etc.* However, all iterative methods may lead to a local minimum, which is not a global one. In particular, the result may depend on the starting values.

### *5.6. Analytic approximations*

There are examples of practically important tensors for which approximations can be described by analytic tools. In these cases, error estimates can also be provided. The first example is based on approximations by means of exponential sums. For instance, we obtain approximations for the function $1/\|\mathbf{x}\|_2$ ($\mathbf{x} \in \mathbb{R}^d$, $\|\cdot\|_2$ Euclidean norm: see Section 5.8). In the case of $d = 3$, this is the important Newton or Coulomb potential. The same tool yields exponentially accurate approximations of the inverse $\mathbf{A}^{-1}$ for operators (matrices) $\mathbf{A}$ arising from separable differential equations (see Section 5.10).

The sparse grid technique is known as a method which can deal with dimensions larger than $d = 3$. There is an analysis providing error bounds for this approach. In Section 5.11 we mention how to interpret the sparse grid results as tensors in the $r$-term format.

### *5.7. Exponential sums*

#### *5.7.1. Basic statements*

Exponential sums are expressions of the form

$$E_r(t) = \sum_{\nu=1}^{r} \omega_\nu \exp(-\alpha_\nu t) \quad (t \in \mathbb{R}), \tag{5.11a}$$

with $2r$ (real or complex) parameters $\omega_\nu$ and $\alpha_\nu$. Exponential sums are a tool to approximate certain univariate functions; for details about their computation see Braess and Hackbusch (2009).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Assume that a univariate function $f$ in an interval $I \subset \mathbb{R}$ is approximated by some exponential sum $E_r$ with respect to the supremum norm in $I$:

$$\|f - E_r\|_{I,\infty} \le \varepsilon. \tag{5.11b}$$

Then the multivariate function

$$F(\mathbf{x}) = F(x_1, \dots, x_d) := f\left(\sum_{j=1}^{d} \phi_j(x_j)\right), \tag{5.11c}$$

obtained by the substitution $t = \sum_{j=1}^{d} \phi_j(x_j)$, is approximated equally well by $F_r(\mathbf{x}) := E_r\big(\sum_{j=1}^{d} \phi_j(x_j)\big)$,

$$\|F - F_r\|_{\mathbf{I},\infty} \le \varepsilon \quad \text{for } \mathbf{I} := \bigtimes_{i=1}^{d} I_j, \tag{5.11d}$$

provided that $t$ varies in $I$:

$$\left\{\sum_{j=1}^{d} \phi_j(x_j) : x_j \in I_j\right\} \subset I \quad \text{with } I \text{ from (5.11b)}. \tag{5.11e}$$

For instance, condition (5.11e) holds for $\phi_j(x_j) = x_j$ and $I_j = I = [0, \infty)$.

By the property of the exponential function, we have

$$\begin{aligned} F_r(\mathbf{x}) := E_r\left(\sum_{j=1}^{d} \phi_j(x_j)\right) &= \sum_{\nu=1}^{r} \omega_\nu \exp\left(-\alpha_\nu \sum_{j=1}^{d} \phi_j(x_j)\right) \\ &= \sum_{\nu=1}^{r} \omega_\nu \prod_{j=1}^{d} \exp(-\alpha_\nu \phi_j(x_j)). \end{aligned} \tag{5.11f}$$

Expressing the multivariate function $E_r$ as a tensor product of univariate functions, we arrive at

$$F_r = \sum_{\nu=1}^{r} \omega_\nu \bigotimes_{j=1}^{d} E_\nu^{(j)} \in \mathcal{R}_r \quad \text{with } E_\nu^{(j)}(x_j) := \exp(-\alpha_\nu \phi_j(x_j)). \tag{5.11g}$$

Equation (5.11g) is an $r$-term representation of the tensor

$$F_r \in C(\mathbf{I}) = {}_\infty\bigotimes_{j=1}^{d} C(I_j).$$

The left suffix $\infty$ indicates completion with respect to the supremum norm in $\mathbf{I} \subset \mathbb{R}^d$. The following simple but important conclusion shows that it is enough to analyse the univariate function $f$ and its approximation by $E_r$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Conclusion 5.4.** The multivariate function $F_r(\mathbf{x})$ has representation rank $r$ independently of the dimension $d$. Also, the approximation error (5.11d) is independent of the dimension $d$, provided that (5.11b) and (5.11e) are valid.

The approximation of $f(\xi)$ by a sum of *Gaussians*,

$$G_r(\xi) = \sum_{\nu=1}^{r} \omega_\nu \, \mathrm{e}^{-\alpha_\nu \xi^2},$$

is equivalent to the approximation of $f(\sqrt{\xi})$ by

$$E_r(t) := \sum_{\nu=1}^{r} \omega_\nu \, \mathrm{e}^{-\alpha_\nu t}.$$

Conversely, the substitution $t = \|\mathbf{x}\|^2$ in (5.11c) yields an approximation by Gaussians:

$$F_r(\mathbf{x}) := E_r\left(\sqrt{\sum_{j=1}^{d} x_j^2}\right) = \sum_{\nu=1}^{r} \omega_\nu \exp\left(-\alpha_\nu \sum_{j=1}^{d} x_j^2\right) = \sum_{\nu=1}^{r} \omega_\nu \prod_{j=1}^{d} \mathrm{e}^{-\alpha_\nu x_j^2},$$

$$i.e., \quad F_r = \sum_{\nu=1}^{r} \omega_\nu \bigotimes_{j=1}^{d} G_\nu^{(j)} \quad \text{with } G_\nu^{(j)}(x_j) := \exp(-\alpha_\nu x_j^2). \tag{5.12}$$

Inequality (5.11b) implies

$$\|F - F_r\|_{D,\infty} \le \varepsilon, \quad \text{where } D := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \in I\}.$$

*5.7.2. Quadrature-based exponential sum approximations*

Approximations by exponential sums may be based on quadrature rules.[13] Assume that a function $f$ with domain $I \subset \mathbb{R}$ is defined by the Laplace transform:

$$f(x) = \int_0^\infty \mathrm{e}^{-tx} g(t)\, \mathrm{d}t \quad \text{for } x \in I.$$

Any quadrature method $Q(F) := \sum_{\nu=1}^{r} \omega'_\nu F(t_\nu)$ for a suitable integrand $F$ defined on $[0, \infty)$ yields an exponential sum of the form (5.11a):

$$f(x) \approx Q(\mathrm{e}^{-\bullet x} g) := \sum_{\nu=1}^{r} \underbrace{\omega'_\nu g(t_\nu)}_{=:\,\omega_\nu} \mathrm{e}^{-t_\nu x} \in \mathcal{R}_r.$$

The dot in $\mathrm{e}^{-\bullet x}$ indicates the variable $t$ involved in the integration. Note that the quadrature error $f(x) - Q(\mathrm{e}^{-\bullet x} g)$ is to be controlled for all parameter values $x \in I$.

[13] Quadrature-based approximation is very common in computational quantum chemistry. For a discussion from the mathematical side see Beylkin and Monzón (2010).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

A possible choice of $Q$ is the *sinc quadrature*. For this purpose one chooses a suitable substitution $t = \varphi(\tau)$ with $\varphi : \mathbb{R} \to [0, \infty)$ to obtain an integral over $\mathbb{R}$:

$$f(x) = \int_{-\infty}^{\infty} \mathrm{e}^{-\varphi(\tau)x}\, g(\varphi(\tau))\, \varphi'(\tau)\, \mathrm{d}\tau.$$

The sinc quadrature can be applied to analytic functions defined on $\mathbb{R}$:

$$\int_{-\infty}^{\infty} F(x)\, \mathrm{d}x \approx T(F,h) := h \sum_{k=-\infty}^{\infty} F(kh) \approx T_N(F,h) := h \sum_{k=-N}^{N} F(kh).$$

$T(F,h)$ can be interpreted as the infinite trapezoidal rule with step size $h$, while $T_N(F,h)$ is a truncated finite sum.

The error analysis of $T(F,h)$ depends on the behaviour of the holomorphic function $F(z)$ in the complex strip $\mathfrak{D}_\delta := \{z \in \mathbb{C} : |\mathrm{Im}\, z| < \delta\}$ and the norm

$$\|f\|_{\mathfrak{D}_\delta} = \int_{\partial \mathfrak{D}_\delta} |f(z)||\mathrm{d}z|.$$

A typical error bound of $T_N(F,h)$ is of the form $C_1 \exp(-\sqrt{2\pi\delta\alpha N})$ with $C_1 = C_1(\|F\|_{\mathfrak{D}_\delta})$ involving the width $\delta$ of $\mathfrak{D}_\delta$, while $\alpha$ describes the decay of $F$: $|F(x)| \le O(\exp(-\alpha|x|))$. For a precise analysis see Stenger (1993) or Hackbusch (2009, § D.4). Sinc quadrature applied to the integrand $F(t) = F(t;x) := \mathrm{e}^{-\varphi(t)x}\, g(\varphi(t))\, \varphi'(t)$ from above yields

$$T_N(F,h) := h \sum_{k=-N}^{N} \mathrm{e}^{-\varphi(kh)x}\, g(\varphi(kh))\, \varphi'(kh).$$

The right-hand side is an exponential sum (5.11a) with $r := 2N+1$ and

$$\omega_\nu := h\, g\big(\varphi((\nu-1-N)h)\big)\, \varphi'\big((\nu-1-N)h\big), \quad \alpha_\nu := \varphi\big((\nu-1-N)h\big).$$

Since the integrand $F(\bullet; x)$ depends on the parameter $x \in I$, the error analysis must be performed *uniformly* in $x \in I$ to prove an estimate (5.11b): $\|f - E_r\|_{I,\infty} \le \varepsilon$.

Even when the obtainable error bounds possess almost optimal asymptotic behaviour, they are inferior to the best approximations discussed next.

### 5.8. *Approximation of* $1/x$ *and* $1/\sqrt{x}$

Negative powers $x^{-\lambda}$ belong to the class of those functions which can be well approximated by exponential sums in $(0, \infty)$. Because of their importance, we shall consider the particular functions $1/x$ and $1/\sqrt{x}$. For the general theory of approximation by exponentials we refer to Braess (1986). The first

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Table 5.1.

| $r$ | $\|1/x - E_{r,[1,10]}\|_{[1,10],\infty}$ |
|---|---|
| 1 | $8.556 \times 10^{-2}$ |
| 2 | $8.752 \times 10^{-3}$ |
| 3 | $7.145 \times 10^{-4}$ |
| 4 | $5.577 \times 10^{-5}$ |
| 5 | $4.243 \times 10^{-6}$ |
| 6 | $3.173 \times 10^{-7}$ |
| 7 | $2.344 \times 10^{-8}$ |

statement concerns the existence of a best approximation and its stability[14] expressed by positivity of its terms.

**Theorem 5.5 (Braess 1986, p. 194).** Given the function $f(x) = x^{-\lambda}$ with $\lambda > 0$ in an interval $I = [a, b]$ (including $b = \infty$) with $a > 0$, and $r \in \mathbb{N}$, there is a unique best approximation $E_{r,I}(x) = \sum_{\nu=1}^{r} \omega_{\nu,I} \exp(-\alpha_{\nu,I} x)$ with

$$\varepsilon(f, I, r) := \|f - E_{r,I}\|_{I,\infty} = \inf\left\{ \left\| f - \sum_{\nu=1}^{r} b_\nu \, \mathrm{e}^{-\beta_\nu x} \right\|_{I,\infty} : b_\nu, \beta_\nu \in \mathbb{R} \right\}.$$

Moreover, this $E_{r,I}$ has positive coefficients: $\omega_\nu, \alpha_\nu > 0$ for $1 \le \nu \le r$.

In the case of $f(x) = 1/x$, substitution $x = at$ $(1 \le t \le b/a)$ shows that the best approximation for $I = [a, b]$ can be derived from the best approximation in $[1, b/a]$ via the transform

$$\omega_{\nu,[a,b]} := \frac{\omega_{\nu,[1,b/a]}}{a}, \quad \alpha_{\nu,[a,b]} := \frac{\alpha_{\nu,[1,b/a]}}{a}, \quad \varepsilon(f, [a, b], r) = \frac{\varepsilon(f, [1, b/a], r)}{a}. \tag{5.13a}$$

In the case of $f(x) = 1/\sqrt{x}$, the relations are

$$\omega_{\nu,[a,b]} = \frac{\omega_{\nu,[1,b/a]}}{\sqrt{a}}, \quad \alpha_{\nu,[a,b]} := \frac{\alpha_{\nu,[1,b/a]}}{a}, \quad \varepsilon(f, [a, b], r) = \frac{\varepsilon(f, [1, b/a], r)}{\sqrt{a}}. \tag{5.13b}$$

Therefore, it suffices to study the best approximation on standardized intervals $[1, R]$ for $R \in (1, \infty)$. The coefficients $\{\omega_\nu, \alpha_\nu : 1 \le \nu \le r\}$ for various values of $R$ and $r$ are given in Hackbusch (2005).

For examples of convergence, we first consider a fixed interval $[1, R] = [1, 10]$. The error $\|1/x - E_{r,[1,10]}\|_{[1,10],\infty}$ is shown in Table 5.1. We observe an exponential decay $O(\exp(-cr))$ with $c > 0$.

[14] Stability refers to the condition number $\varkappa((v_i^{(j)})_{1\le i\le r}^{1\le j\le d})$ from (5.9). With respect to the maximum norm we have $\varkappa((v_i^{(j)})_{1\le i\le r}^{1\le j\le d}) = 1$, since all vectors $v_i^{(j)}$ are non-negative.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Table 5.2.

| $r$ | $R_r^*$ | $\varepsilon(1/x,[1,\infty),r)$ |
|---|---|---|
| 9 | 28387 | $2.611\times10^{-5}$ |
| 16 | $2.027\times10^{+6}$ | $3.659\times10^{-7}$ |
| 25 | $1.513\times10^{+8}$ | $4.898\times10^{-9}$ |
| 36 | $1.162\times10^{+10}$ | $6.382\times10^{-11}$ |
| 49 | $9.074\times10^{+11}$ | $8.172\times10^{-13}$ |

Table 5.3.

| $r$ | $R_r^*$ | $\varepsilon(1/\sqrt{x},[1,\infty),r)$ |
|---|---|---|
| 9 | $7.994\times10^{+6}$ | $3.072\times10^{-4}$ |
| 16 | $4.129\times10^{+9}$ | $1.352\times10^{-5}$ |
| 25 | $2.17\times10^{+12}$ | $5.898\times10^{-7}$ |
| 36 | $1.15\times10^{+15}$ | $2.564\times10^{-8}$ |
| 49 | $6.10\times10^{+17}$ | $1.116\times10^{-9}$ |

If $R$ varies from 1 to $\infty$, there is a certain finite value $R^* = R_r^*$ depending on $r$, such that $\varepsilon(f,[1,R],r)$ as a function of $R$ strictly increases in $[1,R^*]$, whereas the approximant $E_{r,[1,R]}$ as well as the error $\varepsilon(f,[1,R],r)$ is constant for $R$ in $[R^*,\infty)$. This implies that the approximation $E_{r,[1,R^*]}$ is already the best approximation in the semi-infinite interval $[1,\infty)$. Table 5.2 shows $R_r^*$ and $\varepsilon(1/x,[1,R_r^*],r) = \varepsilon(1/x,[1,\infty),r)$. We verify that $\varepsilon(\frac{1}{x},[1,\infty),r)$ behaves like the function $25\exp(-\pi\sqrt{2r})$. The function $f(x) = 1/\sqrt{x}$ has similar asymptotics, as shown in Table 5.3. Here, $\varepsilon(1/\sqrt{x},[1,\infty),r)$ behaves like $4\exp(-\pi\sqrt{r})$. The observed asymptotic behaviour mentioned above is better than the upper bound in the next theorem.

**Theorem 5.6.** Let $f(x) = x^{-\lambda}$ with $\lambda > 0$. The asymptotic behaviour of the error $\varepsilon(f,I,r)$ is

$$\varepsilon(f,I,r) \le \begin{cases} C\exp(-cr) & \text{if } I=[a,b]\subset(0,\infty),\\ C\exp(-c\sqrt{r}) & \text{if } I=[a,\infty),\ a>0,\end{cases}$$

where the constants $C,c>0$ depend on $I$. For instance, for $\lambda = 1/2$ and $a = 1$, the upper bounds are

$$\varepsilon(1/\sqrt{x},[1,R],r) \le 8\sqrt{2}\,\exp\bigl(-\pi^2 r/\sqrt{\log(8R)}\bigr),$$
$$\varepsilon(1/\sqrt{x},[1,\infty),r) \le 8\sqrt{2}\,\exp\bigl(-\pi\sqrt{r/2}\bigr).$$

For general $a > 0$ use (5.13a,b).

*Proof.* For details about the constants see Braess and Hackbusch (2005, 2009). The latter estimates can be found in Braess and Hackbusch (2009, equations (33) and (34)). □

*5.8.1. Multivariate functions derived from $1/\sqrt{x}$*

The function

$$\mathbf{P}(\mathbf{x}) := \frac{1}{\|\mathbf{x}\|} = \frac{1}{\|\sum_{j=1}^3 x_j^2\|} \quad \text{for } \mathbf{x}\in\mathbb{R}^3$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

is called the *Newton potential* in the context of gravity and the *Coulomb potential* in the context of an electrical field. Mathematically, $4\pi\mathbf{P}$ is the singularity function of the Laplace operator $\Delta = \sum_{j=1}^{d} \partial^2/\partial x_j^2$ for $d = 3$. It usually appears in a convolution integral $\mathbf{P} \star \mathbf{f}$. If $\mathbf{f}$ is the mass (charge) density,

$$4\pi \int_{\mathbb{R}^3} \frac{\mathbf{f}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}\,\mathrm{d}\mathbf{y} = 4\pi(\mathbf{P}\star\mathbf{f})(\mathbf{x})$$

describes the gravitational (electrical) field induced by $\mathbf{f}$.

Obviously, it is impossible to approximate $\mathbf{P}$ uniformly on the whole $\mathbb{R}^3$ by exponential sums. Instead, we choose some $\eta > 0$ which will be fixed in Lemma 5.7. Take an optimal approximation $E_r$ of $1/\sqrt{t}$ on $I := [\eta^2, \infty)$. Following the strategy from (5.12), we substitute $t = \|\mathbf{x}\|^2 = \sum_{j=1}^{3} x_j^2$ and obtain

$$E_{r,I}(\|\mathbf{x}\|^2) = \sum_{\nu=1}^{r} \omega_{\nu,I} \prod_{j=1}^{3} \exp(-\alpha_{\nu,I}\, x_j^2),$$

that is, $E_{r,[\eta^2,\infty)}(\|\cdot\|^2) = \sum_{\nu=1}^{r} \omega_{\nu,I} \bigotimes_{j=1}^{3} E_\nu^{(j)} \in \mathcal{R}_r$ with $E_\nu^{(j)}(\xi) = \mathrm{e}^{-\alpha_{\nu,I}\,\xi^2}$. The uniform estimate

$$\left|\mathbf{P}(\mathbf{x}) - E_{r,[\eta^2,\infty)}(\|\mathbf{x}\|^2)\right| \le \varepsilon\left(\frac{1}{\sqrt{\cdot}}, [\eta^2,\infty), r\right) = \frac{\varepsilon}{\eta},$$
$$\text{where } \eta \le \|\mathbf{x}\| < \infty \text{ and } \varepsilon := \varepsilon\left(\frac{1}{\sqrt{\cdot}}, [1,\infty), r\right),$$

excludes the neighbourhood $U_\eta := \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| \le \eta\}$ of the singularity. In $U_\eta$ we use

$$\left|\mathbf{P}(\mathbf{x}) - E_{r,[\eta^2,\infty)}(\|\mathbf{x}\|^2)\right| \le \mathbf{P}(\mathbf{x}) \quad \text{for } \mathbf{x} \in U_\eta \quad \text{and} \quad \int_{U_\eta} \mathbf{P}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 2\pi\eta^2.$$

**Lemma 5.7.** Assume $\|\mathbf{f}\|_{L^1(\mathbb{R}^3)} \le C_1$ and $\|\mathbf{f}\|_{L^\infty(\mathbb{R}^3)} \le C_\infty$. Then

$$\left|\int_{\mathbb{R}^3} \frac{\mathbf{f}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}\,\mathrm{d}\mathbf{y} - \int_{\mathbb{R}^3} E_{r,[\eta^2,\infty)}(\|\mathbf{x}-\mathbf{y}\|^2)\mathbf{f}(\mathbf{y})\,\mathrm{d}\mathbf{y}\right| \le 2\pi\eta^2 C_\infty + \frac{\varepsilon}{\eta} C_1$$

holds with

$$\varepsilon := \varepsilon\left(\frac{1}{\sqrt{\cdot}}, [1,\infty), r\right)$$

for all $\mathbf{x} \in \mathbb{R}^3$. The error bound is minimized for $\eta = \sqrt[3]{(C_1\varepsilon)/(4\pi C_\infty)}$:

$$\left\|\int_{\mathbb{R}^3} \frac{\mathbf{f}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}\mathrm{d}\mathbf{y} - \int_{\mathbb{R}^3} E_{r,[\eta^2,\infty)}(\|\mathbf{x}-\mathbf{y}\|^2)\mathbf{f}(\mathbf{y})\mathrm{d}\mathbf{y}\right\|_{\mathbb{R}^3,\infty} \le \underbrace{\tfrac{3}{2}2^{\frac{2}{3}}\sqrt[3]{\pi}}_{=3.4873} \sqrt[3]{C_1^2 C_\infty}\,\varepsilon^{\frac{2}{3}}.$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Inserting the asymptotic behaviour $\varepsilon = 8\sqrt{2}\exp(-\pi\sqrt{r/2})$ from Theorem 5.6, we obtain a bound of the same form $C\exp(-c\sqrt{r})$ with $c = \sqrt{2}\pi/3$. The observed behaviour is better:

$$O\left(\exp\left(-\frac{2\pi}{3}\sqrt{r}\right)\right).$$

We conclude from Lemma 5.7 that the convolution $\mathbf{P} \star \mathbf{f}$ may be replaced by $E_{r,[\eta^2,\infty)}(\|\cdot\|^2) \star \mathbf{f}$, while the accuracy is still exponentially improving.

In the following, we assume for simplicity that $\mathbf{f}$ is an elementary tensor:

$$\mathbf{f}(\mathbf{y}) = f_1(y_1) \cdot f_2(y_2) \cdot f_3(y_3).$$

The convolution with $E_r(\|\mathbf{x}-\mathbf{y}\|^2)$ can be reduced to three one-dimensional convolutions:

$$\begin{aligned}\int_{\mathbb{R}^3} \frac{\mathbf{f}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}\,\mathrm{d}\mathbf{y} &\approx \int_{\mathbb{R}^3} E_{r,I}(\|\mathbf{x}-\mathbf{y}\|^2)\mathbf{f}(\mathbf{y})\,\mathrm{d}\mathbf{y} \\ &= \sum_{\nu=1}^{r} \omega_{\nu,I} \prod_{j=1}^{3} \int_{\mathbb{R}} \exp(-\alpha_{\nu,I}\,(x_j - y_j)^2) f_j(y_j)\,\mathrm{d}y_j.\end{aligned}$$

For numerical examples related to integral operators involving the Newton potential, see Hackbusch and Khoromskij (2007).

### *5.9. Application to operators and matrices*

Under suitable conditions, univariate functions can be extended to matrix–valued arguments. For instance, functions defined on the spectrum $\sigma(M)$ of a diagonalizable matrix $M$ can be extended to $f(M)$ (Hackbusch 2009, § 13). The next proposition studies the conditioning of $f(M)$ with respect to perturbations of $f$.

**Proposition 5.8.** Let $M = UDU^{\mathsf{H}}$ ($U$ unitary, $D$ diagonal) and assume that $f$ and $\tilde{f}$ are defined on the spectrum $\sigma(M)$. Then the approximation error with respect to the spectral norm $\|\cdot\|_2$ is bounded by

$$\|f(M) - \tilde{f}(M)\|_2 \le \|f - \tilde{f}\|_{\sigma(M),\infty}. \tag{5.14}$$

The estimate extends to self-adjoint operators. For diagonalizable matrices $M = TDT^{-1}$, the right-hand side becomes $\|T\|_2\|T^{-1}\|_2\|f - \tilde{f}\|_{\sigma(M),\infty}$.

*Proof.* Since $f(M) - \tilde{f}(M) = Uf(D)U^{\mathsf{H}} - U\tilde{f}(D)U^{\mathsf{H}} = U[f(D) - \tilde{f}(D)]U^{\mathsf{H}}$ and unitary transformations do not change the spectral norm,

$$\begin{aligned}\|f(M) - \tilde{f}(M)\|_2 &= \|f(D) - \tilde{f}(D)\|_2 = \max\{|f(\lambda) - \tilde{f}(\lambda)| : \lambda \in \sigma(M)\} \\ &= \|f - \tilde{f}\|_{\sigma(M),\infty}\end{aligned}$$

follows. □

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The supremum norm on the right-hand side in (5.14) cannot be relaxed to an $L^p$ norm with $p < \infty$. This fact is what makes the construction of best approximations with respect to the supremum norm so important.

Under stronger conditions on $f$ and $\tilde{f}$, general operators $M \in \mathcal{L}(V, V)$ can be admitted (Hackbusch 2009, Satz 13.2.4). The next proposition applies to general matrices (not necessarily diagonalizable).

**Proposition 5.9.** Let $f$ and $\tilde{f}$ be holomorphic in a complex domain $\Omega$ containing $\sigma(M)$ for some operator $M \in \mathcal{L}(V, V)$. Then

$$\|f(M) - \tilde{f}(M)\|_2 \leq \frac{1}{2\pi} \oint_{\partial\Omega} |f(\zeta) - \tilde{f}(\zeta)| \, \|(\zeta I - M)^{-1}\|_2 \, \mathrm{d}\zeta.$$

*Proof.* Use Cauchy's formula:

$$f(M) := \frac{1}{2\pi \mathrm{i}} \int_{\partial D} (\zeta I - M)^{-1} f(\zeta) \, \mathrm{d}\zeta. \qquad \square$$

*5.10. Application to the inverse of a matrix*

The inverse of $M$ can be considered as the application of the function $f(x) = 1/x$ to $M$, that is, $f(M) = M^{-1}$. Assume that $M$ is Hermitian (self-adjoint) and has a positive spectrum $\sigma(M) \subset [a, b] \subset (0, \infty]$. As approximation $\tilde{f}$ we choose the best exponential sum

$$E_{r,I}(x) = \sum\nolimits_{\nu=1}^{r} \omega_{\nu,I} \exp(-\alpha_{\nu,I} x) \quad \text{on } I,$$

where $I \supset [a, b]$. Then

$$E_{r,I}(M) = \sum\nolimits_{\nu=1}^{r} \omega_{\nu,I} \exp(-\alpha_{\nu,I} M)$$

approximates $M^{-1}$ exponentially well:

$$\|f(M) - \tilde{f}(M)\|_2 \leq \varepsilon\left(\frac{1}{x}, I, r\right). \tag{5.15}$$

The approximation of $M^{-1}$ seems rather impractical, since matrix exponentials $\exp(-t_\nu M)$ must be evaluated. However, the interesting applications are matrices which are sums of certain Kronecker products. The matrix

$$\mathbf{M} = \sum_{j=1}^{d} I \otimes \cdots \otimes M^{(j)} \otimes \cdots \otimes I \in \mathcal{R}_d, \quad M^{(j)} \in \mathbb{K}^{I_j \times I_j} \tag{5.16}$$

(factor $M^{(j)}$ at the $j$th position) has the exponential

$$\exp(\mathbf{M}) = \bigotimes_{j=1}^{d} \exp(M^{(j)}). \tag{5.17}$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Let $M^{(j)}$ be positive definite with extreme eigenvalues $0 < \lambda_{\min}^{(j)} \le \lambda_{\max}^{(j)}$ for $1 \le j \le d$. Since the spectrum of $\mathbf{M}$ is the sum $\sum_{j=1}^{d} \lambda^{(j)}$ of all $\lambda^{(j)} \in \sigma(M^{(j)})$, the interval $I = [a, b]$ containing the spectrum $\sigma(\mathbf{M})$ is given by

$$a := \sum_{j=1}^{d} \lambda_{\min}^{(j)} > 0 \quad \text{and} \quad b := \sum_{j=1}^{d} \lambda_{\max}^{(j)}.$$

In the case of an unbounded self-adjoint operator, $b = \infty$ is chosen. Let

$$E_{r,I}(x) = \sum\nolimits_{\nu=1}^{r} \omega_{\nu,I} \exp(-\alpha_{\nu,I} x)$$

be the optimal exponential sum approximation of $1/x$ on $I$. These preparations lead us to the following statement.

**Proposition 5.10.** Let $M^{(j)} \in \mathbb{K}^{I_j \times I_j}$ be positive definite matrices with extreme eigenvalues $\lambda_{\min}^{(j)}$ and $\lambda_{\max}^{(j)}$. Then the inverse $\mathbf{M}^{-1}$ of $\mathbf{M}$ from (5.16) can be approximated by

$$\mathbf{B} := \sum_{\nu=1}^{r} \omega_{\nu,I} \bigotimes_{j=1}^{d} \exp(-\alpha_{\nu,I} M^{(j)}).$$

The error is given by

$$\|\mathbf{M}^{-1} - \mathbf{B}\|_2 \le \varepsilon\left(\frac{1}{x}, [a, b], r\right)$$

with $a := \sum_{j=1}^{d} \lambda_{\min}^{(j)}$, and $b := \sum_{j=1}^{d} \lambda_{\max}^{(j)}$.

*Proof.* Apply (5.15) with $f(x) = 1/x$ and $\tilde{f} = E_{r,I}$ to $\mathbf{M}$. For $\exp(-\alpha_{\nu,I}\mathbf{M})$ appearing in $\mathbf{B} := E_{r,I}(\mathbf{M})$ use the representation (5.17 ) with the error estimate $\|\mathbf{M}^{-1} - \mathbf{B}\|_2 \le \varepsilon\big(\frac{1}{x}, [a, b], r\big)$. □

We still need to compute the $\exp(-\alpha_{\nu,I} M^{(j)})$. As described in Hackbusch (2009, § 13.3.1) and Gavrilyuk, Hackbusch and Khoromskij (2002), the hierarchical matrix technique allows us to approximate $\exp(-\alpha_{\nu,I} M^{(j)})$ with a cost that is almost linear in $\#I_j$. The total number of arithmetic operations is

$$O\left(r \sum_{j=1}^{d} \#I_j \log^* \#I_j\right).$$

For $\#I_j = n$ $(1 \le j \le d)$, this expression is $O(rdn \log^* n)$ and depends only linearly on $d$. For identical $M^{(j)} = M^{(k)}$ $(1 \le j, k \le d)$, the cost $O(rdn \log^* n)$ reduces to $O(rn \log^* n)$.

Proposition 5.10 can in particular be applied to the Laplace operator and its discretizations, as detailed below.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Remark 5.11.**

(a) The negative Laplace operator[15] $\Delta : H_0^1([0,1]^d) \to H^{-1}([0,1]^d)$ has a $d$-term format with $M^{(j)} = -\partial^2/\partial x_j^2$ and $\lambda_{\min}^{(j)} = \pi^2$, $\lambda_{\max}^{(j)} = \infty$.

(b) A finite difference[16] discretization in an equidistant grid of step size $1/n$ yields the tridiagonal matrix[17] $M^{(j)} = n^{-2} \cdot \mathrm{tridiag}\{-1, 2, -1\}$ with extreme eigenvalues

$$\lambda_{\min}^{(j)} = 4n^2 \sin^2\left(\frac{\pi}{2n}\right) \approx \pi^2, \quad \lambda_{\max}^{(j)} = 4n^2 \cos^2\left(\frac{\pi}{2n}\right) \approx 4n^2.$$

This approach to the inverse allows us to treat cases with large $n$ and $d$. Grasedyck (2004) presents examples with $n = 1024$ and $d \approx 1000$. Note that in this case the matrix is of size $\mathbf{M}^{-1} \in \mathbb{R}^{M \times M}$ with $M \approx 10^{3000}$.

The approximation method can be extended to any separable differential operator in tensor domains $D = \times_{j=1}^d D_j$ with an appropriate spectrum.

**Definition 5.12.** A differential operator $L = \sum_{j=1}^d L_j$ is called *separable* if $L_j$ contains only derivatives with respect to $x_j$ and has coefficients which only depend on $x_j$.

So far, we have applied the exponential sum $E_r \approx 1/x$. Analogous statements can be made about the application of $E_r \approx 1/\sqrt{x}$. Then $r$-term approximations of $\mathbf{A}^{-1/2}$ can be computed.

### *5.11. Sparse grid methods*

For a review of sparse grids we refer to Bungartz and Griebel (2004) and Garcke (2013). Here, we only sketch its relation to tensor representations. To simplify the notation, we assume that the tensor space $\mathbf{V} = \bigotimes_{j=1}^d V_j$ uses identical spaces $V = V_j$ possessing a nested sequence of subspaces:

$$V = V_{(\ell)} \supset V_{(\ell-1)} \supset \cdots \supset V_{(2)} \supset V_{(1)}.$$

Typical examples are finite element spaces $V_{(\ell)}$ of functions, say, on the interval $[0,1]$ using the step size $2^{-\ell}$ and $n := 2^\ell$ subintervals. While the usual uniform discretization by $\mathbf{V} = \otimes^d V_{(\ell)}$ has a dimension of order $2^{\ell d} = n^d$, the

[15] The reference to $H_0^1([0,1]^d)$ indicates vanishing Dirichlet values.

[16] A finite element discretization with piecewise linear elements in the same grid leads to a finite difference matrix $M^{(j)}$ with an additional scaling by a factor of $1/n$, but the identity matrices in $I \otimes \cdots \otimes M^{(j)} \otimes \cdots \otimes I$ are to be replaced by the mass matrix $\mathrm{tridiag}\{1/6, 2/3, 1/6\}$. Therefore, the construction of the approximate inverse has to be modified: see Hackbusch (2012, § 9.7.2.6).

[17] In this case, a cheap and exact evaluation of $\exp(M^{(j)})$ can be obtained by diagonalization of $M^{(j)}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

sparse grid approach uses the sum of tensor spaces,

$$\mathbf{V}_{\mathrm{sg},\ell} = \sum_{\sum_{j=1}^{d} \ell_j = \ell+d-1} \bigotimes_{j=1}^{d} V_{(\ell_j)}.$$

The background is the estimation of the interpolation error[18] by $O(2^{-2\ell}\ell^{d-1})$ for suitably regular functions (Bungartz and Griebel 2004, Theorem 3.8). This is to be compared with $\dim(\mathbf{V}_{\mathrm{sg}}) \approx 2^{\ell}\ell^{d-1} = n\log^{d-1} n$ (Bungartz and Griebel 2004, (3.63)). The basis vectors in $\mathbf{V}_{\mathrm{sg},\ell}$ are elementary tensors $\bigotimes_{j=1}^{d} b_{k,\ell_j}^{(j)}$, where $\ell_j$ denotes the level, $b_{k,\ell_j}^{(j)} \in V_{(\ell_j)}$. Since the number of terms is limited by the dimension of $\mathbf{V}_{\mathrm{sg},\ell}$, the tensor $\mathbf{v} \in \mathbf{V}_{\mathrm{sg},\ell}$ belongs to $\mathcal{R}_r$ with $r = \dim(\mathbf{V}_{\mathrm{sg},\ell}) \approx n\log^{d-1} n$. However, the size of the representation rank $r$ can be reduced to $O(n^{(d-1)/d})$ without changing the sparse grid approximation (Hackbusch 2012, § 7.6.5). Hence, the sparse grid error bound from above can be achieved by tensors from $\mathcal{R}_r$ with representation rank $r = O(n^{(d-1)/d})$.

### *5.12. Conclusion*

If a tensor can be represented in the $r$-term format $\mathbf{v} = \sum_{\nu=1}^{r} v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$ with moderate $r$, this is usually the best possible choice. The storage size $rdn$ is better than for the alternative formats, provided that $r$ compares with the different ranks of the other formats. The computational cost of the operations is either $O(rdn)$ or $O(r^2dn)$, whereas the other formats lead to higher powers of $r$.

There are important tensors which can be represented very well in the $r$-term format, that is, the accuracy improves exponentially with $r$.

Since there is no orthogonality of the vectors[19] $(v_\nu^{(j)})_{\nu=1,\dots,r}$, one has to ensure that the condition number $\varkappa\big((v_i^{(j)})\big)$ from Section 5.4 is not too large.

The main disadvantage is the approximation within $\mathcal{R}_r$, which need not exist as best approximation. The truncation procedure may require a stabilization. In any case, it is an iterative process which is not fully reliable.

## 6. Tensor subspace format

The following tensor subspace format is often called the Tucker format (Tucker 1966). Algebraically, it uses the concept of suitable subspaces as explained in Section 6.1. For the numerical realization, the subspaces are described by means of bases (see Section 6.2).

[18] Any $L^p$ norm with $2 \le p \le \infty$ can be chosen.

[19] One can force the vectors $(v_\nu^{(j)})_{\nu=1,\dots,r}$ to be orthogonal, but then $r$ increases greatly.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*6.1. Tensor subspaces*

Among all subspaces of a tensor spaces $\mathbf{V} = \bigotimes_{j=1}^{d} V_j$, there are special subspaces of tensor product form,

$$\mathbf{U} = \bigotimes_{j=1}^{d} U_j, \quad U_j \subset V_j,$$

which we call *tensor subspaces* (*i.e.*, $\mathbf{U}$ is a subspace of $\mathbf{V}$ *and* a tensor space). Given a tensor $\mathbf{v} \in \mathbf{V}$, we may ask for a tensor subspace $\mathbf{U}$ such that $\mathbf{v} \in \mathbf{U}$. Even when the spaces $V_j$ are infinite-dimensional, $U_j$ is finite-dimensional for the tensors of interest.

**Remark 6.1.** For algebraic tensors $\mathbf{v} \in \mathbf{V} = {}_a\bigotimes_{j=1}^{d} V_j$ there are finite-dimensional subspaces $U_j \subset V_j$ such that $\mathbf{v} \in \mathbf{U}$. In particular, the inequality $\dim(U_j) \leq \operatorname{rank}(\mathbf{v})$ holds.

*Proof.* Given any $r$-term representation

$$\mathbf{v} = \sum_{\nu=1}^{r} \bigotimes_{j=1}^{d} v_\nu^{(j)},$$

set $U_j := \operatorname{span}\{v_\nu^{(j)} : 1 \leq \nu \leq r\}$, which implies $\dim(U_j) \leq r$. Since $r = \operatorname{rank}(\mathbf{v})$ is a possible choice, the inequality follows. □

As we shall see in more detail in Section 6.4, any algebraic tensor $\mathbf{v}$ possesses subspaces $U_j^{\min}(\mathbf{v})$ of minimal dimension ('minimal subspaces'). Minimality is defined by the following two conditions:

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

In the case of the finite-dimensional spaces $V_j = \mathbb{K}^{n_j}$, the matricization $\mathcal{M}_j(\mathbf{v}) \in \mathbb{K}^{n_j \times n_{[j]}}$ (see Section 2.6) can be used to characterize $U_j^{\min}(\mathbf{v})$ by the range of the matrix:

$$U_j^{\min}(\mathbf{v}) = \operatorname{range}(\mathcal{M}_j(\mathbf{v})).$$

This identity implies that

$$\operatorname{rank}_j(\mathbf{v}) := \dim(U_j^{\min}(\mathbf{v})) \tag{6.1}$$

is the Tucker rank introduced in (2.9). For all cases different from $V_j = \mathbb{K}^{n_j}$, we shall use (6.1) as definition of the $j$th rank $r_j$ (in the general case the definition of $U_j^{\min}(\mathbf{v})$ is provided in Section 6.4). The previous properties can be summarized in the following remark.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

**Remark 6.2.** The statement of Remark 6.1 can be strengthened: among all subspaces $U_j$ with $\mathbf{v} \in \bigotimes_{j=1}^d U_j$, there are minimal subspaces $U_j^{\min}(\mathbf{v})$, whose dimension is $\operatorname{rank}_j(\mathbf{v})$.

The proof of Remark 6.1 uses the subspaces $U_j = \operatorname{span}\{v_\nu^{(j)} : 1 \le \nu \le r\}$ of an $r$-term representation. In the case of $r = \operatorname{rank}(\mathbf{v})$, it can be shown that $U_j = U_j^{\min}(\mathbf{v})$ (Hackbusch 2012, Proposition 7.8).

For a given tuple $\mathbf{r} = (r_1, \dots, r_d) \in \mathbb{N}_0^d$ we define the set

$$\mathcal{T}_{\mathbf{r}} := \left\{ \mathbf{v} \in \bigotimes_{j=1}^d U_j : U_j \text{ subspace with } \dim(U_j) \le r_j \right\}. \tag{6.2}$$

**Remark 6.3.**

(a) $\mathbf{v} \in \mathcal{T}_{\mathbf{r}}$ holds if and only if $\operatorname{rank}_j(\mathbf{v}) \le r_j$ for $1 \le j \le d$.

(b) $\mathcal{T}_{(1,\dots,1)}$ is the set of elementary tensors and coincides with $\mathcal{R}_1$ from (2.4).

(c) $\mathcal{T}_{\mathbf{r}} \subset \mathcal{T}_{\mathbf{s}}$ holds for tuples $\mathbf{r}$ and $\mathbf{s}$ with $r_j \le s_j$.

(d) For $d = 2$, the identity $\mathcal{T}_{(r,r)} = \mathcal{R}_r$ is valid for all $r \in \mathbb{N}_0$.

*Proof.* By definition, $\mathbf{v} \in \mathcal{T}_{\mathbf{r}}$ implies the inequality $\operatorname{rank}_j(\mathbf{v}) \le r_j$. The opposite implication follows by Remark 6.2. Parts (b) and (c) are trivial. Part (d) is left to the reader. □

*6.2. Practical representation and storage*

Assume $\mathbf{v} \in \bigotimes_{j=1}^d U_j$. The subspaces $U_j$ have to be described as a span of vectors $\{b_i^{(j)} : 1 \le i \le r_j\}$. In the most general case, this set may be a frame, but usually one chooses a basis (in that case, $r_j$ coincides with $\dim(U_j)$). In the latter case, the tensors $\bigotimes_{j=1}^d b_{i_j}^{(j)}$ for all $1 \le i_j \le r_j$ form a basis of $\bigotimes_{j=1}^d U_j$ (see Proposition 2.1(b)). Hence, there are unique coefficients $a[i_1, \dots, i_d]$ such that

$$\mathbf{v} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{a}[i_1, \dots, i_d] \bigotimes_{j=1}^d b_{i_j}^{(j)}. \tag{6.3}$$

Representation (6.3) of $\mathbf{v}$ is called a *tensor subspace format* or *Tucker format*.

Among all possible choices for a basis, an *orthonormal basis* [20] promises numerical stability and further advantageous properties. In that case, (6.3) may be called an *orthonormal tensor subspace format*.

[20] Here, $V_j$ is assumed to be a Hilbert or pre-Hilbert space, so that orthonormality can be defined.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The following remark summarizes some of the properties of the format.

**Remark 6.4.**

(a) The coefficients $a[i_1, \ldots, i_d]$ in (6.3) form the so-called *core tensor* $\mathbf{a} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_j}$.

(b) If $V_j = \mathbb{K}^{n_j}$, the basis vectors $b_{i_j}^{(j)}$ $(1 \le i_j \le r_j,\ 1 \le j \le d)$ require a storage of $\sum_{j=1}^{d} r_j n_j \le rdn$, where $n := \max_j n_j$ and $r := \max_j r_j$.

(c) The storage of the core tensor $\mathbf{a}$ amounts to $\prod_{j=1}^{d} r_j \le r^d$.

(d) The minimal $r_j$ of any tensor subspace representation (6.3) is the $j$th rank of $\mathbf{v}$ (then $U_j = U_j^{\min}(\mathbf{v})$).

(e) If $V_j = \mathbb{K}^{n_j}$, a singular value decomposition of the matricization

$$\mathcal{M}_j(\mathbf{v}) = \sum_{i=1}^{\operatorname{rank}_j(\mathbf{v})} \sigma_i^{(j)} b_i^{(j)} (v_i^{(j)})^{\mathsf{T}} \quad \text{with} \quad \sigma_i^{(j)} > 0$$

yields an orthonormal basis $\{b_i^{(j)} : 1 \le i \le r_j\}$ of $U_j = U_j^{\min}(\mathbf{v})$.

(f) If $\{b_i^{(j)} : 1 \le i \le r_j\}$, $1 \le j \le d$, are bases, the different ranks of $\mathbf{v} \in \mathbf{V}$ and $\mathbf{a} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_j}$ coincide:

$$\operatorname{rank}(\mathbf{v}) = \operatorname{rank}(\mathbf{a}) \quad \text{and} \quad \operatorname{rank}_j(\mathbf{v}) = \operatorname{rank}_j(\mathbf{a}) \quad \text{for } 1 \le j \le d.$$

(g) If $\{b_i^{(j)} : 1 \le i \le r_j\}$, $1 \le j \le d$, are orthonormal bases, the scalar product $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{V}}$ in $\mathbf{V}$ coincides with the Euclidean scalar product $\langle \mathbf{a}_v, \mathbf{a}_w \rangle$ of the corresponding core tensors. The corresponding norms satisfy $\|\mathbf{v}\|_{\mathbf{V}} = \|\mathbf{a}\|$.

*Proof.* Statements (a)–(d) and (g) are either trivial or have already been mentioned.

Concerning (e), note that $\{b_i^{(j)} : 1 \le i \le \operatorname{rank}_j(\mathbf{v})\}$ spans the range of $\mathcal{M}_j(\mathbf{v})$, which coincides with $U_j^{\min}(\mathbf{v})$.

Part (f) follows from the fact that $\mathbf{u} \in \bigotimes_{j=1}^{d} U_j \mapsto \mathbf{a} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_j}$ is a tensor space isomorphism (see Remarks 2.7 and 2.10). □

The disadvantage of the tensor subspace representation can be seen from part (c). If $d$ becomes large, the storage (and the computational cost of operations) increases exponentially. If $d \le 3$, the size $r^d$ may be tolerable; otherwise, the hierarchical format from Section 7 is more advisable.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*6.3. Higher-order SVDs*

Remark 6.4(e) indicates a possibility of computing an orthonormal basis of $U_j^{\min}(\mathbf{v})$. Equivalently, one can diagonalize the matrix[21]

$$\mathcal{M}_j(\mathbf{v})\mathcal{M}_j(\mathbf{v})^{\mathsf{H}} = \sum_{i=1}^{\operatorname{rank}_j(\mathbf{v})} (\sigma_i^{(j)})^2\, b_i^{(j)} (b_i^{(j)})^{\mathsf{H}}.$$

The basis $\{b_i^{(j)} : 1 \le i \le \operatorname{rank}_j(\mathbf{v})\}$ from Remark 6.4(e) is called the *HOSVD basis* (HOSVD: higher-order singular value decomposition).

If $\mathbf{v}$ is given in full tensor format, the generation of $\mathcal{M}_j(\mathbf{v})\mathcal{M}_j(\mathbf{v})^{\mathsf{H}}$ is rather costly: $O(n_j \prod_{k=1}^d n_k)$ operations. If $\mathbf{v}$ is represented by (6.3) with respect to another basis (or frame), the generation of $\mathcal{M}_j(\mathbf{v})\mathcal{M}_j(\mathbf{v})^{\mathsf{H}}$ costs $O(r_j \prod_{k=1}^d r_k)$ scalar products $\langle \mathbf{b_i}, \mathbf{b_{i'}} \rangle$ and the same number of arithmetic operations. Of course, if the bases are orthonormal, the computation of $\langle \mathbf{b_i}, \mathbf{b_{i'}} \rangle$ can be avoided.

*6.4. Minimal subspaces*

The minimal subspace $U_j^{\min}(\mathbf{v})$ is defined in Section 6.1 for the case $V_j = \mathbb{K}^{n_j}$. For the general case we give two equivalent definitions. First, $\mathbf{v} \in {}_a\bigotimes_{j=1}^d V_j$ is assumed to be an algebraic tensor. We recall that $\mathbf{V}'_{[j]} = (\bigotimes_{k\neq j} V_k)' = L(\bigotimes_{k\neq j} V_k, \mathbb{K})$ is the algebraic dual space. A functional $\varphi_{[j]} \in \mathbf{V}'_{[j]}$ gives rise to a mapping $\hat{\varphi}_{[j]} \in L(\mathbf{V}, V_j)$ via

$$\hat{\varphi}_{[j]}\left(\bigotimes_{k=1}^d u^{(k)}\right) := \varphi_{[j]}\left(\bigotimes_{k\neq j} u^{(k)}\right) \cdot u^{(j)}$$

(see (2.3)). The general definition of the minimal subspace is

$$U_j^{\min}(\mathbf{v}) := \left\{\hat{\varphi}_{[j]}(\mathbf{v}) : \varphi_{[j]} \in \mathbf{V}'_{[j]}\right\}. \tag{6.4}$$

Now we assume normed spaces $(\mathbf{V}, \|\cdot\|)$ and $(V_j, \|\cdot\|_j)$. Furthermore, $\mathbf{V}_{[j]}$ must be equipped with a norm $\|\cdot\|_{[j]}$ such that

$$\|\cdot\| \text{ is not weaker than the injective norm } \|\cdot\|_{\vee(V_j,\mathbf{V}_{[j]})}. \tag{6.5}$$

Condition (6.5) holds in particular if $\|\cdot\|$ is a reasonable crossnorm (see Definition 3.5).

**Lemma 6.5.** Assume (6.5) and $\varphi_{[j]} \in \mathbf{V}^*_{[j]}$ (*i.e.*, $\varphi_{[j]}$ is a *continuous* functional). Then, $\hat{\varphi}_{[j]} \in \mathcal{L}(\mathbf{V}, V_j)$ is also continuous.

[21] In the case of $\mathbb{K} = \mathbb{C}$, $\cdot^{\mathsf{H}}$ denotes the complex conjugate version of the transpose $\cdot^{\mathsf{T}}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*Proof.* By definition, $\|\mathbf{v}\|_{\vee(V_j,\mathbf{V}_{[j]})} = \sup_{\varphi_j} \sup_{\varphi_{[j]}} |(\varphi_j \circ \varphi_{[j]})(\mathbf{v})|$ holds with suprema taken over $\varphi_j \in V_j^*$ and $\varphi_{[j]} \in \mathbf{V}_{[j]}^*$ with $\|\varphi_j\|_j^* = \|\varphi_{[j]}\|_{[j]}^* = 1$. Since

$$\sup_{\varphi_j} \sup_{\varphi_{[j]}} |(\varphi_j \circ \varphi_{[j]})(\mathbf{v})| = \sup_{\varphi_{[j]}} \sup_{\varphi_j} |\varphi_j(\hat{\varphi}_{[j]}(\mathbf{v}))|$$

and $\sup_{\varphi_j} |\varphi_j(v_j)| = \|v_j\|_j$, we conclude that

$$\|\mathbf{v}\|_{\vee(V_j,\mathbf{V}_{[j]})} = \sup_{\varphi_{[j]}} \|\hat{\varphi}_{[j]}(\mathbf{v})\|_j.$$

Hence, any $\hat{\varphi}_{[j]}$ derived from $\varphi_{[j]} \in \mathbf{V}_{[j]}^*$ with $\|\varphi_{[j]}\|_{[j]}^* \le 1$ satisfies

$$\|\hat{\varphi}_{[j]}(\mathbf{v})\|_j \le \|\mathbf{v}\|_{\vee(V_j,\mathbf{V}_{[j]})} \le C\|\mathbf{v}\|,$$

where the last inequality follows from (6.5). For general $\varphi_{[j]} \in \mathbf{V}_{[j]}^*$ we have $\|\hat{\varphi}_{[j]}(\mathbf{v})\|_j \le C\|\mathbf{v}\|\|\varphi_{[j]}\|_{[j]}^*$, that is, $\|\hat{\varphi}_{[j]}\|_{V_j \leftarrow \mathbf{V}} \le C\|\varphi_{[j]}\|_{[j]}^*$ and $\hat{\varphi}_{[j]} \in \mathcal{L}(\mathbf{V}, V_j)$. □

Since $\varphi_{[j]} \in \mathbf{V}_{[j]}^*$ is continuous, it is also well-defined for topological tensors $\mathbf{v} \in \mathbf{V}$. The second definition replaces $\mathbf{V}'_{[j]}$ by $\mathbf{V}_{[j]}^*$:

$$U_j^{\min}(\mathbf{v}) = \{\hat{\varphi}_{[j]}(\mathbf{v}) : \varphi_{[j]} \in \mathbf{V}_{[j]}^*\}. \tag{6.6}$$

Although, in general, $\mathbf{V}_{[j]}^*$ is smaller than $\mathbf{V}'_{[j]}$, both definitions (6.4) and (6.6) coincide for algebraic tensors $\mathbf{v}$. The proof uses the fact that $U_j^{\min}(\mathbf{v})$ is finite-dimensional and, by the Hahn–Banach theorem, functionals on finite-dimensional subspaces can be extended to continuous functionals. If $\mathbf{v}$ is a topological tensor but not algebraic, at least two of the subspaces $U_j^{\min}(\mathbf{v})$ are infinite-dimensional.

Next, we consider a sequence $\mathbf{v}_n \in \mathbf{V}$ of tensors of a normed tensor space $\|\cdot\| \bigotimes_{j=1}^d V_j$, which converges weakly to $\mathbf{v} \in \mathbf{V}$:

$$\mathbf{v}_n \rightharpoonup \mathbf{v}, \quad \textit{i.e.}, \quad \varphi(\mathbf{v}_n) \to \varphi(\mathbf{v}) \quad \text{for all functionals } \varphi \in \mathbf{V}^*.$$

Each $\mathbf{v}_n$, as well as $\mathbf{v}$, gives rise to subspaces $U_j^{\min}(\mathbf{v}_n)$ and $U_j^{\min}(\mathbf{v})$. Their dimensions are subject of the following theorem.

**Theorem 6.6.** Under condition (6.5), weak convergence $\mathbf{v}_n \rightharpoonup \mathbf{v}$ implies

$$\dim(U_j^{\min}(\mathbf{v})) \le \liminf_{n\to\infty} \dim(U_j^{\min}(\mathbf{v}_n)). \tag{6.7}$$

*Proof.* Restrict the sequence to a subsequence such that $\dim(U_j^{\min}(\mathbf{v}_n)$ is weakly increasing. Then the right-hand side of inequality (6.7) becomes $\lim_{n\to\infty} \dim(U_j^{\min}(\mathbf{v}_n))$. If the limit is infinity, (6.7) is a trivial estimate. Otherwise, for a finite limit $N := \lim_{n\to\infty} \dim(U_j^{\min}(\mathbf{v}_n))$ we disprove the inequality $\dim(U_j^{\min}(\mathbf{v})) \ge N+1$. In the latter case, select $N+1$ functionals

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

$\varphi_{[j],\nu} \in \mathbf{V}^*_{[j]}$ such that

$$v^{(j)}_\nu := \varphi_{[j],\nu}(\mathbf{v}) \in U^{\min}_j(\mathbf{v}) \quad (1 \le \nu \le N+1)$$

are linearly independent. Then there are functionals $\varphi_{j,\nu} \in V_j^*$ such that $\varphi_{j,\nu}(v^{(j)}_\mu) = \delta_{\nu\mu}$. Hence, $\det(\varphi_{j,\nu}(\varphi_{[j],\mu}(\mathbf{v}))) = 1$. The composition $\varphi_{j,\nu} \circ \varphi_{[j],\mu}$ belongs to $\mathbf{V}^*$, since $\varphi_{[j],\mu} : \mathbf{V} \to V_j$ and $\varphi_{j,\nu} : V_j \to \mathbb{K}$ are continuous. Hence, $(\varphi_{j,\nu} \circ \varphi_{[j],\mu})(\mathbf{v}_n) \to (\varphi_{j,\nu} \circ \varphi_{[j],\mu})(\mathbf{v})$ implies

$$\det(\varphi_{j,\nu}(\varphi_{[j],\mu}(\mathbf{v}_n))) \to 1.$$

Therefore, for $n$ sufficiently large, the $N+1$ vectors $\varphi_{[j],\mu}(\mathbf{v}_n) \in U^{\min}_j(\mathbf{v}_n)$ are linearly independent, in contradiction to $\dim(U^{\min}_j(\mathbf{v}_n)) \le N$. □

*6.5. Closedness of $\mathcal{T}_\mathbf{r}$ and existence of best approximations*

As seen in Section 5.4, non-closedness of $\mathcal{R}_r$ causes numerical problems. In contrast to $\mathcal{R}_r$, we can prove that $\mathcal{T}_\mathbf{r}$ is closed. It is even *weakly closed*, that is, weak convergence $\mathbf{v}_n \rightharpoonup \mathbf{v}$ of $\mathbf{v}_n \in \mathcal{T}_\mathbf{r}$ already implies that $\mathbf{v} \in \mathcal{T}_\mathbf{r}$.

**Theorem 6.7.** Assume (6.5). Then $\mathcal{T}_\mathbf{r}$ is weakly closed for any $\mathbf{r} \in \mathbb{N}_0^d$.

*Proof.* We note that $\mathbf{v}_n \in \mathcal{T}_\mathbf{r}$ implies

$$\dim(U^{\min}_j(\mathbf{v}_n) \le r_j \quad \text{and} \quad \liminf \dim(U^{\min}_j(\mathbf{v}_n)) \le r_j.$$

From (6.7) we conclude $\dim(U^{\min}_j(\mathbf{v})) \le r_j$, that is, $\mathbf{v} \in \bigotimes_{j=1}^d U^{\min}_j(\mathbf{v}) \subset \mathcal{T}_\mathbf{r}$. □

Weak convergence is important, as can be seen from the proof of the next statement.

**Theorem 6.8.** Assume that $(\mathbf{V}, \|\cdot\|)$ is reflexive and satisfies (6.5). Then, for any $\mathbf{v} \in \mathbf{V}$, the minimization problem

$$\inf_{\mathbf{u} \in \mathcal{T}_\mathbf{r}} \|\mathbf{v} - \mathbf{u}\| = \|\mathbf{v} - \mathbf{u}_{\mathsf{best}}\|$$

has a solution $\mathbf{u}_{\mathsf{best}} \in \mathcal{T}_\mathbf{r}$.

*Proof.* Choose a sequence $\mathbf{u}_n \in \mathcal{T}_\mathbf{r}$ with $\|\mathbf{v} - \mathbf{u}_n\| \to \inf_{\mathbf{u} \in \mathcal{T}_\mathbf{r}} \|\mathbf{v} - \mathbf{u}\|$. In the case of reflexive Banach spaces, there exists a weakly convergent subsequence $\mathbf{u}_n \rightharpoonup \mathbf{u}_{\mathsf{best}}$ for some $\mathbf{u}_{\mathsf{best}} \in \mathbf{V}$ with the property $\inf_{\mathbf{u} \in \mathcal{T}_\mathbf{r}} \|\mathbf{v} - \mathbf{u}\| = \|\mathbf{v} - \mathbf{u}_{\mathsf{best}}\|$. Theorem 6.7 implies that $\mathbf{u}_{\mathsf{best}} \in \mathcal{T}_\mathbf{r}$. □

The requirements of Theorem 6.8 (reflexivity and (6.5)) are satisfied in particular if $\mathbf{V}$ is finite-dimensional.

Supposition (6.5) does not hold for tensor spaces like $H^1(\Omega)$, as mentioned in Remark 3.9. Nevertheless, the statements of Theorems 6.6–6.8 can be proved. In this case, equip the algebraic tensor space $\mathbf{V}^{\mathrm{alg}}_{[j]} := \bigotimes_{k \ne j} V_k$ with

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

the $L^2$ norm on $I_1 \times \cdots \times I_{k-1} \times I_{k+1} \times \cdots \times I_d$. Let $\mathbf{V}_{[j]}$ be the completion with respect to this norm. $V_j$ carries the norm $H^1(I_j)$. Then one verifies that the injective norm $\|\cdot\|_{\vee(V_j, \mathbf{V}_{[j]})}$ satisfies condition (6.5). Hence, Theorems 6.6, 6.7 and 6.8 can be applied.

The fact that a best approximation also exists in the infinite-dimensional case indicates that the conditioning of the minimization problem does not deteriorate with increasing dimension.

### *6.6. Truncation by HOSVD*

Now we assume the special case of the Euclidean norm (and Euclidean scalar product) in $V_j$ and $\mathbf{V}$. The following 'truncation' of $\mathbf{v} \in \mathcal{T}_\mathbf{s}$ to $\mathcal{T}_\mathbf{r}$ with $\mathbf{r} \le \mathbf{s}$ is an approximation $\mathbf{v}$ of by $\mathbf{u} \in \mathcal{T}_\mathbf{r}$, which can be determined explicitly by means of the HOSVD bases.

Let $\{b_i^{(j)} : 1 \le i \le \mathrm{rank}_j(\mathbf{v})\}$ be the HOSVD bases of $\mathbf{v}$, and transform the tensor with respect to these (orthonormal) bases:

$$\mathbf{v} = \sum_{i_1=1}^{\mathrm{rank}_1(\mathbf{v})} \cdots \sum_{i_d=1}^{\mathrm{rank}_d(\mathbf{v})} \mathbf{a}[i_1, \ldots, i_d] \bigotimes_{j=1}^{d} b_{i_j}^{(j)}.$$

Fix any new representation ranks $r_j \le \mathrm{rank}_j(\mathbf{v})$, which define the tuple $\mathbf{r} = (r_1, \ldots, r_d)$. Then the truncation to $\mathcal{T}_\mathbf{r}$ is defined by

$$\mathbf{u}_{\mathrm{HOSVD}} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{a}[i_1, \ldots, i_d] \bigotimes_{j=1}^{d} b_{i_j}^{(j)}; \tag{6.8}$$

that is, all terms with $i_j > r_j$ for at least one $j$ are omitted. Note that the indices $i_j > r_j$ belong to the smallest singular values $\sigma_{i_j}^{(j)}$ of the singular value decomposition of $\mathcal{M}_j(\mathbf{v})$ (see Section 6.3). Here $\mathbf{u}_{\mathrm{HOSVD}}$ is the orthogonal projection of $\mathbf{v}$ onto the subspace spanned by $\bigotimes_{j=1}^{d} b_{i_j}^{(j)}$ for $1 \le i_j \le r_j$.

The next lemma describes an error bound as well as an estimation by the best approximation error. As a result, the truncation result is *quasi-optimal*, that is, optimal up to a fixed factor.

**Theorem 6.9.** The approximation $\mathbf{u}_{\mathrm{HOSVD}}$ defined above satisfies the estimate

$$\|\mathbf{v} - \mathbf{u}_{\mathrm{HOSVD}}\| \le \sqrt{\sum_{j=1}^{d} \sum_{i=r_j+1}^{\mathrm{rank}_j(\mathbf{v})} (\sigma_i^{(j)})^2} \le \sqrt{d} \|\mathbf{v} - \mathbf{u}_{\mathrm{best}}\|, \tag{6.9}$$

where $\mathbf{u}_{\mathrm{best}} \in \mathcal{T}_\mathbf{r}$ yields the minimal error, that is,

$$\|\mathbf{v} - \mathbf{u}_{\mathrm{best}}\| = \min_{\mathbf{u} \in \mathcal{T}_\mathbf{r}} \|\mathbf{v} - \mathbf{u}\|.$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

For a proof see Hackbusch (2012, Theorem 10.3). The first part of (6.9) is given by De Lathauwer, De Moor and Vandewalle (2000*a*, Property 10), while the second part is shown by Grasedyck (2010*a*).

There are two different practical applications of this result. First, one can fix the rank tuple $\mathbf{r}$ and apply the truncation. Then (6.9) yields an upper bound of the truncation error. Second, one can fix a tolerance $\varepsilon$ and collect the smallest $\sigma_i^{(j)}$ such that

$$\sum_{j=1}^{d} \sum_{i=r_j+1}^{\mathrm{rank}_j(\mathbf{v})} (\sigma_i^{(j)})^2 \le \varepsilon^2.$$

Then the truncation is performed with respect to the adaptively determined ranks $r_j$.

The truncation from above is symmetric with respect to the directions $j$. There exists another variant which is based on a certain sequence of directions. For the ordering $j = 1, \dots, d$ the sequential truncation is defined as follows. Again, we assume that desired ranks $r_j$ are given. We start with the representation

$$\mathbf{v} = \sum_{i_1=1}^{s_1} \cdots \sum_{i_d=1}^{s_d} \mathbf{a}_0[i_1, \dots, i_d] \bigotimes_{j=1}^{d} b_{i_j,\mathsf{old}}^{(j)}.$$

- $j = 1$. Determine the HOSVD basis $\{b_i^{(1)} : 1 \le i \le \mathrm{rank}_1(\mathbf{v})\}$ of $\mathbf{v}$. Transform the representation so that

$$\mathbf{v} = \sum_{i_1=1}^{\mathrm{rank}_1(\mathbf{v})} \sum_{i_2=1}^{s_2} \cdots \sum_{i_d=1}^{s_d} \mathbf{a}_1[i_1, \dots, i_d] b_{i_1}^{(1)} \otimes \bigotimes_{j=2}^{d} b_{i_j,\mathsf{old}}^{(j)}.$$

  Set

$$\mathbf{v}_1 = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{s_2} \cdots \sum_{i_d=1}^{s_d} \mathbf{a}_1[i_1, \dots, i_d]\, b_{i_1}^{(1)} \otimes \bigotimes_{j=2}^{d} b_{i_j,\mathsf{old}}^{(j)}.$$

- $j = 2$. Determine the HOSVD basis $\{b_i^{(2)} : 1 \le i \le \mathrm{rank}_2(\mathbf{v}_1)\}$ of $\mathbf{v}_1$. Transform $\mathbf{v}_1$ with respect to the new HOSVD basis:

$$\mathbf{v}_1 = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{\mathrm{rank}_2(\mathbf{v}_1)} \sum_{i_3=1}^{s_3} \cdots \sum_{i_d=1}^{s_d} \mathbf{a}_2[i_1, \dots, i_d] b_{i_1}^{(1)} \otimes b_{i_2}^{(1)} \otimes \bigotimes_{j=3}^{d} b_{i_j,\mathsf{old}}^{(j)},$$

  and set

$$\mathbf{v}_2 = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \sum_{i_3=1}^{s_3} \cdots \sum_{i_d=1}^{s_d} \mathbf{a}_2[i_1, \dots, i_d] b_{i_1}^{(1)} \otimes b_{i_2}^{(1)} \otimes \bigotimes_{j=3}^{d} b_{i_j,\mathsf{old}}^{(j)}.$$

- $j = 3, \dots, d$ is analogous to $j = 1, 2$. The final result is $\tilde{\mathbf{u}}_{\mathrm{HOSVD}} := \mathbf{v}_d \in \mathcal{T}_{\mathbf{r}}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

One advantage of the second variant is that the cost is lower, since the intermediate tensors $\mathbf{v}_j$ already have smaller ranks, $r_1, \ldots, r_{j-1}$. The corresponding error analysis yields the following result (see Hackbusch (2012, Theorem 10.5)). Note that the first part in (6.10) is an equation, not an inequality as in (6.9).

**Theorem 6.10.** The singular values $\tilde{\sigma}_i^{(j)}$ appearing in the second variant satisfy $\tilde{\sigma}_i^{(j)} \le \sigma_i^{(j)}$, where the latter values are those from (6.9). Then

$$\|\mathbf{v} - \tilde{\mathbf{u}}_{\mathrm{HOSVD}}\| = \sqrt{\sum_{j=1}^{d} \sum_{i>r_j} (\tilde{\sigma}_i^{(j)})^2} \le \sqrt{d}\|\mathbf{v} - \mathbf{u}_{\mathrm{best}}\|. \tag{6.10}$$

**Remark 6.11.** The truncation errors (6.9) and (6.10) are expressed in the norm corresponding to the underlying scalar product (usually the Euclidean norm $\|\cdot\|_2$). Although in high dimensions there is no useful estimate of the maximum norm $\|\cdot\|_\infty$ by $\|\cdot\|_2$, one can prove $\|\cdot\|_\infty$ estimates for smooth functions (Hackbusch 2013).

If one is aiming for the best approximation, then iterative methods are available: see De Lathauwer, De Moor and Vandewalle (2000*c*, Alg. 4.2), Ishteva, De Lathauwer, Absil and Van Huffel (2009) and Eldén and Savas (2009). However, there is no guarantee that the limit is the global minimum.

### *6.7. Transfer from $\mathcal{R}_r$ into $\mathcal{T}_{\mathbf{r}}$*

If $\mathbf{v}$ is given in the $r$-term format $\sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}$, there is a trivial translation into the subspace format. Define frames by means of $b_i^{(j)} := v_i^{(j)}$ (note that the vectors $(v_i^{(j)})$ need not be linearly independent) and define the core tensor by $\mathbf{a}[i, \ldots, i] = 1$ for all $1 \le i \le r$ and $\mathbf{a}[\ldots] = 0$ otherwise. Then (6.3) is the subspace representation of $\mathbf{v}$. The core tensor has a special 'diagonal' form. In particular, $\mathbf{a}$ is sparse (only $r$ of the $r^d$ entries may be non-zero). Unfortunately, this property is lost after a change of bases (*e.g.*, after an orthonormalization).

### *6.8. Hybrid format*

The representation (6.3) of $\mathbf{v}$ involves the core tensor $\mathbf{a} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_j}$, which by itself is described in the full format. Therefore, it requires a storage size of $\prod_{j=1}^{d} r_j$. In particular, for large $d$, this is unsatisfactory. One may try to represent $\mathbf{a}$ by a format other than the full format. One verifies that a tensor subspace format for $\mathbf{a}$ does not help, since altogether one obtains a new tensor subspace representation of $\mathbf{v}$, which is improved only if the subspace $U_j$ is of higher dimension than necessary, that is, if $U_j \supsetneqq U_j^{\min}(\mathbf{v})$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

However, one may try to find an $r$-term representation of $\mathbf{a}$. Then the storage cost is reduced to $r\sum_{j=1}^{d} r_j$. The resulting format is called the *hybrid format*.

The hybrid format of $\mathbf{v}$ can be equally well obtained from the $r$-term representation of $\mathbf{v} = \sum_{\nu=1}^{r} v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$. The $r$ vectors $(v_\nu^{(j)})$ spanning the subspace $U_j$ need not be linearly independent. Instead one chooses a basis of $U_j$ and writes each $v_\nu^{(j)}$ as a linear combination of the basis. In the end, one obtains the hybrid representation of $\mathbf{v}$.

### *6.9. Conclusion*

The tensor subspace format is more closely related to matrix properties, since the subspaces may be defined via matricization. A consequence is the weak closedness of the set $\mathcal{T}_{\mathbf{r}}$, which also establishes the existence of a best approximation in infinite-dimensional cases.

Stability of the tensor subspace representation can be enforced by the choice of orthonormal bases $\{b_i^{(j)} : 1 \le i \le r_j\}$ of the subspaces.

Only simple and reliable linear algebra tools are applied, for instance the singular value decomposition or equivalently the diagonalization of $r_j \times r_j$ matrices with $r_j$ not too large. They yield the HOSVD basis, which allows us to define the minimal subspaces and the quasi-optimal truncation to lower dimensions than $r_j$.

The drawback is the appearance of the core tensor $\mathbf{a} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_j}$, whose memory requirement increases exponentially in $d$. The use of the hybrid format of Section 6.8 may be a help in this respect, but imports the drawbacks of the $r$-term format.

## 7. Hierarchical tensor format

### *7.1. Introduction*

We want to keep the advantages of the tensor subspace format, but we have to overcome the appearance of the costly core tensor. We follow the idea of replacing spaces by subspaces and apply it also to intermediate tensor spaces. The concept is demonstrated for the particular example of $\mathbf{V} = \bigotimes_{j=1}^{4} V_j$. This tensor space is isomorphic to $(V_1 \otimes V_2) \otimes (V_3 \otimes V_4)$, which is a tensor space of order two generated by $\mathbf{V}_{\{1,2\}} := V_1 \otimes V_2$ and $\mathbf{V}_{\{3,4\}} := V_3 \otimes V_4$. This setting allows us to introduce again subspaces of $\mathbf{V}_{\{1,2\}}$ and $\mathbf{V}_{\{3,4\}}$. More precisely, we start with subspaces $U_j \subset V_j$ (as for the tensor subspace format). Then $U_1 \otimes U_2$ is already a subspace of $\mathbf{V}_{\{1,2\}}$. To achieve an improvement compared with the tensor subspace format, we introduce a (proper) subspace $\mathbf{U}_{\{1,2\}}$ of $U_1 \otimes U_2$. Similarly, $\mathbf{U}_{\{3,4\}} \subset U_3 \otimes U_4$ is chosen. For any $\mathbf{v} \in \mathbf{U}_{\{1,2\}} \otimes \mathbf{U}_{\{3,4\}}$, we can introduce the one-dimensional subspace $\mathbf{U}_{\{1,2,3,4\}} := \operatorname{span}\{\mathbf{v}\} \subset \mathbf{U}_{\{1,2\}} \otimes \mathbf{U}_{\{3,4\}}$. The

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

following picture illustrates the construction:

$$\begin{array}{ccccc} & & \mathbf{U}_{\{1,2,3,4\}} \subset \mathbf{U}_{\{1,2\}} \otimes \mathbf{U}_{\{3,4\}} \subset \mathbf{V} & & \\ & \swarrow & & \searrow & \\ \mathbf{U}_{\{1,2\}} \subset U_1 \otimes U_2 & & & & \mathbf{U}_{\{3,4\}} \subset U_3 \otimes U_4 \\ \swarrow & \searrow & & & \swarrow & \searrow \\ U_1 \subset V_1 & & U_2 \subset V_2 & & U_3 \subset V_3 & & U_4 \subset V_4 \end{array} \tag{7.1}$$

In diagram (7.1) two structural elements are of importance: (a) the binary tree structure, (b) the association of each vertex with a subspace. The tree will be defined in Section 7.2, while the subspaces are discussed in Section 7.3.

Under a different name, the hierarchical format has been described by Shi, Duan and Vidal (2006). Compare also Lubich (2008, p. 45). Computational issues and truncation are discussed in Hackbusch and Kühn (2009). Another name for this format is the 'hierarchical Tucker format', since the formation of subspaces is also applied in the tensor subspace (= Tucker) format.

### *7.2. Dimension partition tree*

The tree in (7.1) is related to the set

$$D = \{1, \dots, d\}$$

of directions ($d = 4$ in (7.1)). It describes how the set $D$ is partitioned into disjoint subsets:

$$\begin{array}{ccccccc} & & & \{1,2,3,4\} & & & \\ & & \swarrow & & \searrow & & \\ & \{1,2\} & & & & \{3,4\} & \\ & \swarrow & \searrow & & \swarrow & \searrow & \\ \{1\} & & \{2\} & & \{3\} & & \{4\} \end{array}$$

For the later application of singular value decompositions it will be important that the tree is a *binary* tree. Further conditions are not required (*e.g.*, the tree need not be balanced). The vertices of the tree which are non-empty subsets of $D$ will be denoted by $\alpha$.

The exact definition of the *dimension partition tree* $T_D$ is given by the following properties:

- $D$ is the root of $T_D$,
- $\alpha \in T_D$ is a leaf if and only if $\#\alpha = 1$, that is, $\alpha = \{j\}$ for some $j \in D$,
- if $\alpha$ is no leaf, it has two sons $\alpha_1$ and $\alpha_2$ with $\alpha = \alpha_1 \dot\cup \alpha_2$,

where $\dot\cup$ denotes the disjoint union. The leaves define the set $\mathcal{L}(T_D)$. Any binary tree has $2d-1$ vertices ($d$ leaves and $d-1$ vertices from $T_D \backslash \mathcal{L}(T_D)$).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*7.3. Nested subspaces and $\mathcal{H}_{\mathfrak{r}}$*

For all $\alpha \in T_D$ we introduce the notation $\mathbf{V}_\alpha := {}_a\bigotimes_{j\in\alpha} V_j$. In the case of leaves $\alpha = \{j\}$, $\mathbf{V}_{\{j\}} = V_j$ are identical notations. Next, we consider subspaces $\mathbf{U}_\alpha \subset \mathbf{V}_\alpha$ for all vertices $\alpha \in T_D$. As indicated in (7.1), they have to satisfy the nestedness condition

$$\mathbf{U}_\alpha \subset \mathbf{U}_{\alpha_1} \otimes \mathbf{U}_{\alpha_2} \quad (\alpha \in T_D\backslash\mathcal{L}(T_D),\ \alpha_1, \alpha_2 \text{ sons of } \alpha). \tag{7.2a}$$

Their dimensions are denoted by

$$r_\alpha := \dim(\mathbf{U}_\alpha) \quad (\alpha \in T_D) \tag{7.2b}$$

and collected in the tuple

$$\mathfrak{r} := (r_\alpha)_{\alpha\in T_D}. \tag{7.2c}$$

Usually, at the root $\alpha = D$, $\mathbf{U}_D$ coincides with the span of the tensor $\mathbf{v}$ to be represented. Therefore,

$$r_D = 1 \tag{7.2d}$$

is the standard value. In exceptional cases, where more than one tensor has to be represented in the same format, $r_D > 1$ may occur.

In analogy to the sets $\mathcal{R}_r$ and $\mathcal{T}_{\mathbf{r}}$ introduced earlier, we now define the set $\mathcal{H}_{\mathfrak{r}}$ for a tuple $\mathfrak{r}$:

$$\mathcal{H}_{\mathfrak{r}} = \{\mathbf{v} \in \mathbf{U}_D \subset \mathbf{V} : \text{there are subspaces } (\mathbf{U}_\alpha)_{\alpha\in T_D} \text{ satisfying (7.2a–c)}\}.$$

An equivalent definition is

$$\mathcal{H}_{\mathfrak{r}} = \{\mathbf{v} \in \mathbf{V} :\ \mathrm{rank}_\alpha(\mathbf{v}) \le r_\alpha \text{ for all } \alpha \in T_D\} \tag{7.2e}$$

(see (2.10) for $\mathrm{rank}_\alpha$), as we shall see next.

*7.4. Minimal subspaces*

In Section 6.4 we introduced the minimal subspaces $U_j^{\min}(\mathbf{v})$ for $j \in D$ corresponding to the leaves of $T_D$. Now we extend the definition to all $\alpha \in T_D$. The complement of $\alpha$ is denoted by $\alpha^c := D\backslash\alpha$. For $\alpha = D$, we define $\mathbf{U}_D^{\min}(\mathbf{v}) := \mathrm{span}\{\mathbf{v}\}$. Otherwise, we generalize (6.4) by

$$\mathbf{U}_\alpha^{\min}(\mathbf{v}) := \big\{\hat\varphi_{\alpha^c}(\mathbf{v}) : \varphi_{\alpha^c} \in \mathbf{V}'_{\alpha^c}\big\},$$

where $\hat\varphi_{\alpha^c} \in L(\mathbf{V}, \mathbf{V}_\alpha)$ is defined via

$$\hat\varphi_{\alpha^c}\Big(\bigotimes\nolimits_{j\in D} v^{(j)}\Big) := \varphi_{\alpha^c}\Big(\bigotimes\nolimits_{j\in\alpha^c} v^{(j)}\Big) \cdot \bigotimes\nolimits_{j\in\alpha} v^{(j)}$$

(see (2.3)). One easily verifies that

$$\mathbf{U}_\alpha^{\min}(\mathbf{v}) \subset \mathbf{U}_{\alpha_1}^{\min}(\mathbf{v}) \otimes \mathbf{U}_{\alpha_2}^{\min}(\mathbf{v}) \quad (\alpha \in T_D\backslash\mathcal{L}(T_D),\ \alpha_1, \alpha_2 \text{ sons of } \alpha),$$

that is, the minimal subspaces satisfy the nestedness condition (7.2a) (Hackbusch 2012, Corollary 6.18).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

By definition (2.10), the corresponding dimension $r_\alpha = \dim(\mathbf{U}_\alpha^{\min}(\mathbf{v}))$ is the $\alpha$th rank: $r_\alpha = \operatorname{rank}_\alpha(\mathbf{v})$. Altogether, we obtain the following statement.

**Proposition 7.1.** Any algebraic tensor $\mathbf{v} \in \mathbf{V}$ belongs to the set $\mathcal{H}_{\mathfrak{r}}$ with $\mathfrak{r} = (\operatorname{rank}_\alpha(\mathbf{v}))_{\alpha \in T_D}$. The appropriate subspaces are $\mathbf{U}_\alpha = \mathbf{U}_\alpha^{\min}(\mathbf{v})$.

*7.5. Coding of the subspaces*

So far, the set $\mathcal{H}_{\mathfrak{r}}$ has been described algebraically. To obtain the hierarchical format we have to describe its parametrization. There will be three types of data: (a) description of the subspaces $U_j$ $(1 \le j \le d)$, (b) description of the subspaces $\mathbf{U}_\alpha$ $(\alpha \in T_D \backslash \mathcal{L}(T_D))$, (c) description of $\mathbf{v}$.

*7.5.1. Subspaces $U_j$*

Each subspace $U_j$ is described by a basis[22] $\{b_1^{(j)}, \ldots\}$ as in the case of the tensor subspace format (see Section 6.2). The data size is

$$\sum_{j=1}^{d} r_j \dim(V_j) \le rdn, \quad \text{where } n := \max_j \dim(V_j) \quad \text{and} \quad r := \max_j r_j.$$

For $\alpha = \{j\}$ we identify the notations $b_i^{(j)}$ and $\mathbf{b}_i^{(\alpha)}$ as well as $r_\alpha$ and $r_j$.

*7.5.2. Subspaces $\mathbf{U}_\alpha$ for $\alpha \in T_D \backslash \mathcal{L}(T_D)$.*

Since $\alpha \notin \mathcal{L}(T_D)$, there are two sons $\alpha_1, \alpha_2$ and the corresponding subspaces $\mathbf{U}_{\alpha_1}, \mathbf{U}_{\alpha_2}$. By induction, $\mathbf{U}_{\alpha_1}$ and $\mathbf{U}_{\alpha_2}$ are already defined by means of the bases[22] $\{\mathbf{b}_i^{(\alpha_1)} : 1 \le i \le r_{\alpha_1}\}$ and $\{\mathbf{b}_i^{(\alpha_2)} : 1 \le i \le r_{\alpha_2}\}$. $\mathbf{U}_\alpha$ is fixed by a basis $\{\mathbf{b}_i^{(\alpha)} : 1 \le i \le r_\alpha\}$, but since $\mathbf{b}_i^{(\alpha)}$ are tensors of order $\#\alpha > 1$, it is not advisable to store these tensors explicitly. Instead we make use of the nestedness property $\mathbf{U}_\alpha \subset \mathbf{U}_{\alpha_1} \otimes \mathbf{U}_{\alpha_2}$. The tensors

$$\{\mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_j^{(\alpha_2)} : \ 1 \le i \le r_{\alpha_1}, \ 1 \le j \le r_{\alpha_2}\}$$

form a basis of $\mathbf{U}_{\alpha_1} \otimes \mathbf{U}_{\alpha_2}$. Hence, the tensor $\mathbf{b}_\ell^{(\alpha)} \in \mathbf{U}_\alpha$ has a representation

$$\mathbf{b}_\ell^{(\alpha)} = \sum_{i=1}^{r_{\alpha_1}} \sum_{j=1}^{r_{\alpha_2}} c_{ij}^{(\alpha,\ell)} \, \mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_j^{(\alpha_2)}. \tag{7.3}$$

It is sufficient to store the coefficients $c_{ij}^{(\alpha,\ell)}$, which form the *coefficient matrix*

$$C^{(\alpha,\ell)} = (c_{ij}^{(\alpha,\ell)})_{\substack{1 \le i \le r_{\alpha_1} \\ 1 \le j \le r_{\alpha_2}}} \in \mathbb{K}^{r_{\alpha_1} \times r_{\alpha_2}} \quad \text{for } 1 \le \ell \le r_\alpha.$$

[22] In general a *frame* is also possible, but a basis is the standard choice. In the case of a frame, the number of vectors $b_i^{(j)}$ may be larger than $r_j = \dim(U_j)$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The obvious data size of all matrices $\{C^{(\alpha,\ell)} : 1 \le \ell \le r_\alpha\}$ is $r_{\alpha_1} \cdot r_{\alpha_2} \cdot r_\alpha \le r^3$ for $r := \max_\alpha r_\alpha$. Since $\#(T_D \backslash \mathcal{L}(T_D)) = d - 1$, the overall storage cost is smaller than

$$(d-1)r^3 < dr^3.$$

**Remark 7.2.** To obtain a uniquely defined coefficient matrix, we have to introduce an ordering of the sons $\alpha_1, \alpha_2$ of vertex $\alpha$. A reverse ordering would replace the coefficient matrix $C^{(\alpha,\ell)}$ by its transpose $(C^{(\alpha,\ell)})^\mathsf{T}$.

*7.5.3. Tensor* $\mathbf{v}$

The tensor $\mathbf{v} \in \mathbf{V}$, which is to be represented, is of the form

$$\mathbf{v} = \sum_{i=1}^{r_D} c_i^{(D)} \mathbf{b}_i^{(D)}. \tag{7.4}$$

Since the basis $\{\mathbf{b}_i^{(D)}\}$ is (implicitly) given, it suffices to store the coefficient vector $c := (c_i^{(D)}) \in \mathbb{K}^{r_D}$. Note that usually $r_D = 1$ (see (7.2d)). The data size $r_D$ is negligible.

*7.5.4. Storage cost*

Adding the data sizes of the different parts, we obtain the storage cost

$$\sum_{j=1}^{d} r_j \dim(V_j) + \sum_{\alpha \in T_D \backslash \mathcal{L}(T_D)} r_{\alpha_1} \cdot r_{\alpha_2} \cdot r_\alpha + r_D \le drn + dr^3,$$

where $n := \max_j \dim(V_j)$, $r := \max_\alpha r_\alpha$, and $\alpha_1, \alpha_2$ are sons of $\alpha$.

*7.5.5. Transfer from other formats*

As an exercise, we consider the transfer from the $r$-term format into the hierarchical format.

Assume $\mathbf{v} = \sum_{\nu=1}^{r} v_\nu^{(1)} \otimes \cdots \otimes v_\nu^{(d)}$ and define $U_j$ as the span of the frame with $b_i^{(j)} := v_i^{(j)}$ for $1 \le i \le r$. This defines the subspaces $\mathbf{U}_\alpha$ with $r_\alpha := \dim(\mathbf{U}_\alpha) \le r$ for $\alpha \in \mathcal{L}(T_D)$.

By induction, the subspaces $\mathbf{U}_{\alpha_1}, \mathbf{U}_{\alpha_2}$ for the sons $\alpha_1, \alpha_2$ of the vertex $\alpha \in T_D \backslash (\{D\} \cup \mathcal{L}(T_D))$ are described by $\mathbf{b}_i^{(\alpha_1)}$ and $\mathbf{b}_i^{(\alpha_2)}$ $(1 \le i \le r)$. Set

$$\mathbf{U}_\alpha := \operatorname{span}\{\mathbf{b}_i^{(\alpha)} : 1 \le i \le r\}, \quad \text{where } \mathbf{b}_i^{(\alpha)} := \mathbf{b}_i^{(\alpha_1)} \otimes \mathbf{b}_i^{(\alpha_2)}.$$

This again implies $r_\alpha := \dim(\mathbf{U}_\alpha) \le r$. Finally, $\mathbf{U}_D := \operatorname{span}\{\mathbf{v}\}$ yields $r_D = 1$. This proves the following statement.

**Remark 7.3.** $\mathbf{v} \in \mathcal{R}_r$ belongs to $\mathcal{H}_\mathfrak{r}$ with $r_\alpha = r$ for $\alpha \neq D$ and $r_D = 1$. The choice of bases of $U_j \subset V_j$ from above leads to the coefficient matrices $C^{(\alpha,\ell)} = I$ ($r \times r$ identity matrix).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Further, the tensor subspace format $\mathcal{T}_{\mathbf{r}}$ can be embedded into the hierarchical format. Here, we choose $\mathbf{U}_\alpha := \mathbf{U}_{\alpha_1} \otimes \mathbf{U}_{\alpha_2}$ of maximal dimension. As a consequence, the cost is as high as for $\mathcal{T}_{\mathbf{r}}$. The intermediate dimensions are $r_\alpha = \prod_{j\in\alpha} r_j$.

For the embedding of the sparse grid format see Hackbusch and Kühn (2009).

### *7.6. Basis transformations*

#### *7.6.1. Transformation rules*

Since the bases $\{\mathbf{b}_\ell^{(\alpha)} : 1 \le \ell \le r_\alpha\}$ for $\alpha \in T_D \backslash \mathcal{L}(T_D)$ are given only implicitly via the coefficient matrices $C^{(\alpha,\ell)}$ (see (7.3)), a transformation of a basis must be expressed indirectly by a change of the coefficient matrices. A standard transformation of a frame or basis leaves the subspace unchanged. However, in later applications, truncations omitting certain basis vectors are also of interest. To obtain a compact notation, we introduce the tuple $\mathbf{C}_\alpha$ of coefficient matrices:

$$\mathbf{C}_\alpha := \left(C^{(\alpha,\ell)}\right)_{1\le\ell\le r_\alpha} \in \left(\mathbb{K}^{r_{\alpha_1}\times r_{\alpha_2}}\right)^{r_\alpha} \quad \text{for all } \alpha \in T_D\backslash\mathcal{L}(T_D).$$

For a vertex $\alpha \in T_D$ we consider the following 'old' and 'new' frames and subspaces:

$$\mathbf{B}_\alpha^{\mathrm{new}} = \left[\mathbf{b}_{1,\mathrm{new}}^{(\alpha)}, \dots, \mathbf{b}_{r_\alpha^{\mathrm{new}},\mathrm{new}}^{(\alpha)}\right], \quad \mathbf{U}_\alpha^{\mathrm{new}} = \mathrm{range}\{\mathbf{B}_\alpha^{\mathrm{new}}\},$$
$$\mathbf{B}_\alpha^{\mathrm{old}} = \left[\mathbf{b}_{1,\mathrm{old}}^{(\alpha)}, \dots, \mathbf{b}_{r_\alpha^{\mathrm{old}},\mathrm{old}}^{(\alpha)}\right], \quad \mathbf{U}_\alpha^{\mathrm{old}} = \mathrm{range}\{\mathbf{B}_\alpha^{\mathrm{old}}\},$$

The replacement $\mathbf{B}_\alpha^{\mathrm{old}} \mapsto \mathbf{B}_\alpha^{\mathrm{new}}$ creates new coefficient matrices $C_{\mathrm{new}}^{(\alpha,\ell)}$ (see Lemma 7.4). Moreover, if $\alpha \neq D$, the coefficient matrices $C_{\mathrm{old}}^{(\beta,\ell)}$ associated with the father $\beta \in T_D$ of $\alpha$ must be renewed to form $C_{\mathrm{new}}^{(\beta,\ell)}$, since these coefficients refer to $\mathbf{B}_\alpha^{\mathrm{new}}$ (see Lemma 7.5). If $\alpha = D$, the coefficient vector $c^{(D)}$ must be transformed instead (see Lemma 7.7).

We distinguish three different situations.

**Case A.** $\mathbf{B}_\alpha^{\mathrm{old}}$ and $\mathbf{B}_\alpha^{\mathrm{new}}$ generate the *same* subspace $\mathbf{U}_\alpha = \mathbf{U}_\alpha^{\mathrm{new}} = \mathbf{U}_\alpha^{\mathrm{old}}$. Then there are *transformation matrices* $T^{(\alpha)} \in \mathbb{K}^{r_\alpha^{\mathrm{new}}\times r_\alpha^{\mathrm{old}}}$, $S^{(\alpha)} \in \mathbb{K}^{r_\alpha^{\mathrm{old}}\times r_\alpha^{\mathrm{new}}}$ such that

$$\mathbf{B}_\alpha^{\mathrm{old}} = \mathbf{B}_\alpha^{\mathrm{new}} T^{(\alpha)}, \quad \textit{i.e.,} \quad \mathbf{b}_{j,\mathrm{old}}^{(\alpha)} = \sum_{k=1}^{r_\alpha^{\mathrm{new}}} T_{kj}^{(\alpha)}\, \mathbf{b}_{k,\mathrm{new}}^{(\alpha)} \quad (1 \le j \le r_\alpha^{\mathrm{old}}), \tag{7.5a}$$

$$\mathbf{B}_\alpha^{\mathrm{new}} = \mathbf{B}_\alpha^{\mathrm{old}} S^{(\alpha)}, \quad \textit{i.e.,} \quad \mathbf{b}_{k,\mathrm{new}}^{(\alpha)} = \sum_{j=1}^{r_\alpha^{\mathrm{old}}} S_{jk}^{(\alpha)}\, \mathbf{b}_{j,\mathrm{old}}^{(\alpha)} \quad (1 \le k \le r_\alpha^{\mathrm{new}}). \tag{7.5b}$$

In the standard case, $\mathbf{B}_\alpha^{\mathrm{old}}$ and $\mathbf{B}_\alpha^{\mathrm{new}}$ are bases. Then $r_\alpha^{\mathrm{old}} = r_\alpha^{\mathrm{new}} = \dim(\mathbf{U}_\alpha)$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

holds, and $T^{(\alpha)}$ and $S^{(\alpha)}$ are uniquely defined satisfying

$$S^{(\alpha)} := (T^{(\alpha)})^{-1}.$$

In the case of frames, the representation ranks $r_\alpha^{\rm old}, r_\alpha^{\rm new} \geq \dim(\mathbf{U}_\alpha)$ may be different, so that $T^{(\alpha)}$ and $S^{(\alpha)}$ are rectangular matrices. If $r_\alpha^{\rm new} > \dim(\mathbf{U}_\alpha)$ (resp. $r_\alpha^{\rm old} > \dim(\mathbf{U}_\alpha)$), the matrix $T^{(\alpha)}$ (resp. $S^{(\alpha)}$) satisfying (7.5a) (resp. (7.5b)) is not unique.

There may be reasons to change the subspace. In Case B we consider $\mathbf{U}_\alpha^{\rm new} \subset \mathbf{U}_\alpha^{\rm old}$, and in Case C the opposite inclusion $\mathbf{U}_\alpha^{\rm new} \supset \mathbf{U}_\alpha^{\rm old}$.

**Case B.** Assume $\mathbf{U}_\alpha^{\rm new} \subsetneqq \mathbf{U}_\alpha^{\rm old}$. This is a typical step when we truncate the tensor representation. Note that a transformation matrix $S^{(\alpha)}$ satisfying (7.5b) exists, whereas there is no $T^{(\alpha)}$ satisfying (7.5a).

**Case C.** Assume $\mathbf{U}_\alpha^{\rm new} \supsetneqq \mathbf{U}_\alpha^{\rm old}$. This occurs when we enrich $\mathbf{U}_\alpha^{\rm old}$ with further vectors. Then a transformation matrix $T^{(\alpha)}$ satisfying (7.5a) exists, but no $S^{(\alpha)}$ with (7.5b).

In Cases A and B, the transformation matrix $S^{(\alpha)}$ exists. Then (7.5b) proves the following result.

**Lemma 7.4.** If (7.5b) holds for $\alpha \in T_D \backslash \mathcal{L}(T_D)$, the new basis vectors $\mathbf{b}_{k,{\rm new}}^{(\alpha)}$ have coefficient matrices $C_{\rm new}^{(\alpha,k)}$ defined by

$$\mathbf{C}_\alpha^{\rm new} = \mathbf{C}_\alpha^{\rm old} S^{(\alpha)}, \quad \textit{i.e.}, \quad C_{\rm new}^{(\alpha,k)} = \sum_{j=1}^{r_\alpha^{\rm old}} S_{jk}^{(\alpha)} C_{\rm old}^{(\alpha,j)} \quad (1 \leq k \leq r_\alpha^{\rm new}). \tag{7.6}$$

The arithmetic cost of (7.6) is $2 r_\alpha^{\rm new} r_\alpha^{\rm old} r_{\alpha_1} r_{\alpha_2}$ (where $\alpha_1, \alpha_2$ are sons of $\alpha$).

Next, we consider the influence of a transformation upon the coefficient matrices of the father. Here, we rename the father vertex as $\alpha \in T_D \backslash \mathcal{L}(T_D)$ and assume that the bases $\mathbf{B}_{\alpha_1}^{\rm old}$ and $\mathbf{B}_{\alpha_2}^{\rm old}$ for at least one of the sons $\alpha_1, \alpha_2$ of $\alpha$ are changed into $\mathbf{B}_{\alpha_1}^{\rm new}$ and $\mathbf{B}_{\alpha_2}^{\rm new}$. If only one basis is changed, set $S^{(\alpha_i)} = T^{(\alpha_i)} = I$ for the other son. Since the transformation matrix $T^{(\alpha_i)}$ is used, the following lemma applies to Cases A and C.

**Lemma 7.5.** Let $\alpha_1, \alpha_2$ be the sons of $\alpha \in T_D \backslash \mathcal{L}(T_D)$. Basis transformations (7.5a) at the son vertices $\alpha_1, \alpha_2$, that is, $\mathbf{B}_{\alpha_i}^{\rm new} T^{(\alpha_i)} = \mathbf{B}_{\alpha_i}^{\rm old}$ $(i = 1, 2)$, lead to a transformation of the coefficients at vertex $\alpha$ by

$$C_{\rm old}^{(\alpha,\ell)} \mapsto C_{\rm new}^{(\alpha,\ell)} = T^{(\alpha_1)} C_{\rm old}^{(\alpha,\ell)} (T^{(\alpha_2)})^{\mathsf{T}} \quad \text{for } 1 \leq \ell \leq r_\alpha. \tag{7.7}$$

The arithmetic cost for (7.7) is $2 r_\alpha r_{\alpha_1}^{\rm old} r_{\alpha_2}^{\rm old} (r_{\alpha_1}^{\rm new} + r_{\alpha_2}^{\rm new})$. If the basis is changed only at $\alpha_1$ (*i.e.*, $T^{(\alpha_2)} = I$), the cost reduces to $2 r_\alpha r_{\alpha_1}^{\rm new} r_{\alpha_1}^{\rm old} r_{\alpha_2}^{\rm old}$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The next lemma uses the transformation matrix $S^{(\alpha_i)}$ from Cases A, B.

**Lemma 7.6.** Let $\alpha \in T_D \backslash \mathcal{L}(T_D)$ be a vertex with sons $\alpha_1, \alpha_2$. Assume that the coefficient matrices $C_{\rm old}^{(\alpha,\ell)}$ admit a decomposition

$$C_{\rm old}^{(\alpha,\ell)} = S^{(\alpha_1)}\, C_{\rm new}^{(\alpha,\ell)}\, (S^{(\alpha_2)})^{\mathsf{T}} \quad \text{for } 1 \le \ell \le r_\alpha.$$

Then $C_{\rm new}^{(\alpha,\ell)}$ are the coefficient matrices with respect to the new bases

$$\mathbf{B}_{\alpha_i}^{\rm new} := \mathbf{B}_{\alpha_i}^{\rm old}\, S^{(\alpha_i)} \quad (i = 1, 2)$$

at the son vertices (see (7.5b)). Since the frame $\mathbf{B}_{\alpha_i}^{\rm new}$ is not used in computations, no arithmetic operations accrue.

At the root $\alpha = D$, the tensor $\mathbf{v}$ is expressed by

$$\mathbf{v} = \sum_{i=1}^{r_D} c_i^{(D)} \mathbf{b}_i^{(D)} = \mathbf{B}_D c^{(D)}.$$

A change of the basis $\mathbf{B}_D$ is considered next.

**Lemma 7.7.** Assume a transformation by $\mathbf{B}_D^{\rm new}\, T^{(D)} = \mathbf{B}_D^{\rm old}$ (see (7.5a)). Then the coefficient vector $c_{\rm old}^{(D)}$ must be transformed into

$$c_{\rm new}^{(D)} := T^{(D)}\, c_{\rm old}^{(D)}.$$

The arithmetic cost is $2 r_D^{\rm old} r_D^{\rm new}$.

The proofs of the above lemmas are found in Hackbusch (2012, § 11.3.1.4).

*7.6.2. Orthonormalization*

The choice of orthonormal bases has many advantages, one being numerical stability and another being the truncation procedure.

The orthonormalization procedure starts at the leaves and proceeds to the root of $T_D$. At the leaves $\alpha = \{j\}$, the basis (frame) is given explicitly. For instance, the QR algorithm yields a new orthonormal basis together with the transformation[23] $\mathbf{B}_\alpha^{\rm new} T^{(\alpha)} = \mathbf{B}_\alpha^{\rm old}$ for $\alpha = \{j\} \in \mathcal{L}(T_D)$. If $\beta \in T_D$ is the father of $\alpha$, change $C_{\rm old}^{(\beta,\ell)}$ according to Lemma 7.6.

**Lemma 7.8.** Let $\alpha \in T_D \backslash \mathcal{L}(T_D)$. Assume that the bases at the son vertices $\alpha_1, \alpha_2$ are orthonormal. Then the scalar product $\langle \mathbf{b}_\ell^{(\alpha)}, \mathbf{b}_k^{(\alpha)} \rangle$ coincides with the Frobenius scalar product of the coefficient matrices:

$$\langle \mathbf{b}_\ell^{(\alpha)}, \mathbf{b}_k^{(\alpha)} \rangle = \langle C^{(\alpha,\ell)}, C^{(\alpha,k)} \rangle_{\mathsf{F}} := \sum_{i,j} C_{ij}^{(\alpha,\ell)}\, \overline{C_{ij}^{(\alpha,k)}}.$$

[23] $\mathbf{B}_\alpha^{\rm new}$ corresponds to $Q$ and $T^{(\alpha)}$ to $R$ in $\mathbf{B}_\alpha^{\rm old} = QR$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Under the assumption of the previous lemma, one has to orthonormalize the coefficient matrices with respect to the Frobenius scalar product, $\mathbf{C}_\alpha^{\rm new} T^{(\alpha)} = \mathbf{C}_\alpha^{\rm old}$, which is equivalent to $\mathbf{B}_\alpha^{\rm new} T^{(\alpha)} = \mathbf{B}_\alpha^{\rm old}$. As before, $\mathbf{C}_\beta^{\rm old}$ must be updated for the father $\beta$ of $\alpha$ according to Lemma 7.6. In the case of $\alpha = D$, we have to apply Lemma 7.7 instead.

The cost of orthonormalization is asymptotically $2dnr^2 + 4r^4(d-1)$ (Hackbusch 2012, Remark 11.32).

### *7.6.3. HOSVD*

In the case of the tensor subspace format, we have seen in Section 6.6 that the truncation is based on the singular value decomposition of $\mathcal{M}_j(\mathbf{v})$. Similarly, we now have to use the HOSVD basis from $\mathcal{M}_\alpha(\mathbf{v})$ for $\alpha \in T_D$. The HOSVD computation has to take care of the indirect description of $\mathbf{v}$. Furthermore, we do not want to compute the HOSVD basis $\{\mathbf{b}_\ell^{(\alpha)} : 1 \le \ell \le \operatorname{rank}_\alpha(\mathbf{v})\}$ itself, but rather the corresponding coefficient matrices coding this basis.

We recall that the HOSVD basis can be obtained by a diagonalization of $\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^{\mathsf H}$ (see Section 6.3). The key to the following computations is to consider the Hermitian matrices $E_\alpha = \big(e_{ij}^{(\alpha)}\big) \in \mathbb{K}^{r_\alpha \times r_\alpha}$, which satisfy

$$\mathcal{M}_\alpha(\mathbf{v})\mathcal{M}_\alpha(\mathbf{v})^{\mathsf H} = \sum_{i,j=1}^{r_\alpha} e_{ij}^{(\alpha)}\, \mathbf{b}_i^{(\alpha)} \otimes \overline{\mathbf{b}_j^{(\alpha)}} \in \mathbf{V}_\alpha \otimes \mathbf{V}_\alpha,$$

making use of the isomorphism (1.4) between matrices and tensors from $\mathbf{V}_\alpha \otimes \mathbf{V}_\alpha$.

**Theorem 7.9.** Assume that the bases $\{\mathbf{b}_i^{(\alpha)}\}$ are orthonormal[24] and that $r_D = 1$. For $\alpha = D$, the $1 \times 1$ matrix $E_D$ is equal to

$$E_D := c^{(D)}(c^{(D)})^{\mathsf H} \in \mathbb{K}^{r_D \times r_D} \tag{7.8}$$

with $c^{(D)}$ from (7.4). Let $\alpha_1, \alpha_2$ be the sons of $\alpha \in T_D \backslash \mathcal{L}(T_D)$. Given $E_\alpha \in \mathbb{K}^{r_\alpha \times r_\alpha}$, we determine $E_{\alpha_1}$ and $E_{\alpha_2}$ from

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

This result, which is proved in Hackbusch (2012, Theorem 11.38), yields a simple recursion proceeding from the root to the leaves. Note that only the explicitly known data $c^{(D)}$ and $C^{(\alpha,i)}$ are involved.

[24] $E_\alpha$ can be determined for general bases (Hackbusch 2012, Theorem 5.14), but it is more convenient to orthonormalize the bases first and then to apply this theorem.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

It can be shown that the bases $\{\mathbf{b}_i^{(\alpha)}\}$ are HOSVD bases if and only if $E_\alpha$ is diagonal:

$$E_\alpha = \operatorname{diag}\{(\sigma_1^{(\alpha)})^2, \dots, (\sigma_{r_\alpha}^{(\alpha)})^2\}.$$

In the following, $\Sigma_\alpha$ denotes the diagonal matrix containing the singular values $\sigma_i^{(\alpha)}$, $1 \le i \le r_\alpha := \operatorname{rank}_\alpha(\mathbf{v})$. For $\alpha = D$ (start of the recursion) we have

$$\sigma_1^{(D)} := \|\mathbf{v}\|,$$

which coincides with (7.8): $\Sigma_D^2 := E_D := c^{(D)}(c^{(D)})^{\mathsf{H}} \in \mathbb{K}^{1\times 1}$.

Assume that at vertex $\alpha \in T_D \backslash \mathcal{L}(T_D)$ a new basis $\mathbf{B}_\alpha = [\mathbf{b}_1^{(\alpha)}, \dots, \mathbf{b}_{r_\alpha}^{(\alpha)}]$ with $\Sigma_\alpha = \operatorname{diag}\{\sigma_1^{(\alpha)}, \dots, \sigma_{r_\alpha}^{(\alpha)}\}$ is already determined. The corresponding coefficient matrices are gathered in

$$\mathbf{C}^{(\alpha)} = (C^{(\alpha,\ell)})_{1\le\ell\le r_\alpha}$$

(note that in the previous step these matrices have been changed). Form the matrices

$$\mathbf{Z}_{\alpha_1} := \left[\sigma_1^{(\alpha)} C^{(\alpha,1)}, \sigma_2^{(\alpha)} C^{(\alpha,2)}, \dots, \sigma_{r_\alpha}^{(\alpha)} C^{(\alpha,r_\alpha)}\right] \in \mathbb{K}^{r_{\alpha_1}\times(r_\alpha r_{\alpha_2})}, \tag{7.9a}$$

$$\mathbf{Z}_{\alpha_2} := \begin{bmatrix} \sigma_1^{(\alpha)} C^{(\alpha,1)} \\ \vdots \\ \sigma_{r_\alpha}^{(\alpha)} C^{(\alpha,r_\alpha)} \end{bmatrix} \in \mathbb{K}^{(r_\alpha r_{\alpha_1})\times r_{\alpha_2}}, \tag{7.9b}$$

where $\alpha_1$ and $\alpha_2$ are the sons of $\alpha \in T_D$. Compute the left-sided singular value decomposition of $\mathbf{Z}_{\alpha_1}$ and the right-sided one of $\mathbf{Z}_{\alpha_2}$, that is, the matrices $U$ and $V$ in[25]

$$\mathbf{Z}_{\alpha_1} = U\Sigma_{\alpha_1}\hat{V}^{\mathsf{T}} \quad \text{and} \quad \mathbf{Z}_{\alpha_2} = \hat{U}\Sigma_{\alpha_2}V^{\mathsf{T}}. \tag{7.9c}$$

The factors $\hat{V}$ and $\hat{U}$ are not needed. Only the matrices

$$U \in \mathbb{K}^{r_{\alpha_1}\times r_{\alpha_1}^{\mathrm{HOSVD}}}, \quad \Sigma_{\alpha_1} = \operatorname{diag}\left\{\sigma_1^{(\alpha_1)}, \dots, \sigma_{r_{\alpha_1}^{\mathrm{HOSVD}}}^{(\alpha_1)}\right\} \in \mathbb{K}^{r_{\alpha_1}^{\mathrm{HOSVD}}\times r_{\alpha_1}^{\mathrm{HOSVD}}},$$

$$V \in \mathbb{K}^{r_{\alpha_2}\times r_{\alpha_2}^{\mathrm{HOSVD}}}, \quad \Sigma_{\alpha_2} = \operatorname{diag}\left\{\sigma_1^{(\alpha_2)}, \dots, \sigma_{r_{\alpha_2}^{\mathrm{HOSVD}}}^{(\alpha_2)}\right\} \in \mathbb{K}^{r_{\alpha_2}^{\mathrm{HOSVD}}\times r_{\alpha_2}^{\mathrm{HOSVD}}}$$

are of interest, where $r_{\alpha_i}^{\mathrm{HOSVD}} := \operatorname{rank}(\mathbf{Z}_{\alpha_i}) < r_{\alpha_i}$ may occur. These data are characterized by the diagonalizations of the matrices $E_{\alpha_1}$ and $E_{\alpha_2}$:

$$E_{\alpha_1} = \mathbf{Z}_{\alpha_1}\mathbf{Z}_{\alpha_1}^{\mathsf{H}} = \textstyle\sum_{\ell=1}^{r_\alpha} (\sigma_\ell^{(\alpha)})^2 C^{(\alpha,\ell)} C^{(\alpha,\ell)\mathsf{H}} = U\Sigma_{\alpha_1}^2 U^{\mathsf{H}},$$

$$E_{\alpha_2} = \mathbf{Z}_{\alpha_2}^{\mathsf{T}}\overline{\mathbf{Z}_{\alpha_2}} = \textstyle\sum_{\ell=1}^{r_\alpha} (\sigma_\ell^{(\alpha)})^2 C^{(\alpha,\ell)\mathsf{T}}\overline{C^{(\alpha,\ell)}} = V\Sigma_{\alpha_2}^2 V^{\mathsf{H}}.$$

[25] In the special case of $\alpha = D$, the matrices $\mathbf{Z}_{\alpha_1}$ and $\mathbf{Z}_{\alpha_2}$ coincide because of $r_D = 1$. Therefore, $\Sigma_{\alpha_1} = \Sigma_{\alpha_2}$ follows.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The inclusions $\mathrm{range}(C^{(\alpha,\ell)}) \subset \mathrm{range}(U)$ and $\mathrm{range}(C^{(\alpha,\ell)\mathsf{T}}) = \mathrm{range}(V)$ are valid by construction; hence, $C^{(\alpha,\ell)}$ allows a representation $C^{(\alpha,\ell)} = UC^{(\alpha,\ell)}_{\mathrm{HOSVD}}V^{\mathsf{T}}$. Since $U$ and $V$ are orthogonal matrices, the coefficient matrices at vertex $\alpha$ are transformed by

$$C^{(\alpha,\ell)} \mapsto C^{(\alpha,\ell)}_{\mathrm{HOSVD}} := U^{\mathsf{H}} C^{(\alpha,\ell)} \overline{V} \in \mathbb{K}^{r^{\mathrm{HOSVD}}_{\alpha_1} \times r^{\mathrm{HOSVD}}_{\alpha_2}} \quad (1 \le \ell \le r_\alpha). \tag{7.9d}$$

According to Lemmas 7.6 and 7.4, the bases and coefficient matrices at the son vertices $\alpha_1$, $\alpha_2$ transform as follows:

$$\begin{aligned} \mathbf{B}_{\alpha_1} \mapsto \mathbf{B}^{\mathrm{HOSVD}}_{\alpha_1} &:= \mathbf{B}_{\alpha_1} U \text{ and} \quad & \mathbf{B}_{\alpha_2} \mapsto \mathbf{B}^{\mathrm{HOSVD}}_{\alpha_2} &:= \mathbf{B}_{\alpha_2} V, \\ \mathbf{C}^{\mathrm{HOSVD}}_{\alpha_1} &:= \mathbf{C}_{\alpha_1} U \text{ and} & \mathbf{C}^{\mathrm{HOSVD}}_{\alpha_2} &:= \mathbf{C}_{\alpha_2} V. \end{aligned} \tag{7.9e}$$

Again, we write $\mathbf{B}_{\alpha_i}$ and $\mathbf{C}_{\alpha_i}$ instead of $\mathbf{B}^{\mathrm{HOSVD}}_{\alpha_i}$ and $\mathbf{C}^{\mathrm{HOSVD}}_{\alpha_i}$ and redefine $r_{\alpha_i}$ by

$$r_{\alpha_i} := r^{\mathrm{HOSVD}}_{\alpha_i}.$$

To compute the HOSVD bases at all vertices, the algorithm starts at the root and proceeds to the leaves. The underlying computational step at vertex $\alpha \in T_D \backslash \mathcal{L}(T_D)$ is abbreviated as follows:

> **procedure HOSVD**$(\alpha)$; (for $\alpha \in T_D \backslash \mathcal{L}(T_D)$ with sons $\alpha_1, \alpha_2$)
> transform $C^{(\alpha,\ell)}$ $(1 \le \ell \le r_\alpha)$ according to (7.9d);
> transform $C^{(\alpha_i,\ell)}$ $(1 \le \ell \le r_{\alpha_i};\ i = 1, 2)$ according to (7.9e);
> define $\Sigma_{\alpha_1} := \mathrm{diag}\{\sigma_1^{(\alpha_i)}, \ldots, \sigma_{r_{\alpha_i}}^{(\alpha_i)}\}$ for $i = 1, 2$
> according to (7.9c) with possibly new $r_{\alpha_1}$, $r_{\alpha_2}$;

The complete computation of HOSVD bases at all vertices of $T_D$ is performed by the call **HOSVD**$^*(D)$ of the recursive procedure **HOSVD**$^*$ defined by

> **procedure HOSVD**$^*(\alpha)$;
> if $\alpha \notin \mathcal{L}(T_D)$ then
> begin **HOSVD**$(\alpha)$; for all sons $\sigma$ of $\alpha$ do **HOSVD**$^*(\sigma)$ end; (7.10)

As a result of **HOSVD**$^*(D)$ we obtain the coefficient matrices $C^{(\alpha,\ell)}$ which define the HOSVD bases implicitly. The diagonal matrices $\Sigma_\alpha$ contain the singular values.

The total cost of **HOSVD**$^*(D)$ is asymptotically $3(d-2)r^4 + 2dr^2n$, where $r := \max_\alpha r_\alpha$ and $n := \max_j \dim(V_j)$.

Algorithm (7.10) uses a recursion over the tree $T_D$. Computations at the sons of a vertex are completely independent. This allows an easy parallelization, which reduces the computational time by a factor of $d/\log_2 d$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*7.7. Truncation by HOSVD*

The truncation is analogous to the procedure in Section 6.6. Apply the procedure **HOSVD**$^*(D)$ from (7.10) to install the HOSVD bases for the hierarchical representation of $\mathbf{v} \in \mathcal{H}_{\mathfrak{s}}$. Let $\mathfrak{r} = (r_\alpha)_{\alpha \in T_D}$ be another rank tuple with $\mathfrak{r} \le \mathfrak{s}$. Then the truncation of $\mathbf{v} \in \mathcal{H}_{\mathfrak{s}}$ to $\mathbf{u}_{\mathsf{HOSVD}} \in \mathcal{H}_{\mathfrak{r}}$ is obtained by omitting all contributions connected with basis vectors $\mathbf{b}_i^{(\alpha)}$ for $i > r_\alpha$ (as in (6.8)). Since in the hierarchical representation the basis vectors are replaced by the coefficient matrices, the following actions are required:

$$\begin{aligned}
&\text{for all } \alpha \in T_D \text{ do}\\
&\quad \text{for } \ell > r_\alpha \text{ delete } C^{(\alpha,\ell)};\\
&\quad \text{for } \ell \le r_\alpha \text{ replace } C^{(\alpha,\ell)} \in \mathbb{K}^{s_{\alpha_1} \times s_{\alpha_2}} \text{ by the}\\
&\qquad \text{submatrix } \big(c_{i,j}^{(\alpha,\ell)}\big)_{1 \le i \le r_{\alpha_1},\, 1 \le j \le r_{\alpha_2}} \in \mathbb{K}^{r_{\alpha_1} \times r_{\alpha_2}};\\
&\text{for } j = 1 \text{ to } d \text{ do delete bases vectors } b_i^{(j)} \text{ for } i > r_j;
\end{aligned} \tag{7.11}$$

Note that the remaining coefficients are unchanged; only those referring to $\{\mathbf{b}_\ell^{(\alpha)} : r+1 \le \ell \le s_\alpha\}$ are deleted by reducing the size of $\mathbf{C}_\alpha$.

After an HOSVD truncation the orthonormality of the bases is lost. This is easily seen from Lemma 7.8. The truncation maps the matrices $C^{(\alpha,\ell)}$ into submatrices. Frobenius orthonormality of the original matrices is not inherited by the submatrices.

The error analysis yields a result quite similar to Theorem 6.9. The increased number $2d-3$ (compared with $d$ in (6.9)) is due to the fact that there are more vertices in $T_D$, which gives rise to more projections.

**Theorem 7.10.** The approximation $\mathbf{u}_{\mathrm{HOSVD}} \in \mathcal{H}_{\mathfrak{r}}$ defined above is quasi-optimal:

$$\|\mathbf{v} - \mathbf{u}_{\mathrm{HOSVD}}\| \le \sqrt{\sum_\alpha \sum_{i > r_\alpha} (\sigma_i^{(\alpha)})^2} \le \sqrt{2d-3}\,\|\mathbf{v} - \mathbf{u}_{\mathrm{best}}\|. \tag{7.12}$$

Here $\sigma_i^{(\alpha)}$ are the singular values of $\mathcal{M}_\alpha(\mathbf{v})$. The sum $\sum_\alpha$ is taken over all vertices $\alpha \in T_D \backslash \{D\}$ except that only one son of $D$ is involved.

Again there is a sequential variant. Although the established estimate (*e.g.*, in Hackbusch 2012, Theorem 11.61) looks worse than (7.12), in practice the following approximation $\tilde{\mathbf{u}}_{\mathrm{HOSVD}}$ is better than $\mathbf{u}_{\mathrm{HOSVD}}$ from Theorem 7.10.

In the previous version, the HOSVD bases are determined at all vertices before the truncation starts. Now the HOSVD basis computation and reduction is combined in a single vertex before proceeding to the sons. The call of **HOSVD-TrSeq**$(D, \mathfrak{r})$ performs the computation starting from the

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

root $D$ and proceeding to the leaves (Hackbusch and Kühn 2009):

**procedure HOSVD-TrSeq**$(\alpha, \mathfrak{r})$;
if $\alpha \notin \mathcal{L}(T_D)$ then
begin **HOSVD**$(\alpha)$; let
  **REDUCE**$(\alpha_1, r_{\alpha_1})$; **REDUCE**$(\alpha_2, r_{\alpha_2})$;
  **HOSVD-TrSeq**$(\alpha_1, \mathfrak{r})$; **HOSVD-TrSeq**$(\alpha_2, \mathfrak{r})$
end;

Here, $\alpha_1, \alpha_2$ are the sons of $\alpha$. **HOSVD**$(\alpha)$ determines the HOSVD basis at $\alpha \in T_D$ (assuming that the previous bases are already orthonormal). The call of **REDUCE**$(\beta, r_\beta)$ deletes all connections to the basis vectors $\mathbf{b}_i^{(\beta)}$ for $i > r_\beta$ (*i.e.*, the coefficient matrices are deleted or the corresponding rows or columns are omitted). The advantage of this variant is that the HOSVD computation in **HOSVD**$(\alpha)$ always refers to the tensor actually modified. One may hope that the use of the actual HOSVD is more efficient. In any case, it is cheaper than the global version (7.11) since the size of the bases is already reduced.

Above we have fixed the destination rank $\mathfrak{r}$. Instead one can prescribe a tolerance $\varepsilon > 0$ and choose $r_\alpha$ adaptively. The statement of Remark 6.11 also holds for the HOSVD truncations in the hierarchical format.

### *7.8. Closedness of $\mathcal{H}_\mathfrak{r}$ and existence of best approximations*

The statements of Section 6.5 can be repeated without modification. The proofs follow the same lines.

**Theorem 7.11.** Assume (6.5). Then $\mathcal{H}_\mathfrak{r}$ is weakly closed for any $\mathfrak{r} \in \mathbb{N}_0^{T_D}$.

**Theorem 7.12.** Assume that $(\mathbf{V}, \|\cdot\|)$ is reflexive and satisfies (6.5). Then, for any $\mathbf{v} \in \mathbf{V}$, the minimization problem

$$\inf_{\mathbf{u} \in \mathcal{H}_\mathfrak{r}} \|\mathbf{v} - \mathbf{u}\| = \|\mathbf{v} - \mathbf{u}_{\mathsf{best}}\|$$

has a solution $\mathbf{u}_{\mathsf{best}} \in \mathcal{H}_\mathfrak{r}$.

Theorem 7.11 ensures that instabilities as in Section 5.4 do not occur. The stable behaviour can be quantified by the following sensitivity analysis.

### *7.9. Sensitivity analysis*

The hierarchical representation describes a tensor $\mathbf{v} \in \mathcal{H}_\mathfrak{r}$ by means of bases $\{b_i^{(j)}\}$ at the leaves, coefficient matrices $C^{(\alpha,\ell)}$ at vertices $\alpha \in T_D \backslash \mathcal{L}(T_D)$, and a factor $c_1^D$. We may ask how perturbations of these quantities influence the tensor $\mathbf{v}$. An analysis of the orthonormal case is given in Hackbusch (2012, § 11.3.4.2). Here we concentrate on the case where all bases are

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

HOSVD bases. Fix some $\tau \in T_D$ and consider a perturbation[26] $\delta\mathbf{B}_\tau$ of the basis $\mathbf{B}_\tau$ described by $\delta\mathbf{B}_\tau = [\delta\mathbf{b}_1^{(\tau)}, \ldots, \delta\mathbf{b}_{r_\tau}^{(\tau)}]$. The singular values $\sigma_i^{(\tau)}$ from $\Sigma_\tau$ are used for the formulation of the error:

$$\varepsilon := \sqrt{\sum_{i=1}^{r_\tau} \big(\sigma_i^{(\tau)} \|\delta\mathbf{b}_i^{(\tau)}\|\big)^2} \quad (\tau \in T_D). \tag{7.13}$$

The weighting by $\sigma_i^{(\tau)}$ in (7.13) allows for the fact that basis vectors corresponding to small singular values are of less importance.

**Theorem 7.13.** Under the assumptions from above, $\delta\mathbf{b}_i^{(\tau)}$ causes an error $\delta\mathbf{v}$ of $\mathbf{v}$, which can be estimated by

$$\|\delta\mathbf{v}\| \le \varepsilon.$$

The proof and a generalization to so-called weak HOSVD bases are given in Hackbusch (2012, Theorem 11.53).

## 8. Operations in the hierarchical format

In Section 5.2, tensor operations are discussed for the $r$-term format. Now we have to explain how to perform these operations for tensors given in the hierarchical format. An auxiliary procedure is explained in Section 8.1.

The operation counts use the bounds

$$r := \max_\alpha r_\alpha \quad \text{and} \quad n := \max_j \dim(V_j).$$

Proofs, in particular for the cost of the operations, can be found in Hackbusch (2012, Chap. 13).

### *8.1. Joining two hierarchical tensor representation systems*

Let $\mathbf{v}' \in \mathcal{H}_{\mathfrak{r}'}$ and $\mathbf{v}'' \in \mathcal{H}_{\mathfrak{r}''}$ be two tensors represented by different bases

$$(\mathbf{B}'_\alpha)_{\alpha\in\mathcal{L}(T_D)}, \quad (\mathbf{B}''_\alpha)_{\alpha\in\mathcal{L}(T_D)}, \quad (\mathbf{C}'_\alpha)_{\alpha\in T_D\setminus\mathcal{L}(T_D)}, \quad (\mathbf{C}''_\alpha)_{\alpha\in T_D\setminus\mathcal{L}(T_D)},$$

but with a common dimension partition tree $T_D$. Let

$$\{\mathbf{U}'_\alpha\}_{\alpha\in T_D} \quad \text{and} \quad \{\mathbf{U}''_\alpha\}_{\alpha\in T_D}$$

be the different subspaces spanned by

$$(\mathbf{B}'_\alpha)_{\alpha\in\mathcal{L}(T_D)} \quad \text{and} \quad (\mathbf{B}''_\alpha)_{\alpha\in\mathcal{L}(T_D)},$$

respectively. We want to construct a hierarchical system $\mathcal{H}_{\mathfrak{r}}$ which allows a simultaneous representation of $\mathbf{v}'$ and $\mathbf{v}''$. The desired subspaces $\{\mathbf{U}_\alpha\}_{\alpha\in T_D}$

[26] A typical perturbation is $\delta\mathbf{b}_i^{(\tau)} = \mathbf{b}_i^{(\tau)}$ (*i.e.*, $\mathbf{b}_i^{(\tau)} \mapsto \mathbf{b}_i^{(\tau)} - \delta\mathbf{b}_i^{(\tau)} = 0$).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

may be defined by

$$\mathbf{U}_\alpha := \mathbf{U}'_\alpha + \mathbf{U}''_\alpha \quad \text{for } \alpha \in T_D.$$

Then, $\mathcal{H}_\mathfrak{r}$ represents all tensors of $\mathbf{U}_D = \mathbf{U}'_D + \mathbf{U}''_D$, that is, the tensors $\mathbf{v}', \mathbf{v}''$ from above belong to the new, common basis systems. The proposed subspaces $\mathbf{U}_\alpha$ make sense, since they satisfy the nestedness condition (see (7.2a)):

$$\mathbf{U}'_\alpha + \mathbf{U}''_\alpha \subset (\mathbf{U}'_{\alpha_1} + \mathbf{U}''_{\alpha_1}) \otimes (\mathbf{U}'_{\alpha_2} + \mathbf{U}''_{\alpha_2}).$$

The least requirement is that

$$\mathbf{B}'_\alpha = \big[\mathbf{b}_1^{\prime(\alpha)}, \dots, \mathbf{b}_{r'_\alpha}^{\prime(\alpha)}\big] \in (\mathbf{U}'_\alpha)^{r'_\alpha} \quad \text{and} \quad \mathbf{B}''_\alpha = \big[\mathbf{b}_1^{\prime\prime(\alpha)}, \dots, \mathbf{b}_{r''_\alpha}^{\prime\prime(\alpha)}\big] \in (\mathbf{U}''_\alpha)^{r''_\alpha}$$

are frames spanning the respective subspaces $\mathbf{U}'_\alpha$ and $\mathbf{U}''_\alpha$. Since no linear independence is required, the simple definition

$$\mathbf{B}_\alpha := \big[\mathbf{b}_1^{\prime(\alpha)}, \mathbf{b}_2^{\prime(\alpha)}, \dots, \mathbf{b}_{r'_\alpha}^{\prime(\alpha)}, \mathbf{b}_1^{\prime\prime(\alpha)}, \mathbf{b}_2^{\prime\prime(\alpha)}, \dots, \mathbf{b}_{r''_\alpha}^{\prime\prime(\alpha)}\big]$$

yields a frame generating $\mathbf{U}_\alpha := \mathbf{U}'_\alpha + \mathbf{U}''_\alpha$. The cardinality $r_\alpha := r'_\alpha + r''_\alpha$ is increased, but an advantage is the easy construction of the coefficients. Consider a vertex $\alpha \in T_D \backslash \mathcal{L}(T_D)$ with sons $\alpha_1, \alpha_2$. The coefficient matrix $C'^{(\alpha,\ell)}$ representing $\mathbf{b}_\ell^{\prime(\alpha)}$ and the matrix $C''^{(\alpha,\ell)}$ representing $\mathbf{b}_\ell^{\prime\prime(\alpha)}$ yield the block diagonal matrix

$$C^{(\alpha,\ell)} := \begin{bmatrix} C'^{(\alpha,\ell)} & 0 \\ 0 & C''^{(\alpha,\ell)} \end{bmatrix},$$

representing the new columns of $\mathbf{B}_\alpha$ by those from $\mathbf{B}_{\alpha_1}$ and $\mathbf{B}_{\alpha_2}$.

If $\mathbf{v}' \in \mathbf{U}'_D$ is coded by the coefficient vector $c'^{(D)} \in \mathbb{K}^{r'_D}$, the new coefficient is $\left[\begin{smallmatrix} c'^{(D)} \\ 0 \end{smallmatrix}\right]$, while $\mathbf{v}'' \in \mathbf{U}''_D$ coded by $c''^{(D)} \in \mathbb{K}^{r''_D}$ is expressed by the coefficient vector $\left[\begin{smallmatrix} 0 \\ c''^{(D)} \end{smallmatrix}\right]$.

Because we are using frames, $r_\alpha$ may be larger than $\dim(\mathbf{U}_\alpha)$. Therefore, it is better to orthonormalize according to Section 7.6.2. Note that in this case $r_D = 2$ occurs. The joining of frames requires only a rearrangement of data but no arithmetic operation. The orthonormalization costs about $2dnr^2 + 4r^4(d-1)$ operations.

### *8.2. Addition*

The first trivial case is characterized by $\mathbf{v}', \mathbf{v}'' \in \mathcal{H}_\mathfrak{r}$, that is, both tensors have a common representation $\mathbf{v}' = \sum_{i=1}^{r_D} c'_i \mathbf{b}_i^{(D)}$ and $\mathbf{v}'' = \sum_{i=1}^{r_D} c''_i \mathbf{b}_i^{(D)}$. Obviously, $\mathbf{v} = \sum_{i=1}^{r_D} (c'_i + c''_i)\mathbf{b}_i^{(D)}$ requires only $r_D$ additions.

Otherwise, apply the procedure from Section 8.1 to obtain a common representation. In the latter case, the addition increases the ranks $r_\alpha$ and a truncation is advisable.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*8.3. Scalar product*

We assume that $\mathbf{V} = \bigotimes_{j=1}^{d} V_j$ is equipped with the canonical scalar product (see Section 3.4.1). For $\mathbf{V}_\alpha = \bigotimes_{j\in\alpha} V_j$, the canonical scalar product is denoted by $\langle\cdot,\cdot\rangle_\alpha$. For $\alpha = \{j\}$ we retain the scalar product[27] $\langle\cdot,\cdot\rangle_{\{j\}} = \langle\cdot,\cdot\rangle_j$ of $V_j$.

We start with the scalar product $\langle \mathbf{u}, \mathbf{v}\rangle$ of $\mathbf{u} \in \mathcal{H}_\mathfrak{r}$ and an elementary tensor $\mathbf{v} = \bigotimes_{j=1}^{d} v^{(j)}$. Define $\mathbf{v}^{(\alpha)} := \bigotimes_{j\in\alpha} v^{(j)}$ and use the recursion

$$\big\langle \mathbf{b}_\ell^{(\alpha)}, \mathbf{v}^{(\alpha)}\big\rangle_\alpha = \sum_{i=1}^{r_{\alpha_1}}\sum_{j=1}^{r_{\alpha_2}} c_{ij}^{(\alpha,\ell)} \big\langle \mathbf{b}_i^{(\alpha_1)}, \mathbf{v}^{(\alpha_1)}\big\rangle_{\alpha_1} \big\langle \mathbf{b}_j^{(\alpha_2)}, \mathbf{v}^{(\alpha_2)}\big\rangle_{\alpha_2}, \quad \text{(see (7.3))}$$

where $\alpha_1, \alpha_2$ are the sons of $\alpha$. The final result is

$$\sum_{i=1}^{r_D} c_i^D \big\langle \mathbf{b}_i^{(D)}, \mathbf{v}\big\rangle.$$

The recursion has an asymptotic cost of $2(d-1)r^3 + 2rn$ operations.

In the case of the scalar product $\langle \mathbf{u}, \mathbf{v}\rangle$ of two tensors in hierarchical format, we have to distinguish two cases. If $\mathbf{u}, \mathbf{v} \in \mathcal{H}_\mathfrak{r}$ have a common representation involving the same orthonormal bases, the immediate result is

$$\langle \mathbf{u}', \mathbf{u}''\rangle = \big\langle c'^{(D)}, c''^{(D)}\big\rangle.$$

Otherwise, we have to determine all scalar products

$$\beta_{\ell k}^{(\alpha)} := \big\langle \mathbf{b}_\ell'^{(\alpha)}, \mathbf{b}_k''^{(\alpha)}\big\rangle_\alpha$$

from the recursion

$$\beta_{\ell k}^{(\alpha)} = \sum_{i=1}^{r'_{\alpha_1}}\sum_{j=1}^{r'_{\alpha_2}}\sum_{m=1}^{r''_{\alpha_1}}\sum_{n=1}^{r''_{\alpha_2}} c_{ij}'^{(\alpha,\ell)}\overline{c_{mn}''^{(\alpha,k)}}\beta_{im}^{(\alpha_1)}\beta_{jn}^{(\alpha_2)}.$$

Note that $\beta_{\ell k}^{(\alpha)}$ can be explicitly determined at the leaves $\alpha \in \mathcal{L}(T_D)$. Then

$$\langle \mathbf{u}', \mathbf{u}''\rangle = \sum_{\ell=1}^{r'_D}\sum_{k=1}^{r''_D} c_\ell'^{(D)}\overline{c_k''^{(D)}}\beta_{\ell k}^{(D)}$$

is the final result. The computational work is about $2dr^2n + 6(d-1)r^4 + 2r^2$.

Another possibility is to construct a common orthonormal representation for $\mathbf{u}', \mathbf{u}''$, so that the first case applies. In general, the latter variant is a bit more expensive, but may be advantageous in special cases.

[27] Note that the scalar product $\langle\cdot,\cdot\rangle_j$ is not necessarily the Euclidean one, but the indicated cost corresponds to the case of Euclidean scalar products.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*8.4. General binary operation*

Here we consider tensor spaces

$$\mathbf{V} = \bigotimes_{j=1}^{d} V_j, \quad \mathbf{W} = \bigotimes_{j=1}^{d} W_j, \quad \mathbf{X} = \bigotimes_{j=1}^{d} X_j,$$

and any bilinear operation

$$\boxdot : \mathbf{V} \times \mathbf{W} \to \mathbf{X},$$

which satisfies[28]

$$\left(\bigotimes_{j=1}^{d} v^{(j)}\right) \boxdot \left(\bigotimes_{j=1}^{d} w^{(j)}\right) = \bigotimes_{j=1}^{d} \big(v^{(j)} \boxdot w^{(j)}\big), \quad v^{(j)} \boxdot w^{(j)} \in X_j.$$

Let $\mathbf{v} \in \mathcal{H}_{\mathfrak{r}'}(\mathbf{V})$ and $\mathbf{w} \in \mathcal{H}_{\mathfrak{r}''}(\mathbf{W})$ be two tensors described in two different hierarchical formats, but with the same dimension partition tree $T_D$. Since

$$\mathbf{v} = \sum_{\ell} c'^{(D)}_{\ell} \mathbf{b}'^{(D)}_{\ell} \quad \text{and} \quad \mathbf{w} = \sum_{k} c''^{(D)}_{k} \mathbf{b}''^{(D)}_{k},$$

we start from

$$\mathbf{v} \boxdot \mathbf{w} = \sum_{\ell=1}^{r'_D} \sum_{k=1}^{r''_D} c'^{(D)}_{\ell} c''^{(D)}_{k}\, \mathbf{b}'^{(D)}_{\ell} \boxdot \mathbf{b}''^{(D)}_{k}$$

and use the recursion

$$\mathbf{b}'^{(\alpha)}_{\ell} \boxdot \mathbf{b}''^{(\alpha)}_{k} = \sum_{i=1}^{r'_{\alpha_1}} \sum_{j=1}^{r'_{\alpha_2}} \sum_{m=1}^{r''_{\alpha_1}} \sum_{n=1}^{r''_{\alpha_2}} c'^{(\alpha,\ell)}_{ij} c''^{(\alpha,k)}_{mn} \big(\mathbf{b}'^{(\alpha_1)}_{i} \boxdot \mathbf{b}''^{(\alpha_1)}_{m}\big) \otimes \big(\mathbf{b}'^{(\alpha_2)}_{j} \boxdot \mathbf{b}''^{(\alpha_2)}_{n}\big)$$

(see (7.3)), which terminates at the leaves of $T_D$ .

In the first approach we accept the *frame* $\mathfrak{b}^{(\alpha)}$ consisting of the $r'_\alpha r''_\alpha$ vectors $\mathbf{b}'^{(\alpha)}_{\ell} \boxdot \mathbf{b}''^{(\alpha)}_{k}$ $(1 \le \ell \le r'_\alpha,\ 1 \le k \le r''_\alpha)$ describing the subspace $\mathbf{U}_\alpha$. Denote the elements of $\mathfrak{b}^{(\alpha)}$ by $\mathbf{b}^{(\alpha)}_m$ with $m \in J_\alpha := \{1, \dots, r'_\alpha\} \times \{1, \dots, r''_\alpha\}$, that is, $\mathbf{b}^{(\alpha)}_m = \mathbf{b}'^{(\alpha)}_{\ell} \boxdot \mathbf{b}''^{(\alpha)}_{k}$ if $m = (\ell, k)$. Then we obtain

$$\mathbf{b}^{(\alpha)}_m = \sum_{p \in J_{\alpha_1}} \sum_{q \in J_{\alpha_2}} c^{(\alpha,m)}_{pq} \mathbf{b}^{(\alpha_1)}_p \otimes \mathbf{b}^{(\alpha_2)}_q \quad \text{with } c^{(\alpha,m)}_{pq} := c'^{(\alpha,\ell)}_{p_1 q_1} c''^{(\alpha,k)}_{p_2 q_2}$$

for $p = (p_1, p_2) \in J_{\alpha_1}, q = (q_1, q_2) \in J_{\alpha_2}$. The new coefficient matrix $C^{(\alpha,m)}$ is the Kronecker product

$$C^{(\alpha,m)} = C'^{(\alpha,\ell)} \otimes C''^{(\alpha,k)} \quad \text{for } m = (\ell, k).$$

[28] The map $\boxdot_j = \boxdot : V_j \times W_j \to X_j$ on the right-hand side is also denoted by $\boxdot$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The final result,

$$\mathbf{v} \boxdot \mathbf{w} = \sum_{m \in J_D} c_m^{(D)} \mathbf{b}_m^{(D)} \quad \text{with} \quad c_m^{(D)} := c_{m_1}^{\prime(D)} c_{m_2}^{\prime\prime(D)} \quad \text{for } m = (m_1, m_2),$$

is represented in $\mathcal{H}_{\mathfrak{r}}(\mathbf{X})$ with representation ranks $r_\alpha := r'_\alpha r''_\alpha$. Altogether, the computational cost amounts to $dr^2 N_{\boxdot} + (d-1)r^4 + 1$. Here, $N_{\boxdot}$ is a bound for the cost of $u \boxdot v$ for any $u, v \in V_j$.

An additional orthonormalization of the frame requires $2dnr^4 + 4dr^8$ operations. Because of the strongly increased ranks, a truncation is advisable.

### *8.5. Hadamard product, convolution, matrix–matrix multiplication, matrix–vector multiplication*

All these operations are of the form $\mathbf{v} \boxdot \mathbf{w}$, as discussed above. Therefore the same recursion is used. The concrete operations $u \boxdot v$ are to be applied only at the leaves (*i.e.*, for $u, v \in V_j$).

### *8.6. Entry-wise evaluation*

For $\alpha \subset D = \{1, \dots, d\}$, the index $\mathbf{i}_\alpha$ belongs to $\mathbf{I}_\alpha = \times_{j \in \alpha} I_j$. The evaluation of

$$\beta_\ell^{(\alpha)} := \mathbf{b}_\ell^{(\alpha)}[\mathbf{i}_\alpha] = \sum_{i=1}^{r_{\alpha_1}} \sum_{j=1}^{r_{\alpha_2}} c_{ij}^{(\alpha,\ell)} \, \mathbf{b}_i^{(\alpha_1)}[\mathbf{i}_{\alpha_1}] \mathbf{b}_j^{(\alpha_2)}[\mathbf{i}_{\alpha_2}]$$

of the basis vector $\mathbf{b}_\ell^{(\alpha)}$ is performed recursively from the leaves to the root:

procedure $\mathbf{eval}^*(\alpha, \mathbf{i})$;
for $\ell := 1$ to $r_\alpha$ do if $\alpha = \{j\}$ then $\beta_\ell^{(\alpha)} := b_\ell^{(j)}[i_j]$ else
begin $\mathbf{eval}^*(\alpha_1, \mathbf{i})$; $\mathbf{eval}^*(\alpha_2, \mathbf{i})$; $\beta_\ell^{(\alpha)} := \sum_{i=1}^{r_{\alpha_1}} \sum_{j=1}^{r_{\alpha_2}} c_{ij}^{(\alpha,\ell)} \beta_i^{(\alpha_1)} \beta_j^{(\alpha_2)}$ end;

Here, $\alpha_1, \alpha_2$ are the sons of $\alpha$. Call $\mathbf{eval}^*(D, \mathbf{i})$. Then,

$$\mathbf{v}[\mathbf{i}] = \sum_{\ell=1}^{r_D} c_\ell^{(D)} \beta_\ell^{(D)}$$

is the final result. The computational cost is $2dr^3$.

The same recursion applies to other functionals $\bigotimes_{j=1}^d \varphi^{(j)}$ with $\varphi^{(j)} \in V_j'$.

### *8.7. Functions of tensors, fixed-point iterations*

In Section 5.9 we discussed functions $f(A)$ of matrices; see also Higham (2008). Since a matrix $A$ may be a Kronecker tensor, this is also an example of a function of a tensor. In particular cases there are fixed-point iterations

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

converging to $f(A)$. In the case of $A^{-1}$, the Newton method yields the iteration[29]

$$X_{m+1} := 2X_m - X_m A X_m, \tag{8.1}$$

which shows locally quadratic convergence. A possible starting value is $X_0 := I$. Equation (8.1) is an example of a fixed-point iteration.

If the desired tensor satisfies $X^* = \Phi(X^*)$, the sequence

$$X_{m+1} := \Phi(X_m)$$

converges to $X^*$ if $\Phi$ is contractive. Assuming that the evaluation of $\Phi$ involves only the operations studied in this section, $\Phi(X_m)$ is available. However, since the operations cause an increase of the representation ranks, the next iteration must be preceded by a truncation:

$$\tilde{X}_{m+1} := T(\Phi(X_m)) \quad (T\text{: truncation}).$$

The resulting iteration is called *truncated iteration* and has been studied in Hackbusch, Khoromskij and Tyrtyshnikov (2008); see also Hackbusch (2009, § 14.3.2). In essence, the error decreases as in the original iteration until the iterates reach an $X^*$ neighbourhood of the size of the truncation error ($X^*$ is the exact solution).

A useful and seemingly simple (nonlinear) function is the maximum of a tensor $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^{d} \mathbb{R}^{I_j} \cong \mathbb{R}^{\mathbf{I}}$ ($\mathbf{I} = \times_{j=1}^{d} I_j$):

$$\max(\mathbf{v}) := \max\{\mathbf{v}_\mathbf{i} : \mathbf{i} \in \mathbf{I}\}.$$

Since $\min(\mathbf{v}) = -\max(-\mathbf{v})$, this function allows us to determine the maximum norm $\|\mathbf{v}\|_\infty$ of a tensor. The implementation is trivial for an elementary tensor:

$$\max\left(\bigotimes_{j=1}^{d} v^{(j)}\right) = \prod_{j=1}^{d} \max(v^{(j)}).$$

However, the implementation for general tensors is not straightforward. A possible approach is based on the reformulation as an eigenvalue problem: see Espig, Hackbusch, Litvinenko, Matthies and Zander (2013, § 4.1). The tensor $\mathbf{v} \in \mathbf{V}$ corresponds to a multiplication operator $\mathbf{G}(\mathbf{v})$ defined by $\mathbf{G}(\mathbf{v})(\mathbf{f}) = \mathbf{v} \odot \mathbf{f}$. Let $\mathbf{I}^* := \{\mathbf{i} \in \mathbf{I} : \max(\mathbf{v}) = \mathbf{v}_\mathbf{i}\}$ be the index subset where the maximum is attained. Then the eigenvalue problem

$$\mathbf{G}(\mathbf{v})\mathbf{u} = \lambda \mathbf{u} \quad (0 \neq \mathbf{u} \in \mathbf{V})$$

has the maximal eigenvalue $\lambda = \max(\mathbf{v})$. The eigenspace consists of all vectors $\mathbf{u}$ with support in $\mathbf{I}^*$. In particular, if $\mathbf{I}^* = \{\mathbf{i}^*\}$ is a singleton, the

[29] For a suitable modification see Oseledets and Tyrtyshnikov (2005) and Oseledets, Savostyanov and Tyrtyshnikov (2009).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

maximal eigenvalue is a simple one and the eigenvector is a multiple of the unit vector $\mathbf{e}^{(\mathbf{i}^*)}$, which has tensor rank 1. Using the simple vector iteration or more advanced methods, we can determine not only the maximum $\max(\mathbf{v})$ but also the corresponding index.

## 9. Matrix product format

Matrix product states (MPS) have been introduced in quantum physics: see, for example, Affleck, Kennedy, Lieb and Tasaki (1987) and Verstraete and Cirac (2006). Since the re-invention by Oseledets and Tyrtyshnikov (2009*a*) and Oseledets (2011*b*), this representation has been called the *TT format*.[30] The TT representation is a special form of the hierarchical format[31] (see Section 9.3).

### *9.1. Basic TT representation*

For simplicity, we restrict our considerations to the finite-dimensional case of $\mathbf{V} = \bigotimes_{j=1}^{d} V_j$ with $V_j = \mathbb{K}^{I_j}$. The tensor $\mathbf{v} \in \mathbf{V}$ is given component-wise by

$$\mathbf{v}[i_1 i_2 \cdots i_d] = \sum_{k_1=1}^{\rho_1} \cdots \sum_{k_{d-1}=1}^{\rho_{d-1}} v^{(1)}_{i_1 k_1} \cdot v^{(2)}_{k_1 i_2 k_2} \cdot \ldots \cdot v^{(d-1)}_{k_{d-2} i_{d-1} k_{d-1}} \cdot v^{(d)}_{k_{d-1} i_d} \quad (9.1a)$$

for all $(i_1, \ldots, i_d) \in \mathbf{I} := I_1 \times \cdots \times I_d$. The scalars

$$v^{(j)}_{k_{j-1}, i_j, k_j}$$

can be considered as entries of a tensor of order three from $\mathbb{K}^{K_{j-1}} \otimes \mathbb{K}^{I_j} \otimes \mathbb{K}^{K_j}$, where

$$K_j = \{1, \ldots, \rho_j\} \quad \text{for } 0 \le j \le d. \quad (9.1b)$$

In the cases of $j = 1$ and $j = d$ we set

$$\rho_0 = \rho_d = 1, \quad v^{(1)}_{i_1 k_1} = v^{(1)}_{1, i_1 k_1}, \quad v^{(d)}_{k_{d-1} i_d} = v^{(d)}_{k_{d-1} i_d, 1}, \quad (9.1c)$$

since

$$\mathbb{K}^{K_0} \otimes \mathbb{K}^{I_1} \otimes \mathbb{K}^{K_1} \cong \mathbb{K}^{I_1} \otimes \mathbb{K}^{K_1} \quad \text{and} \quad \mathbb{K}^{K_{d-1}} \otimes \mathbb{K}^{I_d} \otimes \mathbb{K}^{K_d} \cong \mathbb{K}^{K_{d-1}} \otimes \mathbb{K}^{I_d}$$

[30] The meaning of 'TT' changed from 'tree tensor' to 'tensor train'.

[31] In fact, the format called the *extended tensor-train decomposition* in Oseledets and Tyrtyshnikov (2009*b*, equation (11)) is exactly the same as a particular form of the hierarchical representation. A comparison of the TT and the general hierarchical format is the subject of Grasedyck and Hackbusch (2011).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

(see Remark 2.2(b)). Rewriting $v^{(j)}_{k_{j-1}i_jk_j}$ as $v^{(j)}_{k_{j-1}k_j}[i_j]$, we reformulate (9.1a) as

$$\mathbf{v}[i_1i_2\cdots i_d] = \sum_{k_1=1}^{\rho_1} \cdots \sum_{k_{d-1}=1}^{\rho_{d-1}} v^{(1)}_{k_1}[i_1] \cdot v^{(2)}_{k_1k_2}[i_2] \cdot \ldots \cdot v^{(d-1)}_{k_{d-2}k_{d-1}}[i_{d-1}] \cdot v^{(d)}_{k_{d-1}}[i_d]. \tag{9.2}$$

Fixing the indices $i_1, \ldots, i_d$, we interpret $v^{(j)}_{k_{j-1}k_j}[i_j]$ as entries of the matrix

$$V^{(j)}[i_j] := \big(v^{(j)}_{k_{j-1}k_j}[i_j]\big)_{k_{j-1}\in K_{j-1},\, k_j\in K_j} \in \mathbb{K}^{K_{j-1}\times K_j} \quad (i_j \in I_j)$$

(using (9.1c)). Then the entries $\mathbf{v}[i_1, i_2, \ldots, i_d]$ can be regarded as matrix products:

$$\mathbf{v}[i_1i_2\cdots i_d] = V^{(1)}[i_1] \cdot V^{(2)}[i_2] \cdot \ldots \cdot V^{(d-1)}[i_{d-1}] \cdot V^{(d)}[i_d] \in \mathbb{K}. \tag{9.3}$$

This representation justifies the term 'matrix product representation'. Note that $V^{(1)}[i_1] \in \mathbb{K}^{1\times\rho_1} \cong \mathbb{K}^{\rho_1}$ is a row vector, while $V^{(d)}[i_d] \in \mathbb{K}^{\rho_{d-1}\times 1} \cong \mathbb{K}^{\rho_{d-1}}$ is a column vector.

For fixed $k_{j-1}$ and $k_j$, the entries $v^{(j)}_{k_{j-1}k_j}[i_j]$ define the vectors $v^{(j)}_{k_{j-1}k_j} \in \mathbb{K}^{I_j} = V_j$ (for all $1 \le j \le d$, bearing in mind (9.1c)). Then (9.1a) is equivalent to

$$\mathbf{v} = \sum_{k_1=1}^{\rho_1} \sum_{k_2=1}^{\rho_2} \cdots \sum_{k_{d-1}=1}^{\rho_{d-1}} v^{(1)}_{1,k_1} \otimes v^{(2)}_{k_1k_2} \otimes v^{(3)}_{k_2k_3} \otimes \cdots \otimes v^{(d-1)}_{k_{d-2}k_{d-1}} \otimes v^{(d)}_{k_{d-1},1}. \tag{9.4}$$

Formulation (9.4) can be used for general spaces $V_j$. Using $K_j$ from (9.1b) together with (9.1c), we shorten the notation (9.4) by

$$\mathbf{v} = \sum_{k_0\in K_0} \cdots \sum_{k_d\in K_d} \bigotimes_{j=1}^{d} v^{(j)}_{k_{j-1}k_j} \quad \text{with } v^{(j)}_{k_{j-1}k_j} \in V_j.$$

**Definition 9.1.** Let $\mathbf{V} = {}_a\bigotimes_{j=1}^{d} V_j$ and fix a tuple

$$\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{d-1}) \in \mathbb{N}^{d-1}.$$

The TT format is defined by

$$\mathbb{T}_{\boldsymbol{\rho}} = \mathbb{T}_{\boldsymbol{\rho}}(\mathbf{V}) := \left\{ \mathbf{v} \in \mathbf{V} : \begin{array}{l} \mathbf{v} = \sum\limits_{\substack{k_i\in K_i \\ (0\le i\le d)}} \bigotimes\limits_{j=1}^{d} v^{(j)}_{k_{j-1}k_j} \text{ with } v^{(j)}_{k_{j-1}k_j} \in V_j, \\ \text{and } \#K_j = \begin{cases} 1 & \text{for } j = 0 \text{ or } j = d, \\ \rho_j & \text{for } 1 \le j \le d-1. \end{cases} \end{array} \right\} \tag{9.5}$$

**Theorem 9.2.** Let $\mathbf{v} \in \mathbb{T}_{\boldsymbol{\rho}}$ with $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{d-1})$. Then the inequalities $\rho_j \ge \rho_j^*$ hold with $\rho_j^* := \operatorname{rank}_{\{1,\ldots,j\}}(\mathbf{v})$ for $1 \le j \le d-1$ (see (2.10)). Furthermore, a representation $\mathbf{v} \in \mathbb{T}_{\boldsymbol{\rho}^*}$ exists for the minimal rank parameters

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

$\boldsymbol{\rho}^* = (\rho_1^*, \ldots, \rho_{d-1}^*)$. Therefore, an equivalent definition of $\mathbb{T}_{\boldsymbol{\rho}}$ is

$$\mathbb{T}_{\boldsymbol{\rho}} = \{\mathbf{v} \in \mathbf{V} : \operatorname{rank}_{\{1,\ldots,j\}}(\mathbf{v}) \le \rho_j \quad \text{for } 1 \le j \le d-1\}.$$

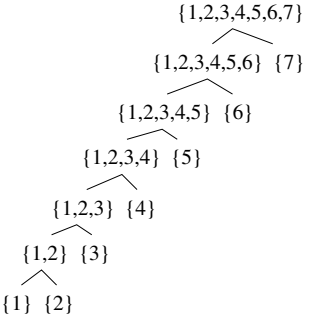
The storage cost is obtained by adding the sizes of the matrices involved in (9.3):

$$\operatorname{storage}_{\mathsf{TT}} = \sum_{j=1}^{d} n_j \rho_{j-1} \rho_j \le n\rho^2 \quad \Big(n := \max_j n_j,\ \rho := \max_j \rho_j\Big)$$

($\rho_d = 1$). This cost has a worse asymptotic than the cost for the hierarchical format (see also Section 9.3).

### *9.2. Related subspaces*

Let $D := \{1, \ldots, d\}$. The dimension partition tree $T_D^{\mathsf{TT}}$ is as follows:




Here $T_D^{\mathsf{TT}}$ consists of leaves $\{j\}$ ($j \in D$) and interior nodes $\{1, \ldots j\}$ for $j \in D \backslash \{1\}$:

$$T_D^{\mathsf{TT}} = \{\{1, \ldots j\}, \{j\} : 1 \le j \le d\}. \tag{9.6a}$$

The first son of $\{1, \ldots, j\}$ is $\{1, \ldots, j-1\}$, and the second one is $\{j\}$:

$$S(\{1, \ldots, j\}) = \{\{1, \ldots, j-1\}, \{j\}\} \tag{9.6b}$$

for $2 \le j \le d$. According to Section 7.3, we have to introduce subspaces $\mathbf{U}_\alpha \subset \mathbf{V}_\alpha$ for all $\alpha \in T_D^{\mathsf{TT}}$. For $j = 1$ we choose

$$U_1 = \mathbf{U}_{\{1\}} = \operatorname{span}\{v_{k_1}^{(1)} : k_1 \in K_1\}. \tag{9.7}$$

Here, the vectors $v_{k_1}^{(1)}$ (and later $v_{k_{j-1}k_j}^{(j)}$) are those from the representation (9.4). For $j > 1$, the trivial choice

$$U_j = V_j \quad \text{for } j \in D \backslash \{1\} = \{2, \ldots, d\}$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

is made. The next interior node is $\{1,2\} \in T_D^{\mathsf{TT}}$. We set

$$\mathbf{v}_{k_2}^{\{1,2\}} := \sum_{k_1=1}^{\rho_1} v_{1,k_1}^{(1)} \otimes v_{k_1 k_2}^{(2)} \quad \text{and} \quad \mathbf{U}_{\{1,2\}} := \operatorname{span}\{\mathbf{v}_{k_2}^{\{1,2\}} : k_2 \in K_2\}.$$

In the general case,

$$\mathbf{v}_{k_j}^{\{1,\dots,j\}} := \sum_{k_0 \in K_0} \cdots \sum_{k_{j-1} \in K_{j-1}} \bigotimes_{\ell=1}^{j} v_{k_{j-1} k_j}^{(\ell)}$$

is obtained recursively by

$$\mathbf{v}_{k_j}^{\{1,\dots,j\}} = \sum_{k_{j-1}=1}^{\rho_{j-1}} \mathbf{v}_{k_{j-1}}^{\{1,\dots,j-1\}} \otimes v_{k_{j-1} k_j}^{(j)} \quad (k_j \in K_j).$$

These tensors define the subspace

$$\mathbf{U}_{\{1,\dots,j\}} := \operatorname{span}\{\mathbf{v}_{k_j}^{\{1,\dots,j\}} : k_j \in K_j\} \quad \text{for } j \in D\backslash\{1\}$$

(the case $j = 1$ has already been stated in (9.7)). Since

$$\mathbf{v}_{k_{j-1}}^{\{1,\dots,j-1\}} \in \mathbf{U}_{\{1,\dots,j-1\}} \quad \text{and} \quad v_{k_{j-1} k_j}^{(j)} \in U_j = V_j,$$

we obtain the inclusion

$$\mathbf{U}_{\{1,\dots,j\}} \subset \mathbf{U}_{\{1,\dots,j-1\}} \otimes U_j \quad \text{for } j \in D\backslash\{1\},$$

which is the nestedness condition (7.2a). Because of $\#K_d = 1$ (see (9.5)), there is only one tensor $\mathbf{v}_{k_d}^{\{1,\dots,d\}} = \mathbf{v}$ which spans $\mathbf{U}_D$. This proves the last requirement of (7.2a–d):

$$\mathbf{v} \in \mathbf{U}_D \quad \text{and} \quad \dim(\mathbf{U}_D) = 1.$$

*9.3. From hierarchical format to TT format*

Now we start from $\mathbf{v} \in \mathcal{H}_{\mathfrak{r}} \subset \bigotimes_{j=1}^{d} \mathbb{K}^{I_j}$ with underlying tree $T_D^{\mathsf{TT}}$ from (9.6a,b) and rank tuple $\mathfrak{r} = (r_\alpha)_{\alpha \in T_D^{\mathsf{TT}}}$. The tensor is represented by the bases $\{b_i^{(j)}\}$ at the leaves, the coefficient matrices $C^{(\{1,\dots,j\},\ell)}$ at the vertices $\{1,\dots,j\} \in T_D^{\mathsf{TT}}\backslash\mathcal{L}(T_D^{\mathsf{TT}})$, and the coefficients $c_i^{(D)}$. Then it can be shown that the matrix product representation (9.2) with $\rho_\ell := r_{\{1,\dots,\ell\}}$ has the following coefficients:

$$v_{k_1}^{(1)} := b_{k_1}^{(1)} \qquad \text{for } j = 1, \tag{9.8a}$$

$$v_{k_{j-1},k_j}^{(j)} := \textstyle\sum_{\ell_j=1}^{r_j} c_{k_{j-1},\ell_j}^{(\{1,\dots,j\},k_j)} b_{\ell_j}^{(j)} \qquad \text{for } 2 \le j \le d-1, \tag{9.8b}$$

$$v_{k_{d-1}}^{(d)} := c_{k_d}^{(D)} \textstyle\sum_{\ell_d=1}^{r_d} c_{k_{d-1},\ell_d}^{(D,1)} b_{\ell_d}^{(j)} \qquad \text{for } j = d, \tag{9.8c}$$

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

with $1 \le k_j \le \rho_j$ for $1 \le j \le d-1$. The difference between the TT format (9.2) and the hierarchical format is the choice of the subspace $U_j$. The hierarchical format uses a subspace $U_j \subset V_j$, which in the best case is the minimal one: $\dim(U_j) = \mathrm{rank}_j(\mathbf{v})$. The TT format works without a proper subspace, that is, with $U_j = V_j$. If we introduce this (maximal) subspace and the canonical basis $b_\ell^{(j)}[i] = \delta_{\ell i}$ in the hierarchical format, (9.8b) and (9.8c) become

$$v^{(j)}_{k_{j-1},k_j}[i_j] = c^{(\{1,\dots,j\},k_j)}_{k_{j-1},i_j} \quad \text{and} \quad v^{(d)}_{k_{d-1}}[i_d] = c^{(D)}_{k_d} c^{(D,1)}_{k_{d-1},i_d}.$$

The latter comment shows the almost identical structure of the hierarchical data and the TT data. This implies that algorithms defined for the hierarchical format transfer to those for the TT format, and *vice versa*. Therefore, it is not necessary to explain how the tensor operations are performed in the TT format.

### *9.4. Cyclic matrix products and tensor network states*

The definition $\rho_0 = \rho_d = 1$ in (9.1c) has the purpose of avoiding summations over these indices. Instead, one can identify the indices of $K_0 = K_d$ and allow $\rho_d > 1$:

$$\mathbf{v} = \sum_{k_1=1}^{\rho_1} \cdots \sum_{k_{d-1}=1}^{\rho_{d-1}} \sum_{k_d=1}^{\rho_d} v^{(1)}_{k_d,k_1} \otimes v^{(2)}_{k_1 k_2} \otimes \cdots \otimes v^{(d-1)}_{k_{d-2}k_{d-1}} \otimes v^{(d)}_{k_{d-1},k_d}. \tag{9.9}$$

This results in a cycle instead of a linear tree. In the following we set $D = \mathbb{Z}_d$, where $0 = d$ and hence $\rho_0 = \rho_d$. Although this tensor representation looks quite similar to (9.4), it has essentially different properties.

**Proposition 9.3.**

(a) If $\rho_j = 1$ holds for at least one direction $j \in D$, the tensor representation (9.9) coincides with the standard TT format (9.4) with the ordering $\{j+1, j+2, \dots, d, 1, \dots, j\}$.

(b) The minimal subspace $U_j^{\min}(\mathbf{v})$ is not related to a single parameter $\rho_k$ in (9.9), that is, $\rho_k$ cannot be interpreted as a subspace dimension.

(c) Inequality $\mathrm{rank}_j(\mathbf{v}) = \dim(U_j^{\min}(\mathbf{v})) \le \rho_{j-1}\rho_j$ holds for (9.9 ).

(d) In general, a cyclic representation with $\mathrm{rank}_j(\mathbf{v}) = \rho_{j-1}\rho_j$ does not exist.

*Proof.* First, assume that $j = d$ in part (a). Then $\rho_0 = \rho_d = 1$ yields (9.4).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Now, fix $j = 1$. Then the representation

$$\mathbf{v} = \sum_{k_1=1}^{\rho_1} \sum_{k_d=1}^{\rho_d} v^{(1)}_{k_d,k_1} \otimes \mathbf{v}^{[1]}_{k_d,k_1}$$

holds with

$$\mathbf{v}^{[1]}_{k_d,k_1} := \sum_{k_2,k_3,\ldots,k_{d-1}} \bigotimes_{\ell=2}^{d} v^{(\ell)}_{k_{\ell-1}k_\ell}.$$

Both indices $k_d$ and $k_1$ enter the definition of

$$U_1^{\min}(\mathbf{v}) = \{\varphi(\mathbf{v}^{[1]}_{k_d,k_1}) : \varphi \in \mathbf{V}'_{[1]}\}$$

in the same way, proving (b). Obviously, the dimension is bounded by $\rho_d\rho_1 = \rho_0\rho_1$, as stated in (c).

Finally, if $\operatorname{rank}_j(\mathbf{v})$ is prime, $\operatorname{rank}_j(\mathbf{v}) = \rho_{j-1}\rho_j$ implies that $\rho_{j-1} = 1$ or $\rho_j = 1$. Hence, by (a), (9.9) cannot be a proper cyclic representation. □

According to Proposition 9.3(a), we call (9.9) a proper representation if all $\rho_j$ are larger than 1. In the cyclic case, the ranks $\rho_j$ are not related to the dimensions of $U_j^{\min}(\mathbf{v})$. Therefore we cannot prove an analogue of Theorem 7.11 to show closedness of the format (9.9). In fact, non-closedness is proved (see Theorem 9.4).

The cycle $\mathbb{Z}_d$ is only one example of tensor representations based on graphs (instead of trees). Examples[32] are given in Hübener, Nebendahl and Dür (2010, p. 5); in particular, multidimensional grid-shaped graphs are considered instead of the one-dimensional chain used in (9.4). For computations in tensor networks we refer to Huckle, Waldherr and Schulte-Herbrüggen (2013) and Espig, Hackbusch, Handschuh and Schneider (2012*a*). Handschuh (2012) describes how tensors represented in one network can be transferred into another network topology.

Whenever the graph contains a cycle, statements as in Proposition 9.3(b)–(d) can be made and instability must be expected because of Theorem 9.4. If a connected graph contains no cycle, it describes a tree and corresponds to the hierarchical format (with the possible generalization that the dimension partition tree $T_D$ is not necessarily binary).

The following result is proved by Landsberg (2012, Theorem 14.1.2.2). It implies that a similar kind of instability may occur as for the $r$-term format (see Conclusion 5.3).

**Theorem 9.4.** In general, a graph-based format containing a cycle is not closed.

[32] In physics, graph-based tensor formats are used, in particular for spin systems. The following names are used: tensor network states, finitely correlated states (FCS), valance-bond solids (VBS), projected entangled pair states (PEPS), *etc.*

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

## 10. Optimization

Many problems can be written as minimization problems of the form

$$\text{find } \mathbf{x} \in \mathbf{V} \text{ such that } J(\mathbf{x}) = \min_{\mathbf{v}\in\mathbf{V}} J(\mathbf{v}). \tag{10.1}$$

Examples are linear systems $\mathbf{A}\mathbf{x} = \mathbf{b}$ with $\mathbf{x}, \mathbf{b} \in \mathbf{V} := \bigotimes_{j=1}^{d} \mathbb{K}^{n_j}$ and $\mathbf{A} \in \mathbf{M} := \bigotimes_{j=1}^{d} \mathbb{K}^{n_j \times n_j}$. Here, $J$ takes the form

$$J(\mathbf{v}) = \langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle - 2\mathrm{Re}\,\langle \mathbf{b}, \mathbf{v}\rangle \tag{10.2}$$

if $\mathbf{A}$ is positive definite (Hackbusch 1994, § 10.1.4). Otherwise, use

$$J(\mathbf{v}) = \|\mathbf{A}\mathbf{v} - \mathbf{b}\|^2 \quad \text{or} \quad \|\mathbf{B}(\mathbf{A}\mathbf{v} - \mathbf{b})\|^2$$

with a preconditioning operator $\mathbf{B}$. The largest eigenvalue of a positive definite matrix $\mathbf{A}$ and the corresponding eigenvector can be determined from the Rayleigh quotient

$$J(\mathbf{v}) = \frac{\langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}.$$

Fix some tensor format and the corresponding rank parameter, that is, the set of tensors is $\mathcal{F} = \mathcal{F}(\mathbf{V}) \in \{\mathcal{R}_r, \mathcal{T}_{\mathbf{r}}, \mathcal{H}_{\mathfrak{r}}, \mathbb{T}_\rho\}$. Instead of problem (10.1), we want to solve

$$\text{find } \mathbf{x}_{\mathcal{F}} \in \mathcal{F} \text{ such that } J(\mathbf{x}_{\mathcal{F}}) = \min_{\mathbf{v}\in\mathcal{F}} J(\mathbf{v}). \tag{10.3}$$

Falcó and Nouy (2012, Theorems 1 and 2) discuss the existence of minimizers. The advantage of the optimization approach is the fact that one does not leave the fixed ranks of the format. In particular, no truncation is needed to reduce an increased rank.

By definition, a tensor $\mathbf{v} \in \mathcal{F}$ depends on certain parameters:

$$\mathbf{v} = \rho(p_1, p_2, \ldots, p_m),$$

which may be grouped suitably. The standard iterative method for solving (10.3) is a Gauss–Seidel-type method, which is called ALS (alternating least squares) in the case of quadratic functions $J(\mathbf{v})$. For fixed $p_2, \ldots, p_m$, one minimizes $J(\rho(p_1, p_2, \ldots, p_m))$ with respect to $p_1$. Then $p_1$ is fixed, and one minimizes with respect to $p_2$, *etc.* This method is discussed in Espig, Hackbusch, Rohwedder and Schneider (2012*b*), Oseledets (2011*a*), Uschmajew (2012), Holtz, Rohwedder and Schneider (2012*b*) and Mohlenkamp (2013).

Since the mapping $\rho$ is not injective, we are interested in the tangent space $\mathcal{T}(\mathbf{v})$ spanned by $\partial\rho/\partial p_i$ and its geometric properties. Here we refer to Uschmajew (2010) and Holtz, Rohwedder and Schneider (2012*a*). The geometric structures in the infinite-dimensional case are discussed by Falcó, Hackbusch and Nouy (2013).

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

The tangent space is also of importance for the solution of the dynamic problem $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}(t) = F(\mathbf{v}(t), t)$. Here, the *Dirac–Frenkel discretization* is applied, which is the variational formulation

$$\mathbf{v}_{\mathcal{F}}(t) \in \mathcal{F} \quad \text{with} \quad \left\langle \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}_{\mathcal{F}}(t) - \mathbf{F}(t, \mathbf{v}_{\mathcal{F}}(t)), \mathbf{t} \right\rangle = 0 \quad \text{for all } \mathbf{t} \in \mathcal{T}(\mathbf{v}_{\mathcal{F}}).$$

For an error analysis of this discretization we refer to Lubich (2005), Koch and Lubich (2010), Arnold and Jahnke (2012), Lubich and Oseledets (2013) and Lubich, Rohwedder, Schneider and Vandereycken (2013).

## 11. Multivariate cross approximation

### *11.1. Matrix case*

The (adaptive) cross approximation is a heuristic method which computes a rank-$k$ approximation to an $n \times m$ matrix and costs $O(knm)$ operations: see Bebendorf (2011) and Hackbusch (2009, § 9.4). Let $M \in \mathbb{K}^{I \times J}$. In essence, we choose $r$ rows (index subset $\tau := \{i_1, \ldots, i_r\}$) and $r$ columns (index subset $\sigma := \{j_1, \ldots, j_r\}$) and determine a matrix $R_r$ of rank $r$ which interpolates on all rows from $\tau$ and all columns from $\sigma$, that is, $R[i, j] = M[i, j]$ for all index pairs with either $i \in \tau$ or $j \in \sigma$. The matrix $R_r$ is defined by

$$R_r = M|_{I \times \sigma} \cdot (M|_{\tau \times \sigma})^{-1} \cdot M|_{\tau \times J}, \tag{11.1}$$

provided that the $r \times r$ matrix $M|_{\tau \times \sigma}$ is regular.

In practice, $R_r$ is determined recursively as a sum of rank-one matrices. The first matrix is $R_1 := E_1 := M[\cdot, j_1]M[i_1, \cdot]/M[i_1, j_1]$. Here, $i_1$ and $j_1$ must be chosen such that $M[i_1, j_1] \neq 0$. This rank-one matrix coincides with $M$ at the first *cross* defined by the index subset $C_1 := \{(i, j) : i = i_1 \text{ or } j = j_1\}$. Therefore the difference $M_1 := M - E_1$ vanishes at $C_1$. Again, $E_2 := M_1[\cdot, j_2]M_1[i_2, \cdot]/M_1[i_2, j_2]$ is a rank-one matrix interpolating $M_1$ at the second cross $C_2$, provided that $M_1[i_2, j_2] \neq 0$. Set $M_2 := M_1 - E_2$ and $R_2 := R_1 + E_2$. $R_2$ coincides with $M$ at $C_1 \cup C_2$. Repeating this procedure, we obtain $R_r$ from above.

If $\operatorname{rank}(M) = r$, there exist sets $\tau := \{i_1, \ldots, i_r\}$, $\sigma := \{j_1, \ldots, j_r\}$ such that $M|_{\tau \times \sigma}$ is regular and the construction recovers $M$, that is, $R_r = M$.

The adaptive variant ACA (adaptive cross approximation) tries to choose the indices $i_\ell$ and $j_\ell$ adaptively. The theoretically best choice is the total pivot strategy $(i_\ell, j_\ell) := \operatorname{argmax}\{|M_{\ell-1}[i, j]| : (i, j) \in I \times J\}$. The required evaluation of all entries is not feasible for large-scale matrices. Instead, pivot strategies are used which require only the evaluation of certain rows or columns of $M$. If, for example, a column $j_\ell$ is selected, then $i_\ell := \operatorname{argmax}\{|M_{\ell-1}[i, j_\ell]| : i \in I\}$ is a natural choice for the $\ell$th row.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

### *11.2. Order $d \geq 3$*

Let $\mathbf{V} = \mathbb{K}^{I_1} \otimes \cdots \otimes \mathbb{K}^{I_d}$ and $\mathbf{I} := I_1 \times \cdots \times I_d$. The matricization $\mathcal{M}_\alpha : \mathbf{V} \mapsto \mathbb{K}^{\mathbf{I}_\alpha \times \mathbf{I}_{\alpha^c}}$ from Section 2.6 allows us to turn the tensor $\mathbf{v}$ into a matrix, to which we can apply the cross approximation from above. The index sets $I$ and $J$ in (11.1) are now $\mathbf{I}_\alpha = \times_{j\in\alpha} I_j$ and $\mathbf{I}_{\alpha^c} = \times_{j\in\alpha^c} I_j$. This shows that the 'rows' and 'columns' are not simple vectors but tensors in $\mathbf{V}_\alpha = \bigotimes_{j\in\alpha} V_j$ and $\mathbf{V}_{\alpha^c}$. Therefore, the factors $M|_{I\times\sigma}$ and $M|_{\tau\times J}$ are not directly representable. However, the matrix $(M|_{\tau\times\sigma})^{-1}$ is still of size $r \times r$. The 'columns' $M[\cdot, \mathbf{i}_{\alpha^c}] \in \mathbf{V}_\alpha$ and the 'rows' $M[\mathbf{i}_\alpha, \cdot] \in \mathbf{V}_{\alpha^c}$ are the basis vectors of the subspaces $\mathbf{U}_\alpha$ and $\mathbf{U}_{\alpha^c}$ required in the hierarchical format. It turns out that all constructions can be described by the parameters of the hierarchical format. The above-mentioned adaptive strategies also apply to the tensor case. Furthermore, the cross approximation can reconstruct a tensor $\mathbf{v}$ of hierarchical rank $\mathfrak{r} := (\operatorname{rank}_\alpha(\mathbf{v}))_{\alpha\in T_D}$ in $\mathcal{H}_\mathfrak{r}$ exactly. A detailed description is given by Ballani (2012), Ballani, Grasedyck and Kluge (2013) and Hackbusch (2012, § 15.4).

Other variants of the cross approximation are proposed by Espig, Grasedyck and Hackbusch (2009) and Oseledets, Savostyanov and Tyrtyshnikov (2010).

### *11.3. Applications*

#### *11.3.1. Coding of multivariate functions*

Let $F(x_1, \ldots, x_d)$ be a function defined for $x_j \in I_j$ and assume that there is an implementation which evaluates $F$ at least with a sufficient accuracy. For instance, $F(x_1, \ldots, x_d)$ may be a complicated integral with parameters $x_j$, for which sufficiently accurate quadrature methods can be applied. If we could evaluate $F|_G$ in a grid $G$ contained in $\mathbf{I} := I_1 \times \cdots \times I_d$, $F$ could be approximated at any $x_1, \ldots, x_d$. The cross approximation procedure can be used to obtain a tensor $\mathbf{v}_F$ describing the grid function $F|_G$. The cost for determining the tensor $\mathbf{v}_F$ is less important, since this computation is done only once. Afterwards, it is cheap to evaluate the tensor $\mathbf{v}_F$ (instead of the original implementation of $F$). An example of such an application is provided by Ballani (2012).

#### *11.3.2. Pointwise functions of tensors*

Let $\mathbf{v} \in \mathbf{V} = \mathbb{K}^{I_1} \otimes \cdots \otimes \mathbb{K}^{I_d}$, while $\varphi : \mathbb{K} \to \mathbb{K}$. Define the function

$$\Phi : \mathbf{V} \to \mathbf{V} \text{ by } \Phi(\mathbf{v})[\mathbf{i}] := \varphi(\mathbf{v}[\mathbf{i}]) \quad \text{for all } \mathbf{i} \in \mathbf{I} = I_1 \times \cdots \times I_d.$$

In the case of multivariate functions $\mathbf{v}$, $\Phi(\mathbf{v})$ is the standard function composition $\varphi \circ \mathbf{v}$. In general, this is a difficult operation, since even for an elementary tensor there is no simple expression for $\Phi(\mathbf{v})$.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

Assume that $\mathbf{v}$ is given in one of the formats. Since the evaluation $\mathbf{i} \mapsto \mathbf{v}[\mathbf{i}]$ is simple, $\varphi(\mathbf{v}[\mathbf{i}])$ is also easy to evaluate. Therefore, $\Phi(\mathbf{v})$ can be approximated by cross approximation.

*11.3.3. Truncation*

The cross approximation may be used as a surrogate of the singular value decomposition. In this sense, Savostyanov, Tyrtyshnikov and Zamarashkin (2012) use the cross approximation to simplify the truncation procedure.

## 12. Tensorization

In this section we transform univariate grid functions into tensors. This justifies the term 'tensorization'. We are free to use any of the tensor formats discussed above to represent the obtained tensors. In some cases, the $r$-term format works (see Remark 12.2 and following text), but mainly the hierarchical representation in form of the TT format is used.[33]

This tensorization was introduced by Oseledets (2010) (applied to matrices instead of vectors). The first examples of data compression via tensorization were presented by Khoromskij (2011). The first analysis was given by Grasedyck (2010*b*).

*12.1. Why regular grids are advantageous*

It was emphasized right at the start of the Introduction (in Section 1.1.1) that grid functions on tensor grids fit into the tensor setting. The use of regular grids seems rather old-fashioned. In fact, the modern approach tries to adapt the grid (or nodal points) to each particular situation. The goal is to obtain the desired accuracy with a minimal number of degrees of freedom. The related techniques include (a) local refinement ($h$-method), (b) the adaptive $hp$-method, and (c) wavelet methods. The price to pay for the (relatively) small number of unknowns is a more complicated system of equations and higher computational overheads. In contrast, we recommend a regular tensor grid if the underlying domain is a Cartesian product. The overhead is minimal. As already seen, the tensor formats help to reduce the size $n^d$ of grid points to $O(dn)$. The factor $n$ is caused, for example, by the basis vectors $b_i^{(j)} \in V_j = \mathbb{K}^n$. Below we shall reduce this factor $n$ by a quantity which, under suitable conditions, can become $O(\log n)$. The 'suitable conditions' are those that allow us to approximate a function (or a grid function on a regular fine grid) by an $hp$-method or by wavelets. This means that the compression, which is part of the design of an $hp$-method or a wavelet method, is now applied to the tensor (after the discretization in the

[33] The combination of tensorization and the TT format is called QTT by Khoromskij (2011). The use of the tensor subspace format does not make sense in this context.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

regular grid). While the compression technique in the various discretization methods increases the programming overhead, compression is now a simple truncation by means of singular value decompositions. Because of its black-box character, the overhead is minimal.

### *12.2. Basic isomorphism and properties*

We restrict ourselves to the vector space $\mathbb{K}^n = \mathbb{K}^I$ with $I = \{0, 1, \dots, n-1\}$ and $n = 2^d$. Let

$$\mathbf{V} := \bigotimes_{j=1}^{d} \mathbb{K}^2,$$

where $\mathbb{K}^2 = \mathbb{K}^J$ with $J = \{0, 1\}$. The isomorphism $\Psi_n : \mathbb{K}^I \to \mathbf{V}$ is given by means of the binary integer representation $k = \sum_{j=1}^{d} i_j 2^{j-1}$ $(0 \le i_j \le 1)$:

$$\begin{array}{rccl} \Psi_n : & \mathbb{K}^n & \to & \mathbf{V} \\ & v & \mapsto & \mathbf{v} \quad \text{with } v_k = \mathbf{v}[i_1 \cdots i_d] \text{ for } k = \sum_{j=1}^{d} i_j 2^{j-1}. \end{array}$$

Obviously, the vector $v$ and the tensor $\mathbf{v}$ have the same data size if we use the full format for $\mathbf{v}$. Instead we may try to use another format, either for the exact tensor $\mathbf{v}$ or for an approximation $\mathbf{v}_\varepsilon$ obtained by truncation. As shown below, this step can reduce the data size $n$ to $O(\log n)$.

**Remark 12.1.** A direct tensorization of matrices is based on the isomorphism $\Psi : \mathbb{K}^{n\times n} \cong \bigotimes_{j=1}^{d} \mathbb{K}^{2\times 2}$ via

$$M[k, \ell] = \Psi(M)[(k_1, \ell_1), \dots, (k_d, \ell_d)],$$

where $k_j$ and $\ell_j$ are the binary digits of $k$ and $\ell$ (Oseledets 2010).

### *12.3. Rank estimates by analytic tools*

#### *12.3.1. Approximation by exponential sums*

**Remark 12.2.** If $v \in \mathbb{K}^n$ has the entries $v_k = \zeta^k$ for some basis $\zeta$, then $\mathbf{v} = \Psi_n(v)$ is the elementary tensor

$$v^{(1)} \otimes \cdots \otimes v^{(d)} \quad \text{with } v^{(j)} = \begin{bmatrix} 1 \\ \zeta^{2^{j-1}} \end{bmatrix}.$$

The data size is reduced from $n$ to $2d = 2 \log_2 n$.

A simple consequence is that any (grid) function that can be approximated by exponential sums with $r$ terms, as in (5.11a), allows an approximation by a tensor from $\mathcal{R}_r$. Since the basis $\zeta$ may be complex, the approximation may also be obtained by $r$ Fourier terms.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

*12.3.2. Ranks of the TT format*

We recall that the minimal rank $\rho_j$ appearing in the TT format is defined by $\operatorname{rank}_{\{1,\dots,j\}}(\mathbf{v})$, the matrix rank of the corresponding matricization $\mathcal{M}_{\{1,\dots,j\}}(\mathbf{v})$. This matricization can be simply expressed by the vector $v$:

$$\mathcal{M}_{\{1,\dots,j\}}(\mathbf{v}) = \begin{bmatrix} v_0 & v_{2^j} & \cdots & v_{2^{d-1}} \\ v_1 & v_{2^j+1} & \cdots & v_{2^{d-1}+1} \\ \vdots & \vdots & & \vdots \\ v_{2^j-1} & v_{2^{j+1}-1} & \cdots & v_{2^d-1} \end{bmatrix}. \tag{12.1}$$

Each of the $2^{d-j}$ columns corresponds to the (grid) function $v$ in the intervals $[k2^j h, (k+1)2^j h)$ for $k = 0, \dots, 2^{d-j} - 1$ and the step size $h = 1/n$.

The data size of a TT tensor is

$$O(d\rho^2), \quad \text{where } \rho := \max_j \rho_j.$$

Note that there is no dependence on $\dim(V_j)$, since this is the constant 2.[34]

*12.3.3. Approximation by global polynomials*

Assume that $v$ can be approximated by a polynomial $v_\varepsilon$ (restricted to the grid) of degree $p$. Then each column in (12.1) is also a polynomial of degree $p$. Therefore their span has at most the dimension $\rho_j \le p + 1$. The data size of the tensor $\mathbf{v}_\varepsilon$ is $O(d(p+1)^2)$.

*12.3.4. hp-approximation*

Assume a (grid) function in $[0, 1]$ with singular (or less regular) behaviour at $x = 0$. A possible *hp*-discretization uses the partition of $[0, 1]$ into the intervals $(0, h]$, $(h, 2h]$, $(2h, 4h]$, $\dots$, $(2^j h, 2^{j+1}h]$, $\dots$, $(1/2, 1)$. If the function is asymptotically smooth (Hackbusch 2012, § 14.2.3), an error of comparable maximum norm can be obtained by an polynomial approximation of equal degree $p$. Let $\mathbf{v}_\varepsilon$ be the tensor corresponding to the grid values of this piecewise polynomial. We verify that all columns except the first one are polynomials of degree $p$. Therefore they span a space of dimension $\le p+1$. Whatever the first column is, it can increase the dimension only by one. This proves the uniform rank bound

$$\rho_j = \dim(\mathcal{M}_{\{1,\dots,j\}}(\mathbf{v}_\varepsilon)) \le p + 2.$$

If the point singularity is not at an end point of $[0, 1]$ but in the interior, two exceptional columns are present and $\rho_j \le p + 3$ follows. The original result of Grasedyck (2010*b*) considers the case of multiple point singularities.

[34] The data size of the hierarchical format contains the cubic factor $r^3$, $r := \max r_\alpha$. One of the $r_\alpha$ is now $r_j \le 2$!

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

### *12.4. Operations*

We recall the isomorphism $\Psi_n : \mathbb{K}^I \to \mathbf{V} = \otimes^d \mathbb{K}^2$ mapping the vector $v$ into the tensor $\mathbf{v}$.

#### *12.4.1. Addition, scalar product, Hadamard product*

The addition and the Hadamard product (both denoted by the symbol $\boxdot$) satisfy $\Psi_n(v \boxdot w) = \Psi_n(v) \boxdot \Psi_n(w)$. The cost of the tensor operations on the right-hand side is discussed in Section 8 and is related to the data size of the tensors (not of the vectors).

Also, the Euclidean scalar product can be transferred directly to the scalar product of the tensors: $\langle v, w\rangle = \langle \Psi_n(v), \Psi_n(w)\rangle$.

If matrices are represented in the TT format according to Remark 12.1, a similar statement holds for the matrix–vector multiplication.

#### *12.4.2. Convolution*

The $d$-variate convolution can be performed in each direction separately:

$$\left(\bigotimes_{j=1}^{d} v^{(j)}\right) \star \left(\bigotimes_{j=1}^{d} w^{(j)}\right) = \bigotimes_{j=1}^{d} (v^{(j)} \star w^{(j)})$$

(see Section 5.2.6). The proof follows directly from the definition of the multidimensional convolution. How the convolution $v \star w$ of two vectors is related to a convolution $\mathbf{v} \star \mathbf{w}$ of the tensors $\mathbf{v} = \Psi_n(v)$ and $\mathbf{w} = \Psi_n(w)$ is another question entirely. It is *almost* possible to reduce $\mathbf{v} \star \mathbf{w}$ to a $d$-fold convolution of the $\mathbb{K}^2$ components. In addition, a carry-over procedure links the $j$th direction with the $(j+1)$th direction. The computational cost is related to the data size of the tensors. Details can be found in Hackbusch (2011) or Hackbusch (2012, § 14.3).

#### *12.4.3. Fast Fourier transform*

There is a close relationship between the FFT algorithm and the binary structure of the tensorization. However, the performance of the algorithm has to be combined with truncation step in between, since otherwise the ranks may explode. A description of the algorithm is given by Dolgov, Khoromskij and Savostyanov (2012*b*) and Hackbusch (2012, § 14.4).

### *12.5. Combination with other formats*

As mentioned above, vectors from $\mathbb{K}^{n_j}$ appear in the tensor subspace format as basis vectors $b_i^{(j)}$ or as vectors $v_i^{(j)}$ in the $r$-term format. Assume that all $n_j$ are powers of two. Then the full representation of $\mathbb{K}^{n_j}$ vectors can be replaced by the TT format. This reduces not only the storage cost but also

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

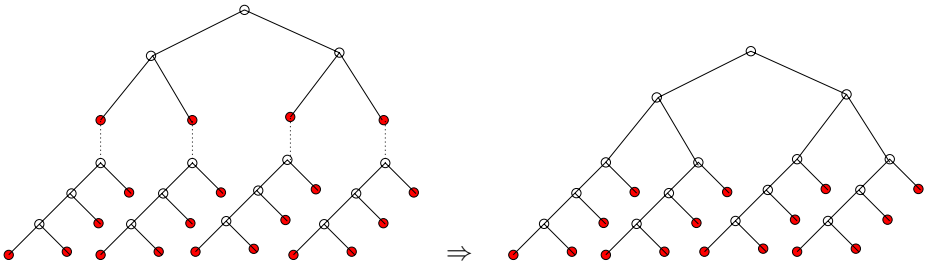



Figure 12.1. (a) Balanced tree with four leaves $V_j = \mathbb{K}^{16}$. (b) Extended tree.

all operations involving the vectors (see Section 12.4). However, in the case of the tensor subspace format, the costly core tensor remains unchanged.

If we combine any hierarchical format (with $\mathbb{K}^{n_j}$ vectors in the leaves) with the TT format, we again obtain an (extended) hierarchical format, as illustrated in Figure 12.1.

## 13. Applications

We only have space for a brief sketch of the main lines of research and references to the literature.

### *13.1. Linear systems of equations*

Consider the system $\mathbf{Au} = \mathbf{f}$, where the 'vectors' $\mathbf{u}, \mathbf{f} \in \mathbf{V}$ and the 'matrix' $\mathbf{A}$ are given in some format. The iterative solution by

$$\mathbf{u}^{m+1} = \mathbf{u}^m - \mathbf{B}(\mathbf{Au}^m - \mathbf{f})$$

has already been mentioned in Section 1.1.5. Assuming that $\mathbf{A}$ is the discretization of an elliptic operator which is an isomorphism from $H^{-1}(\Omega)$ to $H^1_0(\Omega)$, a suitable preconditioning matrix $\mathbf{B}$ is the inverse of the discretization of $-\Delta$. Instead we may use the accurate approximation discussed in Proposition 5.10. However, the iteration $\mathbf{u}^{m+1} = \mathbf{u}^m - \mathbf{B}(\mathbf{Au}^m - \mathbf{f})$, if performed exactly, increases the representation ranks of the respective format until the maximal rank is reached. Therefore, one has to insert truncations $T$ into a fixed or adaptively determined representation rank, for example, $\mathbf{u}^{m+1} = T(\mathbf{u}^m - \mathbf{B} \cdot T(\mathbf{Au}^m - \mathbf{f}))$ (Khoromskij 2009).

The linear iteration discussed above can be accelerated by the conjugate gradient methods or their variants. Note that the orthogonalization required in these methods is performed only approximately if truncation is used. If the convergence of the method is fast enough, the deviation from orthogonality does not harm the results. Different approaches involving different formats are described by Ballani and Grasedyck (2013), Oseledets, Tyrtysh-

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

nikov and Zamarashkin (2011), Kressner and Tobler (2010, 2011*a*, 2011*b*) and Savas and Eldén (2013).

Another way to solve linear systems is based on the minimization of (10.2) or of $J(\mathbf{v}) = \|\mathbf{A}\mathbf{v} - \mathbf{b}\|^2$ or $\|\mathbf{B}(\mathbf{A}\mathbf{v} - \mathbf{b})\|^2$ with $\mathbf{B}$ as above (Espig *et al.* 2012*b*).

The linear system may take the shape of matrix equations. For the treatment of the generalized Lyapunov equation we refer to Benner and Breiten (2013).

### *13.2. Elliptic PDEs with stochastic coefficients*

A source of multivariate problems is the elliptic boundary value problem

$$\operatorname{div} a(x,\omega) \operatorname{grad} u = f(x) \text{ in } \mathcal{D},\ u = 0 \text{ on } \partial\Omega,$$

where $x \in \mathcal{D} \subset R^m$ is the spatial variable, while $\omega \in \Omega$ is a stochastic variable. To ensure ellipticity, one has to assume $0 < a_- \le a(x,\omega) \le a_+ < \infty$ for almost all $(x,\omega) \in \mathcal{D}\times\Omega$. Under standard assumptions, the stochastic behaviour can be approximated by the Karhunen–Loève expansion (infinite singular value decomposition). The truncation to the $M$ largest terms allows us to replace $\omega$ by the parameters $\mathbf{p} = (p_1, \ldots, p_M)$. Finally, we obtain a multi-parametric problem:

$$\operatorname{div} a_M(x,\mathbf{p}) \operatorname{grad} u_M = f(x) \text{ in } \mathcal{D},\ u_M = 0 \text{ on } \partial\Omega,$$

defining solutions $u_M = u_M(x,\mathbf{p})$.

For the algorithms and analysis we refer to the survey by Schwab and Gittelson (2011). Among many publications concerning this topic we mention Khoromskij and Schwab (2011), Dolgov, Kazeev and Khoromskij (2012*a*) and Espig, Hackbusch, Litvinenko, Matthies and Wähnert (2014).

### *13.3. Problems from quantum chemistry*

The Schrödinger equation describes an eigenvalue problem in the subspace of the antisymmetric functions, which are not well suited to the formats discussed so far (see the end of Section 2.1). According to Remark 3.8, fast approximation in a tensor format requires the mixed derivatives to be under control. Here, the pioneering results of Yserentant (2010) are important. The eigenfunctions possess mixed derivatives which, even after exponential weighting, are $L^2$ functions. Further, the results of Flad, Hackbusch and Schneider (2006, 2007) show that the wavefunctions are well representable in tensor formats.

The simpler Hartree–Fock or Kohn–Sham formulations are (nonlinear) eigenvalue problems for a differential equation in $\mathbb{R}^3$. The standard approach in quantum chemistry is approximation by sums of Gaussians multiplied by polynomials. This ansatz yields relatively good approximations

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

for already low dimension, but it does not behave very well asymptotically. Approximations by regular grids (see Section 12.1) promise better convergence, provided that the $\mathbb{R}^3$ functions appearing in the Hartree–Fock equation allow a favourable approximation in tensor formats. Indeed, the test by Chinnamsetty, Espig, Flad and Hackbusch (2010) shows a good approximation in the $r$-term format. The general concept of the tensor approach for Hartree–Fock is described by Flad, Hackbusch, Khoromskij and Schneider (2010). A numerical realization is presented by Khoromskij, Khoromskaia and Flad (2011).

An important part of many-electron problems is the calculation of many-electron integrals. For such a tensor calculation of two-electron integrals see, for example, Benedikt, Auer, Espig and Hackbusch (2011).

## REFERENCES[35]

I. Affleck, T. Kennedy, E. H. Lieb and H. Tasaki (1987), 'Rigorous results on valence-bond ground states in antiferromagnets', *Phys. Rev. Lett.* **59**, 799–802.

C. A. Andersson and R. Bro (2000), 'The $N$-way toolbox for MATLAB', *Chemometrics Intelligent Lab. Syst.* **52**, 1–4.

C. J. Appellof and E. R. Davidson (1981), 'Strategies for analyzing data from video fluorometric monitoring of liquid-chromatographic effluents', *Anal. Chem.* **13**, 2053–2056.

A. Arnold and T. Jahnke (2012), On the approximation of high-dimensional differential equations in the hierarchical Tucker format. Preprint 12/09, KIT, Karlsruhe.

B. W. Bader and T. G. Kolda (2007), MATLAB tensor toolbox, version 2.3. Technical report: http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox

J. Ballani (2012), 'Fast evaluation of singular BEM integrals based on tensor approximations', *Numer. Math.* **121**, 433–460.

J. Ballani and L. Grasedyck (2013), 'A projection method to solve linear systems in tensor format', *Numer. Linear Algebra Appl.* **20**, 27–43.

J. Ballani, L. Grasedyck and M. Kluge (2013), 'Black box approximation of tensors in hierarchical Tucker format', *Linear Algebra Appl.* **438**, 639–657.

M. Bebendorf (2011), 'Adaptive cross approximation of multivariate functions', *Constr. Approx.* **34**, 149–179.

U. Benedikt, A. A. Auer, M. Espig and W. Hackbusch (2011), 'Tensor decomposition in post-Hartree–Fock methods I: Two-electron integrals and MP2', *J. Chem. Phys.* **134**, 054118.

P. Benner and T. Breiten (2013), 'Low rank methods for a class of generalized Lyapunov equations and related issues', *Numer. Math.* **124**, 441–470.

[35] The URLs cited in this work were correct at the time of going to press, but the publisher and the authors make no undertaking that the citations remain live or are accurate or appropriate.

https://doi.org/10.1017/S0962492914000087 Published online by Cambridge University Press

G. Beylkin and M. J. Mohlenkamp (2005), 'Algorithms for numerical analysis in high dimensions', *SIAM J. Sci. Comput.* **26**, 2133–2159.

G. Beylkin and L. Monzón (2010), 'Approximation by exponential sums revisited', *Appl. Comput. Harmon. Anal.* **28**, 131–149.

G. Beylkin, M. J. Mohlenkamp and F. Pérez (2008), 'Approximating a wavefunction as an unconstrained sum of Slater determinants', *J. Math. Phys.* **49**, 032107.

D. Bini, G. Lotti and F. Romani (1980), 'Approximate solutions for the bilinear form computational problem', *SIAM J. Comput.* **9**, 692–697.

D. Braess (1986), *Nonlinear Approximation Theory*, Springer.

D. Braess and W. Hackbusch (2005), 'Approximation of $1/x$ by exponential sums in $[1, \infty)$', *IMA J. Numer. Anal.* **25**, 685–697.

D. Braess and W. Hackbusch (2009), On the efficient computation of high-dimensional integrals and the approximation by exponential sums. In *Multiscale, Nonlinear and Adaptive Approximation* (R. A. DeVore and A. Kunoth, eds), Springer, pp. 39–74.

J. Buczyński and J. M. Landsberg (2014), 'On the third secant variety', *J. Algebraic Combin.*, to appear.

H.-J. Bungartz and M. Griebel (2004), Sparse grids. In *Acta Numerica*, Vol. 13, Cambridge University Press, pp. 147–269.

R. B. Cattell (1944), 'Parallel proportional profiles and other principles for determining the choice of factors by rotation', *Psychometrika* **9**, 267–283.

L. Chiantini and G. Ottaviani (2012), 'On generic identifiability of 3-tensors of small rank', *SIAM J. Matrix Anal. Appl.* **33**, 1018–1037.

S. R. Chinnamsetty, M. Espig, H.-J. Flad and W. Hackbusch (2010), 'Canonical tensor products as a generalization of Gaussian-type orbitals', *Z. Phys. Chem.* **224**, 681–694.

P. Comon (2002), Tensor decomposition: State of the art and applications. In *Mathematics in Signal Processing V* (J. G. McWhirter and I. K. Proudler, eds), Oxford University Press, pp. 1–24.

L. De Lathauwer (2008), 'Decompositions of a higher-order tensor in block terms II: Definitions and uniqueness', *SIAM J. Matrix Anal. Appl.* **30**, 1033–1066.

L. De Lathauwer, B. De Moor and J. Vandewalle (2000*a*), 'A multilinear singular value decomposition', *SIAM J. Matrix Anal. Appl.* **21**, 1253–1278.

L. De Lathauwer, B. De Moor and J. Vandewalle (2000*b*), 'An introduction to independent component analysis', *J. Chemometrics* **14**, 123–149.

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

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

S. Dolgov, V. A. Kazeev and B. Khoromskij (2012*a*), The tensor-structured solution of one-dimensional elliptic differential equations with high-dimensional parameters. Preprint 51, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig.

S. Dolgov, B. Khoromskij and D. V. Savostyanov (2012*b*), 'Superfast Fourier transform using QTT approximation', *J. Fourier Anal. Appl.* **18**, 915–953.

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

M. Espig and W. Hackbusch (2012), 'A regularized Newton method for the efficient approximation of tensors represented in the canonical tensor format', *Numer. Math.* **122**, 489–525.

M. Espig, L. Grasedyck and W. Hackbusch (2009), 'Black box low tensor-rank approximation using fiber-crosses', *Constr. Approx.* **30**, 557–597.

M. Espig, W. Hackbusch, S. Handschuh and R. Schneider (2012*a*), 'Optimization problems in contracted tensor networks', *Comput. Vis. Sci.* **14**, 271–285.

M. Espig, W. Hackbusch, A. Litvinenko, H. G. Matthies and P. Wähnert (2014), 'Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats', *Comput. Math. Appl.* **67**, 818–829.

M. Espig, W. Hackbusch, A. Litvinenko, H. G. Matthies and E. Zander (2013), Efficient analysis of high dimensional data in tensor formats. In Garcke and Griebel (2013), pp. 31–56.

M. Espig, W. Hackbusch, T. Rohwedder and R. Schneider (2012*b*), 'Variational calculus with sums of elementary tensors of fixed rank', *Numer. Math.* **122**, 469–488.

A. Falcó and A. Nouy (2012), 'Proper generalized decomposition for nonlinear convex problems in tensor Banach spaces', *Numer. Math.* **121**, 503–530.

A. Falcó, W. Hackbusch and A. Nouy (2013), Geometric structures in tensor representations. Preprint 9, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig.

H.-J. Flad, W. Hackbusch and R. Schneider (2006), 'Best $N$-term approximation in electronic structure calculations I: One-electron reduced density matrix', *M2AN* **40**, 49–61.

H.-J. Flad, W. Hackbusch and R. Schneider (2007), 'Best $N$-term approximation in electronic structure calculations II: Jastrow factors', *M2AN* **41**, 261–279.

H.-J. Flad, W. Hackbusch, B. Khoromskij and R. Schneider (2010), Concept of data-sparse tensor-product approximation in many-particle modelling. In *Matrix Methods: Theory, Algorithms, Applications* (V. Olshevsky and E. E. Tyrtyshnikov, eds), World Scientific, pp. 313–343.

J. Garcke (2013), Sparse grids in a nutshell. In Garcke and Griebel (2013), pp. 57–80.

J. Garcke and M. Griebel, eds (2013), *Sparse Grids and Applications*, Vol. 88 of *Lecture Notes in Computational Science and Engineering*, Springer.

I. P. Gavrilyuk, W. Hackbusch and B. Khoromskij (2002), '$\mathcal{H}$-matrix approximation for the operator exponential with applications', *Numer. Math.* **92**, 83–111.

L. Grasedyck (2004), 'Existence and computation of a low Kronecker-rank approximant to the solution of a tensor system with tensor right-hand side', *Computing* **72**, 247–266.

L. Grasedyck (2010*a*), 'Hierarchical singular value decomposition of tensors', *SIAM J. Matrix Anal. Appl.* **31**, 2029–2054.

L. Grasedyck (2010*b*), Polynomial approximation in hierarchical Tucker format by vector-tensorization. DFG-SPP 1324. Preprint 43, Philipps-Universität Marburg.

L. Grasedyck and W. Hackbusch (2011), 'An introduction to hierarchical ($\mathcal{H}$-) rank and TT-rank of tensors with examples', *Comput. Methods Appl. Math.* **11**, 291–304.

L. Grasedyck, D. Kressner and C. Tobler (2013), 'A literature survey of low-rank tensor approximation techniques', *GAMM-Mitteilungen* **36**.

W. H. Greub (1975), *Linear Algebra*, Springer.

W. Hackbusch (1994), *Iterative Solution of Large Sparse System of Equations*, Cambridge University Press.

W. Hackbusch (2005), Entwicklungen nach Exponentialsummen. Techn. Bericht 25, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig.

W. Hackbusch (2009), *Hierarchische Matrizen: Algorithmen und Analysis*, Springer.

W. Hackbusch (2011), 'Tensorisation of vectors and their efficient convolution', *Numer. Math.* **119**, 465–488.

W. Hackbusch (2012), *Tensor Spaces and Numerical Tensor Calculus*, Vol. 42 of *Springer Series in Computational Mathematics*, Springer.

W. Hackbusch (2013), '$L^\infty$ estimation of tensor truncations', *Numer. Math.* **125**, 419–440.

W. Hackbusch and B. Khoromskij (2007), 'Tensor-product approximation to operators and functions in high dimensions', *J. Complexity* **23**, 697–714.

W. Hackbusch and S. Kühn (2009), 'A new scheme for the tensor representation', *J. Fourier Anal. Appl.* **15**, 706–722.

W. Hackbusch, B. Khoromskij and E. E. Tyrtyshnikov (2005), 'Hierarchical Kronecker tensor-product approximations', *J. Numer. Math.* **13**, 119–156.

W. Hackbusch, B. Khoromskij and E. E. Tyrtyshnikov (2008), 'Approximate iterations for structured matrices', *Numer. Math.* **109**, 365–383.

S. Handschuh (2012), Changing the topology of tensor networks. Preprint 15, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig.

R. Harshman (1970), 'Foundations of PARAFAC procedure: Models and conditions for an "exploratory" multi-mode analysis', *UCLA Working Papers in Phonetics* **16**, 1–84.

J. Håstad (1990), 'Tensor rank is NP-complete', *J. Algorithms* **11**, 644–654.

R. Henrion (1994), '$N$-way principal component analysis: Theory, algorithm and applications', *Chemometrics Intelligent Lab. Syst.* **25**, 1–23.

N. J. Higham (2008), *Functions of Matrices: Theory and Computation*, SIAM.

F. L. Hitchcock (1927), 'The expression of a tensor or a polyadic as a sum of products', *J. Math. Phys.* **6**, 164–189.

S. Holtz, T. Rohwedder and R. Schneider (2012*a*), 'On manifolds of tensors of fixed TT-rank', *Numer. Math.* **120**, 701–731.

S. Holtz, T. Rohwedder and R. Schneider (2012*b*), 'The alternating linear scheme for tensor optimization in the tensor train format', *SIAM J. Sci. Comput.* **34**, A683–A713.

T. D. Howell (1978), 'Global properties of tensor rank', *Linear Algebra Appl.* **22**, 9–23.

R. Hübener, V. Nebendahl and W. Dür (2010), 'Concatenated tensor network states', *New J. Phys.* **12**, 025004.
T. Huckle, K. Waldherr and T. Schulte-Herbrüggen (2013), 'Computations in quantum tensor networks', *Linear Algebra Appl.* **438**, 750–781.
M. Ishteva, L. De Lathauwer, P.-A. Absil and S. Van Huffel (2009), 'Differential-geometric Newton method for the best rank-$(r_1, r_2, r_3)$ approximation of tensors', *Numer. Algorithms* **51**, 179–194.
B. Khoromskij (2009), 'Tensor-structured preconditioners and approximate inverse of elliptic operators in $\mathbb{R}^d$', *Constr. Approx.* **30**, 599–620.
B. Khoromskij (2011), '$O(d \log N)$-quantics approximation of $N - d$ tensors in high-dimensional numerical modeling', *Constr. Approx.* **34**, 257–280.
B. Khoromskij and C. Schwab (2011), 'Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs', *SIAM J. Sci. Comput.* **33**, 364–385.
B. Khoromskij, V. Khoromskaia and H.-J. Flad (2011), 'Numerical solution of the Hartree–Fock equation in multilevel tensor-structured format', *SIAM J. Sci. Comput.* **33**, 45–65.
O. Koch and C. Lubich (2010), 'Dynamical tensor approximation', *SIAM J. Matrix Anal. Appl.* **31**, 2360–2375.
T. G. Kolda and B. W. Bader (2009), 'Tensor decompositions and applications', *SIAM Rev.* **51**, 455–500.
D. Kressner and C. Tobler (2010), 'Krylov subspace methods for linear systems with tensor product structure', *SIAM J. Matrix Anal. Appl.* **31**, 1688–1714.
D. Kressner and C. Tobler (2011*a*), 'Low-rank tensor Krylov subspace methods for parametrized linear systems', *SIAM J. Matrix Anal. Appl.* **32**, 1288–1316.
D. Kressner and C. Tobler (2011*b*), 'Preconditioned low-rank methods for high-dimensional elliptic PDE eigenvalue problems', *Comput. Methods Appl. Math.* **11**, 363–381.
D. Kressner and C. Tobler (2012), htucker: A Matlab toolbox for tensors in hierarchical Tucker format. Technical report, MATHICSE, EPF Lausanne.
P. M. Kroonenberg (2008), *Applied Multiway Data Analysis*, Wiley.
J. B. Kruskal (1977), 'Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics', *Linear Algebra Appl.* **18**, 95–138.
J. M. Landsberg (2012), *Tensors: Geometry and Applications*, AMS.
C. Lubich (2005), 'On variational approximations in quantum molecular dynamics', *Math. Comp.* **74**, 765–779.
C. Lubich (2008), *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*, EMS, Zürich.
C. Lubich and I. V. Oseledets (2013), A projector-splitting integrator for dynamical low-rank approximation. arXiv 1301.1058
C. Lubich, T. Rohwedder, R. Schneider and B. Vandereycken (2013), 'Dynamical approximation of hierarchical Tucker and tensor-train tensors', *SIAM J. Matrix Anal. Appl.* **34**, 470–494.
M. J. Mohlenkamp (2010), 'A center-of-mass principle for the multiparticle Schrödinger equation', *J. Math. Phys.* **51**, 022112.

M. J. Mohlenkamp (2013), 'Musing on multilinear fitting', *Linear Algebra Appl.* **438**, 834–852.

I. V. Oseledets (2010), 'Approximation of matrices using tensor decomposition', *SIAM J. Matrix Anal. Appl.* **31**, 2130–2145.

I. V. Oseledets (2011*a*), 'DMRG approach to fast linear algebra in the TT-format', *Comput. Methods Appl. Math.* **11**, 382–393.

I. V. Oseledets (2011*b*), 'Tensor-train decomposition', *SIAM J. Sci. Comput.* **33**, 2295–2317.

I. V. Oseledets and E. E. Tyrtyshnikov (2005), 'Approximate inversion of matrices in the process of solving a hypersingular integral equation', *Comput. Math. Math. Phys.* **45**, 302–313.

I. V. Oseledets and E. E. Tyrtyshnikov (2009*a*), 'Breaking the curse of dimensionality, or how to use SVD in many dimensions', *SIAM J. Sci. Comput.* **31**, 3744–3759.

I. V. Oseledets and E. E. Tyrtyshnikov (2009*b*), Tensor tree decomposition does not need a tree. Preprint 2009-08, RAS, Moscow.

I. V. Oseledets, D. V. Savostyanov and E. E. Tyrtyshnikov (2009), 'Linear algebra for tensor problems', *Computing* **85**, 169–188.

I. V. Oseledets, D. V. Savostyanov and E. E. Tyrtyshnikov (2010), 'Cross approximation in tensor electron density computations', *Numer. Linear Algebra Appl.* **17**, 935–952.

I. V. Oseledets, E. E. Tyrtyshnikov and N. L. Zamarashkin (2011), 'Tensor-train ranks for matrices and their inverses', *Comput. Methods Appl. Math.* **11**, 394–403.

B. Savas and L. Eldén (2013), 'Krylov-type methods for tensor computations I', *Linear Algebra Appl.* **438**, 891–918.

D. V. Savostyanov, E. E. Tyrtyshnikov and N. L. Zamarashkin (2012), 'Fast truncation of mode ranks for bilinear tensor operations', *Numer. Linear Algebra Appl.* **19**, 103–111.

C. Schwab and C. J. Gittelson (2011), Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. In *Acta Numerica*, Vol. 20, Cambridge University Press, pp. 291–467.

Y. Shi, L. Duan and G. Vidal (2006), 'Classical simulation of quantum many-body systems with a tree tensor network', *Phys. Rev.* A **74**, 022320.

A. Smilde, R. Bro and P. Geladi (2004), *Multi-way Analysis: Applications in the Chemical Sciences*, Wiley.

M. Sørensen, L. De Lathauwer, P. Comon, S. Icart and L. Deneire (2012), 'Canonical polyadic decomposition with a columnwise orthonormal factor matrix', *SIAM J. Matrix Anal. Appl.* **33**, 1190–1213.

A. Stegeman (2010), 'On the uniqueness of the $n$th order tensor decomposition into rank-1 terms with linear independence in one mode', *SIAM J. Matrix Anal. Appl.* **31**, 2498–2516.

F. Stenger (1993), *Numerical Methods Based on Sinc and Analytic Functions*, Springer.

V. Strassen (1969), 'Gaussian elimination is not optimal', *Numer. Math.* **13**, 354–356.

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

A. Uschmajew (2010), 'Convex maximization problems on non-compact Stiefel manifolds with application to orthogonal tensor approximations', *Numer. Math.* **115**, 309–331.

A. Uschmajew (2012), 'Local convergence of the alternating least squares algorithm for canonical tensor approximation', *SIAM J. Matrix Anal. Appl.* **33**, 639–652.

F. Verstraete and J. I. Cirac (2006), 'Matrix product states represent ground states faithfully', *Phys. Rev.* B **73**, 094423.

S. Winograd (1971), 'On multiplication of $2 \times 2$ matrices', *Linear Algebra Appl.* **4**, 381–388.

H. Yserentant (2010), *Regularity and Approximability of Electronic Wave Functions*, Vol. 2000 of *Lecture Notes in Mathematics*, Springer.