COMPUTATIONAL METHODS IN APPLIED MATHEMATICS, Vol. 6 (2006), No. 2, pp. 194–220

STRUCTURED RANK-$(R_1, \ldots, R_D)$ DECOMPOSITION OF FUNCTION-RELATED TENSORS IN $\mathbb{R}^D$

B. N. KHOROMSKIJ[1]

Abstract — The structured tensor-product approximation of multidimensional nonlocal operators by a two-level rank-$(r_1, \ldots, r_d)$ decomposition of related higher-order tensors is proposed and analysed. In this approach, the construction of the desired approximant to a target tensor is a reminiscence of the *Tucker-type model*, where the canonical components are represented in a fixed (uniform) basis, while the core tensor is given in the *canonical* format. As an alternative, the multilevel nested canonical decomposition is presented. The complexity analysis of the corresponding multilinear algebra shows an almost linear cost in the one-dimensional problem size. The existence of a low Kronecker rank two-level representation is proven for a class of function-related tensors. In particular, we apply the results to d-th order tensors generated by the multivariate functions $|x|^{-2}$, $|x-y|^{-1}$, $e^{-\alpha|x-y|}$, $|x-y|^{-1}e^{-|x-y|}$ and $|x|^{\lambda}\mathrm{sinc}\,(|x|\,|y|)$ with $x, y \in \mathbb{R}^d$.

2000 Mathematics Subject Classification: 65F30, 65F50, 65N35, 65F10.

Keywords: structured matrices, $\mathcal{H}$-matrices, Kronecker products, higher-order tensors, multidimensional integral operators, Newton potential, matrix-valued functions.

1. Introduction

There is a wide range of traditional applications, as well as an increasing number of modern applications involving quantities described by higher-order tensors, which are, in fact, higher-order analogues of vectors and matrices. Naive numerical implementation of the corresponding multilinear algebra suffers from the so-called "curse of dimensionality" which can be relaxed by invoking various Kronecker product formats to represent the fully populated tensors that arise. As a result of over more than thirty year developments, nowadays we have several well-established concepts of data-sparse approximation to higher-order tensors, which are based either on the so-called Tucker model [38] or on the CANDECOMP/PARAFAC (CP) decomposition [3,24]. There are numerous successful applications of the Tucker and CP models in higher-order statistics, independent component analysis, chemometrics, telecommunications, signal processing, data mining, mathematical biology, complexity theory, and in many other fields (see [35] and references therein).

Given a d-th order tensor $\mathcal{A} = [a_{i_1 \ldots i_d}] \in \mathbb{C}^{\mathcal{I}}$, defined on the product index set $\mathcal{I} = I_1 \times \ldots \times I_d$, the Tucker model deals with the following approximation:

$$\mathcal{A}_{(\mathbf{r})} = \sum_{k_1=1}^{r_1} \ldots \sum_{k_d=1}^{r_d} b_{k_1 \ldots k_d} \times_1 V_{k_1}^{(1)} \times_2 \ldots \times_d V_{k_d}^{(d)} \equiv \sum_{\mathbf{k}=1}^{\mathbf{r}} b_{\mathbf{k}} \bigotimes_{\ell=1}^{d} V_{k_\ell}^{(\ell)} \approx \mathcal{A}, \tag{1.1}$$

[1] *Max-Planck-Institute for Mathematics in the Sciences*, Inselstr. 22-26, D-04103 Leipzig, Germany. E-mail: bokh@mis.mpg.de

where the Kronecker factors $V_{k_\ell}^{(\ell)} \in \mathbb{C}^{I_\ell}$ ($k_\ell = 1, \ldots, r_\ell$, $\ell = 1, \ldots, d$) are complex vectors of the respective size $n_\ell = |I_\ell|$, $\mathbf{r} = (r_1, \ldots, r_d)$ (the Tucker rank) and $b_{k_1 \ldots k_d} \in \mathbb{C}$ (see Fig. 1.1 visualising (1.1) for $d = 3$). Here and in the following we use the notation $\times_\ell$ to represent the canonical tensor $\mathcal{U} \equiv \{u_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{J}} = b \times_1 U^{(1)} \times_2 \ldots \times_d U^{(d)} \in \mathbb{C}^{\mathcal{J}}$, defined by $u_{i_1 \ldots i_d} = b \cdot u_{i_1}^{(1)} \cdots u_{i_d}^{(d)}$ with $U^{(\ell)} \equiv \{u_{i_\ell}^{(\ell)}\}_{i_\ell \in I_\ell} \in \mathbb{C}^{I_\ell}$.

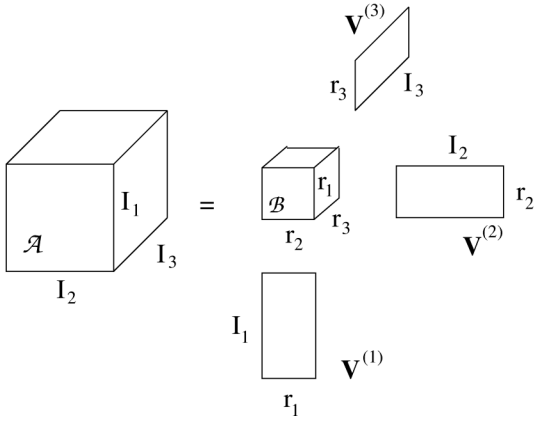


Fig. 1.1. Visualisation of the Tucker model for a third-order tensor

Without loss of generality, we assume that the vectors $\{V_{k_\ell}^{(\ell)}\}$ are orthonormal, i.e.,

$$\langle V_{k_\ell}^{(\ell)}, V_{m_\ell}^{(\ell)} \rangle = \delta_{k_\ell, m_\ell}, \quad k_\ell, m_\ell = 1, \ldots, r_\ell, \quad \ell = 1, \ldots, d,$$

where δ_{k_ℓ, m_ℓ} is Kronecker's delta. In the following, we denote by $\mathfrak{T}_{\mathbf{r}}$ the set of tensors parametrised by (1.1) (i.e., $\mathbf{V}^{(\ell)} = [V_1^{(\ell)} V_2^{(\ell)} \ldots V_{r_\ell}^{(\ell)}]$ is an orthogonal matrix for $\ell = 1, \ldots, d$). We use the shortening in notation

$$\mathcal{A}_{(\mathbf{r})} = \mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \ldots \times_d \mathbf{V}^{(d)}, \tag{1.2}$$

with tensors $\mathbf{V}^{(\ell)} \in \mathbb{R}^{I_\ell \times r_\ell}$ and $\mathcal{B} = \{b_{\mathbf{k}}\} \in \mathbb{R}^{r_1 \times \ldots \times r_d}$, where the latter is called the core tensor. Notice that the representation of elements $\mathcal{A} \in \mathfrak{T}_{\mathbf{r}}$ is still not unique due to the rotational uncertainty in the core tensor $\mathcal{B}$ (see Remark 2.2).

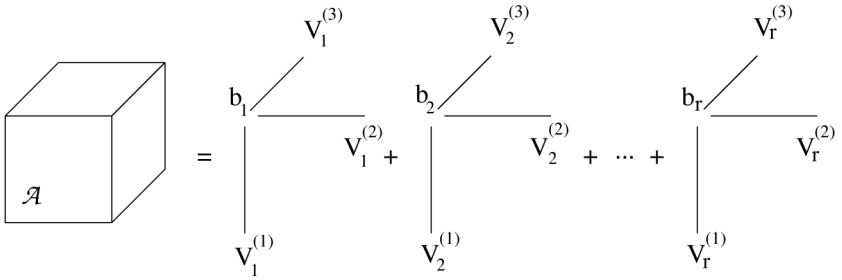
The canonical CP model is a simplified version of the *general* decomposition (1.1) (i.e., without orthogonality constraint) defined by

$$\mathcal{A}_{(r)} = \sum_{k=1}^{r} b_k \times_1 V_k^{(1)} \times_2 \ldots \times_d V_k^{(d)} \equiv \sum_{k=1}^{r} b_k \bigotimes_{\ell=1}^{d} V_k^{(\ell)} \approx \mathcal{A}, \quad b_k \in \mathbb{C}, \tag{1.3}$$

where the Kronecker factors $V_k^{(\ell)} \in \mathbb{C}^{I_\ell}$ are unit-norm vectors. Indeed, decomposition (1.3) can be viewed as a special case of the Tucker model (1.1), where $r = r_1 = \ldots = r_d$ and $b_{k_1 \ldots k_d} = 0$ unless $k_1 = k_2 = \ldots = k_d$, i.e., only the superdiagonal of $\mathcal{B} = \{b_{\mathbf{k}}\}$ is nonzero. In this case, we say that $\mathcal{B} = \operatorname{diag}\{b_1, \ldots, b_r\}$ with $b_k = b_{k \ldots k}$.

The trilinear CP-decomposition is visualised in Fig. 1.2. The minimal number r in representation (1.3) is called the Kronecker rank of a given tensor $\mathcal{A}_{(r)}$; under moderate assumptions, the corresponding decomposition is unique (see [29, 30]). We denote by $\mathfrak{C}_r$ the set of component-wise normalised tensors parametrised by (1.3) and by $\mathfrak{C}_r^{\perp} \subset \mathfrak{C}_r$ the corresponding subset of orthogonally decomposable tensors (i.e., the matrices $[V_1^{(\ell)}, \ldots, V_r^{(\ell)}]$ ($\ell = 1, \ldots, d$) are orthogonal). If we let $r = r_\ell$, $n = n_\ell$ ($\ell = 1, \ldots, d$), then both the CP- and Tucker

models require only drn numbers to represent the canonical components plus r (resp. r^d) memory units for the core tensor. Notice that the Tucker decomposition can be viewed as a CP representation with $r_{CP} = r_1 \cdots r_d$.



F i g. 1.2. Visualisation of the CP-decomposition for a third-order tensor

The methods of optimised data-sparsity discussed in the present paper are mainly oriented toward FEM/BEM applications for partial differential and integral equations in $\mathbb{R}^d$, many-particle modelling based on the electronic Schrödinger [41] and Hartree-Fock equations, evolution Schrödinger equations [27, 32] (molecular dynamics), financial mathematics and stochastic PDEs [33, 34], operator equations in Banach spaces [12, 13] etc. In such applications the parameter d is usually associated with the spatial dimension.

The feasibility of any tensor decomposition method is naturally limited by:

(a) robustness and complexity of computing the canonical components;

(b) data-sparsity and complexity of tensor-tensor operations;

(c) its range of applications.

Clearly, the efficiency of numerical implementation crucially depends on the particular purposes of the multilinear algebra (say, the component identification, matrix-vector or matrix-matrix operations, computation of matrix-valued functions, etc.).

The main computational problem is the approximation of a given higher-order tensor $\mathcal{A}_0$ in a certain set of low-rank structured tensors $\mathbf{S}$. In particular, $\mathbf{S}$ may be one of the classes $\mathfrak{T}_{\mathbf{r}}$, $\mathfrak{C}_r$ or $\mathfrak{C}_r^{\perp}$. The most general approach is to derive the components of $\mathcal{A}_{(\mathbf{r})}$ (resp. $\mathcal{A}_{(r)}$) by straightforward minimisation of the quadratic *cost functional* $f(\mathcal{A}) := \|\mathcal{A} - \mathcal{A}_0\|^2$,

$$\mathcal{A}_{(\mathbf{r})} = \operatorname{argmin} \|\mathcal{A} - \mathcal{A}_0\|^2, \tag{1.4}$$

over all rank-$\mathbf{r}$ (resp. rank-r) tensors $\mathcal{A} \in \mathbf{S}$. Here and in the following we make use of the Frobenius (energy) norm $\|\mathcal{A}\| := \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$ induced by the inner product $\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{(i_1,\ldots,i_d)\in\mathcal{J}} a_{i_1\ldots i_d} b_{i_1\ldots i_d}$. The maximum-norm is defined by $\|\mathcal{A}\|_\infty := \max_{\mathbf{i}\in\mathcal{J}} |a_{\mathbf{i}}|$. Relying on the beneficial features of the tensor-product arithmetic (see [26]), the approximation process can be simplified dramatically if the target tensor $\mathcal{A}_0$ already has a data-sparse representation, say, $\mathcal{A}_0 \in \mathfrak{T}_{\mathbf{R}}$ or $\mathcal{A}_0 \in \mathfrak{C}_R$ with $R \ll n^d$ (resp. $\max_\ell R_\ell \ll n$) (see the discussion in Section 3.3).

There are *algebraic, analytically-based and combined strategies* for computing a Kronecker tensor-product decomposition of a higher-order tensor. *Algebraic* methods are the most general ones. However, they are known to have intrinsic limitations due to (a) multiple local minima of the cost functional, (b) degeneracy of a minimiser (typical of the CP model), and (c) high-dimensional nonlinear optimisation. *Analytically-based representation* methods are efficient for the special class of *function-related* operators/tensors, while all the difficulties accompanying algebraic methods are addressed as soon as one requires further tensor-tensor

operations. *Combined* methods are designed to take advantage of both algebraic and analytic approaches, and, at the same time, to relax their limitations.

Tensor decompositions can be beneficially combined with different well-structured matrix/tensor formats which may include $\mathcal{H}$-matrices, low-rank, Toeplitz/circulant or Fourier based matrices, wavelet sparsity etc. (see the Table). The hierarchical Kronecker tensor-product (HKT) matrix format (see [20, 22]) provides an extension of the CP-model to the case of matrices associated with higher-order tensors. The key idea of the analytically-based HKT-model is a data-sparse hierarchical representation of low-dimensional matrix components corresponding to the canonical factors. The *analytic component estimates* are mainly based on the approximation of a *univariate* generating function by exponential sums.

Structured Kronecker product models (d-th order tensors of size n^d)

Model	Notation	Memory/$A \cdot x$	$A \cdot B$	Comp. tools
Canonical - CP	$\mathfrak{C}_r$	drn	drn^2	ALS/Newton
HKT - CP	$\mathfrak{C}_{\mathcal{H},r}$	$dr\sqrt{n}\log^q n$	$drn\log^q n$	Analytic (quadr.)
Nested - CP	$\mathfrak{C}_{T(I),L}$	$dr^{\log d}n + rd$	$dr^{\log d}n$	SVD/QR/orthog. iter.
Tucker	$\mathfrak{T}_{\mathbf{r}}$	$r^d + drn$	—	Orthogonal ALS
Two-level Tucker	$\mathfrak{T}_{(\mathcal{U},\mathbf{r},q)}$	drq/drr_0qn^2	dr^2q^2 (mem.)	Analyt.(interp.)+ CP

In the present paper we propose two combined models, which extend the general CP and Tucker decompositions to the case of matrices associated with higher-order tensors. First, we introduce the so-called multilevel nested CP model that generally exhibits a higher tensor rank, but provides robust numerical schemes since it avoids nonlinear minimisation (see Section 2.2). This tensor format is denoted by $\mathfrak{C}_{T(\mathcal{I}),L}$, where L is the depth of recursion corresponding to the product cluster tree $T(\mathcal{I})$.

In the second approach, using a combined strategy, we extend the general Tucker model to the case of matrices associated with higher-order tensors by imposing a certain *a priori* structure on the canonical components (e.g., representation in a uniform basis) and combining this with the rank-q CP decomposition of the core tensor. This tensor class will be denoted by $\mathfrak{T}_{(\mathcal{U},\mathbf{r},q)}$, so that

$$\mathfrak{T}_{(\mathcal{U},\mathbf{r},q)} \subset \mathfrak{T}_{\mathbf{r}}, \quad \text{with } \mathcal{B} \in \mathfrak{C}_q.$$

For each $\mathcal{A} \in \mathfrak{T}_{(\mathcal{U},\mathbf{r},q)}$, the core tensor $\mathcal{B} \in \mathfrak{C}_q$ of size $\mathbb{R}^{r_1 \times \cdots \times r_d}$ contains at most $q(r_1 + \ldots + r_d)$ entries (representation coefficients) instead of $r_1 r_2 \cdots r_d$. In the applications considered in this paper, we are able to prove that $r = \max_\ell r_\ell = O(|\log \varepsilon| \log n)$. Since $\mathcal{B}$ is represented by a rank-q CP model, the overall complexity is of order $O(rqd)$ independently of n. Another version of the two-level structure is specified by the choice of $\mathcal{B} \in \mathfrak{T}_{\mathbf{q}}$.

In this approach the canonical components can be constructed analytically dwelling upon the tensor-product *sinc*-interpolation of the *multivariate* (generating) function. Since the core tensor is supposed to have a small size, its CP decomposition can be performed by combining analytic approximations (initial guess) with nonlinear minimisation methods.

Motivations for developing the $\mathfrak{T}_{(\mathcal{U},\mathbf{r},q)}$-model are its applicability to a more general class of discrete nonlocal operators (not necessarily generated by shift-invariant kernels), the possibility to reduce its numerical complexity (logarithmic in both n and ε), and at the same time, it seems to provide more flexibility in the construction of efficient numerical implementations (the canonical decomposition applies only to *low-dimensional* tensors). Notice that all of our constructions can be generalised to complex-valued tensors.

This paper mainly focuses on the following issues:

• brief survey of modern tensor decomposition techniques;

• construction of a two-level rank-$(r_1, \ldots, r_d)$ matrix/tensor format based on model (1.1) as well as on the multilevel nested CP model;

• complexity analysis of matrix/tensor arithmetic in the proposed formats;

• rank estimates for the considered decompositions applied to a class of function-generated tensors, provided with numerical experiments.

The Table presents the computational characteristics of different tensor-product models considered in this paper. Here ALS means the alternating least squares iteration.

The rest of the paper is organised as follows. Section 2 describes the L-level nested rank-$(r_1, \ldots, r_L)$ and the two-level rank-$(r_1, \ldots, r_d)$ Tucker tensor formats, addresses complexity issues and discusses some numerical aspects of the traditional and new approaches. In particular, in Section 2.1, we recall the main features of the HKT matrix decomposition. Section 3 addresses the computational aspects of the CP and the Tucker models. The main result is presented in Section 3.2 which derives the Lagrange equations for the cost functional specifying the orthogonal Tucker decomposition. In Section 4, we apply the results to d-th order tensors generated by the multivariate functions $|x|^{-2}$, $|x-y|^{-1}$, $e^{-\alpha|x-y|}$ and $|x-y|^{-1}e^{-|x-y|}$ with $x, y \in \mathbb{R}^d$, and present some numerical illustrations.

2. Nested CP and rank-$(r_1, \ldots, r_d)$ decompositions

In Section 2, we introduce and analyse a class of nested rank-$(r_1, \ldots, r_L)$ and two-level rank-$(r_1, \ldots, r_d)$ tensor decompositions based on the structured versions of the CP and Tucker models, respectively.

2.1. Rank-r matrix decomposition via the HKT-format. We consider the representation problem for a class of real-valued square matrices related to discrete multidimensional operators posed in $\mathbb{R}^d$, such that $A \in \mathbb{R}^{N\times N}$, $N = n^d$. In general, such matrices can be interpreted as high-order fully populated tensors, which makes the standard matrix arithmetic almost unfeasible. To overcome this difficulty, one needs numerically tractable data-sparse representations of high-dimensional tensors that arise.

The hierarchical Kronecker tensor-product (HKT) format as proposed in [20, 22] reads

$$A = \sum_{k=1}^{r} b_k V_k^{(1)} \otimes \ldots \otimes V_k^{(d)}, \quad b_k \in \mathbb{R}, \tag{2.1}$$

where the Kronecker factors $V_k^{(\ell)} \in \mathbb{R}^{n\times n}$ are $\mathcal{H}$-matrices (see [15 – 19] for the definition, approximation properties and applications of $\mathcal{H}$-matrices). We recall that the Kronecker product of matrices $A \otimes B$ is defined as a block matrix $[a_{ij}B]$, provided that $A = [a_{ij}]$. The operation "$\otimes$" can be applied to arbitrary rectangular matrices (in particular, to row or column vectors) and in the multifactor version as in (2.1).

We write $A \in \mathrm{HKT}_{(r,s)}$ if A is of the form of (2.1) and $V_k^{(\ell)}$ have a hierarchical block partitioning (independent of k) with blocks of rank at most s. Again the minimal number of Kronecker-product terms r involved is referred to as the *Kronecker rank* of the matrix (tensor) at issue.

The approximations of function-related matrices by matrices of the form (2.1) were first studied in [22, 39]. The main results of these papers are the estimates of the form $r =$

$O(\log^2\varepsilon)$ and $r = O(|\log\varepsilon|\log n)$, where ε is a prescribed approximation accuracy. However, if there is no structure in the Kronecker factors, then the storage is $O(drn^2)$, while the matrix-matrix complexity is $O(dr^2n^3)$, which is far from being satisfactory. A possible remedy is the hierarchical ($\mathcal{H}$-matrix) approximation to the Kronecker factors (HKT-approximations) with the advantage of rigorously proved existence theorems [22] with estimates of the form $r = O(\log^2\varepsilon)$, $s = O(\log\varepsilon^{-1})$.

If $A \in \mathrm{HKT}_{(r,s)}$, then only $V_k^{(\ell)}$ need to be stored. Since, by definition, they have the $\mathcal{H}$-format, we arrive at the following complexity bounds (the linear complexity is $O(n^d)$):

- the storage for A is $O(drsn\log n)$, indicating the *superlinear compression property*;
- multiplication of A by a rank-r_1 vector $x = \sum_{k=1}^{r_1} b_k x_k^{(1)} \otimes \ldots \otimes x_k^{(d)}$ requires $O(drr_1sn \times \log n)$ operations;
- the complexity of the matrix-matrix multiplication is $O(dr^2n\log^q n)$.

In this paper we prove the existence results for the low Kronecker rank approximations for the general class of matrices related to functions that are characterised in terms of their Laplace transform.

2.2. Multilevel nested rank-$(r_1,\dots,r_L)$ CP model with $L = \log_2 d$. In this section, we discuss the decomposition of general tensors $\mathcal{A} \in \mathbb{R}^{\mathcal{I}}$. As in the previous section, in the matrix case, we use the standard Kronecker product operation $\otimes$ instead of $\times_\ell$. For ease of presentation, we assume that $d = 2^L$ with $L = \log_2 d \in \mathbb{N}$. The generalisation to the case of arbitrary $d \geqslant 2$ is straightforward.

Let $I_d := \{1,\dots,2^L\}$ and let $T(I_d)$ be a binary cluster tree of I_d. On the tensor-product index set $\mathcal{I}$, we introduce the associated cluster tree $T(\mathcal{I})$, such that for each $\sigma \in T(I_d)$ we define $\tau_\sigma \in T(\mathcal{I})$ by $\tau_\sigma = \times_{\ell\in\sigma} I_\ell$. By construction, we have $T(\mathcal{I}) := \{T^{(\ell)}(\mathcal{I})\}_{\ell=0}^{L}$, where $T^{(\ell)}(\mathcal{I}) := \{\tau_i^{(\ell)} : 1 \leqslant i \leqslant 2^\ell\}$ with standard definition for the clusters $\tau_i^{(\ell)}$ on level ℓ,

$$\tau_i^{(\ell)} := \{(i-1)*2^{L-\ell}, (i-1)*2^{L-\ell}+1, \dots, i*2^{L-\ell}-1\}.$$

In particular, $T^{(0)}(\mathcal{I}) = \mathcal{I}$ is the root, $T^{(L)}(\mathcal{I})$ forms the set of leaves and $\tau_i^{(\ell)} = \tau_{2i-1}^{(\ell+1)} \times \tau_{2i}^{(\ell+1)}$, where $\tau_{2i-1}^{(\ell+1)}$, $\tau_{2i}^{(\ell+1)}$ are the sons of $\tau_i^{(\ell)}$ ($\ell < L$).

Assume we are given a cluster tree $T(\mathcal{I})$ of depth L and the set of rank parameters $r_1,\dots,r_L \in \mathbb{N}$. We introduce the multilevel nested CP model as a class of tensors $\mathcal{A}^{(0)} \in \mathbb{R}^{T^{(0)}(\mathcal{I})}$, defined via a recursive sequence of two-fold Kronecker products, which finally builds the nested decomposition.

Given the components on level L (i.e., on the leaves), we proceed recursively, i.e., for each $\ell = 1,\dots,L$, we represent the corresponding Kronecker factors $U_{m_{\ell-1}}^{(j)} \in \mathbb{R}^{\tau_j^{(\ell-1)}}$, $1 \leqslant j \leqslant 2^{\ell-1}$, by

$$U_{m_{\ell-1}}^{(j)} = \sum_{m_\ell=1}^{r_\ell} \mu_{m_\ell}^{(j)}\, U_{m_\ell}^{(2j-1)} \otimes U_{m_\ell}^{(2j)} \quad \text{with} \quad U_{m_\ell}^{(2j-1)} \in \mathbb{R}^{\tau_{2j-1}^{(\ell)}},\ U_{m_\ell}^{(2j)} \in \mathbb{R}^{\tau_{2j}^{(\ell)}}, \tag{2.2}$$

where $\tau_{2j-1}^{(\ell)}$ and $\tau_{2j}^{(\ell)}$ are the sons of $\tau_j^{(\ell-1)}$, and $\mu_{m_\ell}^{(j)} \in \mathbb{R}$. On zero level, i.e., for $\ell = 0$, define

$$\mathcal{A}^{(0)} \equiv U_{m_0}^{(0)} = \sum_{m_1=1}^{r_1} \mu_{m_1}^{(1)}\, U_{m_1}^{(1)} \otimes U_{m_1}^{(2)} \quad \text{with} \quad U_{m_1}^{(1)} \in \mathbb{R}^{\tau_1^{(1)}},\ U_{m_1}^{(2)} \in \mathbb{R}^{\tau_2^{(1)}}, \tag{2.3}$$

where $\tau_1^{(1)}$ and $\tau_2^{(1)}$ are the sons of $\tau_1^{(0)} = T^{(0)}(\mathcal{I})$, and $\mu_{m_1}^{(1)} \in \mathbb{R}$.

We denote by $\mathfrak{C}_{T(\mathcal{I}),L}$ the set of tensors parametrised as above by the recursive construction of depth L.

Example 2.1. Let $d = 8$. Then $L = 3$ and a three-level nested CP model (2.2) – (2.3) reads as

$$\mathcal{A}^{(0)} = \sum_{m_1=1}^{r_1} \mu_{m_1}^{(1)} U_{m_1}^{(1)} \otimes U_{m_1}^{(2)}$$

with

$$U_{m_1}^{(1)} = \sum_{m_2=1}^{r_2} \mu_{m_2}^{(1)} \Big(\sum_{m_3=1}^{r_3} \mu_{m_3}^{(1)} U_{m_3}^{(1)} \otimes U_{m_3}^{(2)} \Big) \otimes \Big(\sum_{m_3=1}^{r_3} \mu_{m_3}^{(2)} U_{m_3}^{(3)} \otimes U_{m_3}^{(4)} \Big)$$

$$U_{m_1}^{(2)} = \sum_{m_2=1}^{r_2} \mu_{m_2}^{(2)} \Big(\sum_{m_3=1}^{r_3} \mu_{m_3}^{(3)} U_{m_3}^{(5)} \otimes U_{m_3}^{(6)} \Big) \otimes \Big(\sum_{m_3=1}^{r_3} \mu_{m_3}^{(4)} U_{m_3}^{(7)} \otimes U_{m_3}^{(8)} \Big).$$

In the matrix case, we just have $\mathcal{I} = \mathcal{I}_x \times \mathcal{I}_y$ and $T(\mathcal{I}) := \{T^{(\ell)}(\mathcal{I})\}_{\ell=0}^{L}$ with $T^{(\ell)}(\mathcal{I}) = T^{(\ell)}(\mathcal{I}_x) \times T^{(\ell)}(\mathcal{I}_y)$. The corresponding matrix $A \in \mathfrak{C}_{T(\mathcal{I}),L}$ operates as $A : \mathbb{R}^{\mathcal{I}_x} \to \mathbb{R}^{\mathcal{I}_y}$.

Obviously, the Kronecker rank of the recursive decomposition (2.3) is $r_1 r_2^{2^1} r_3^{2^2} \dots r_L^{2^{L-1}}$ (that is r^{d-1} if $r = r_1 = \dots = r_L$).

Definition 2.1. A d-th order tensor is called *supersymmetric* if it is invariant under arbitrary permutations of indices in $\{1, \dots, d\}$.

Examples of supersymmetric tensors include those generated by translation invariant functions (see Section 4).

The following Lemma shows the effective rank reduction via (2.2), (2.3).

Lemma 2.1. *The storage for $\mathcal{A} \in \mathfrak{C}_{T(\mathcal{I}),L}$ can be estimated by $O(dr_1 \dots r_L n + \sum_{\ell=1}^{L} r_\ell \times 2^{\ell-1})$, while for a supersymmetric tensor we arrive at the memory consumption $O(r_1 \dots r_L n + \sum_{\ell=1}^{L} r_\ell)$.*

In the matrix case, multiplication of $\mathcal{A} \in \mathfrak{C}_{T(\mathcal{I}),L}$ by a rank-r_0 vector requires $O(dr_0 r_1 \dots \dots r_L n^2)$ operations. Let $A_1, A_2 \in \mathfrak{C}_{T(\mathcal{I}),L}$ and suppose $r = r_1 = \dots = r_L$, then both $A_1 A_2$ and the Hadamard matrix product $A_1 \odot A_2$ can be computed (and stored) in $O(dr^{2L} n^3)$ operations.

Proof. The storage demand is a direct consequence of the L-level recursive construction. Let $\mathbf{x} \in \mathfrak{C}_{T(\mathcal{I}_x),L}$ and suppose (for ease of exposition) that $\mathbf{x} = x_1 \otimes \dots \otimes x_d$, i.e., $r_1 = \dots \dots = r_L = 1$ holds. Then the product $A\mathbf{x}$ can be calculated recursively in L steps starting from the leaves of $T(\mathcal{I})$, which proves the desired matrix-vector complexity. In the case of general rank, similar arguments apply.

To analyse the matrix-matrix operations, we notice that the product of two rank-$(r_1, \dots \dots, r_L)$ matrices is a rank-$(r_1^2, \dots, r_L^2)$ matrix. Since the matrix-matrix product on the leaves (level L) costs $O(n^3)$, the result follows. □

Notice that any tensor $\mathcal{A} \in \mathfrak{C}_r$ (see the CP decomposition (1.3)) is a particular case of the rank-$(r_1, \dots, r_L)$ representation with $r_1 = r$, $r_2 = \dots = r_L = 1$. Hence, the existence of a low-rank CP approximation also implies the existence of a low-rank nested CP decomposition. Moreover, it provides a lower bound for r, $r_1 \leqslant r$ (r_1 is computable by SVD). For efficient SVD calculation via subspace iteration, the first-level rank r_1 can be a priori estimated based on some analytic arguments. However, a possible drawback of the nested decomposition is the expensive SVD since in general the parameters $r_2, \dots, r_L$ cannot be estimated a priori. Thus, the numerical efficiency of this heuristic model generally depends on the specific application.

2.3. General rank-$(r_1, \dots, r_d)$ matrix decomposition. Let $A \in \mathbb{R}^{\mathcal{I} \times \mathcal{I}}$ be a real-valued matrix defined on the index set $\mathcal{I} := I_1 \times \dots \times I_d$ with $I_\ell = \{1, \dots, n_\ell\}$. The matrix A can be multiplied by the vector $X \in \mathbb{R}^{\mathcal{I}}$, so that $AX \in \mathbb{R}^{\mathcal{I}}$.

The matrix A (resp. a vector X) can also be regarded as a d-th order tensor $\mathcal{A} \in \mathbb{R}^{I_1^2 \times \ldots \times I_d^2}$ (resp. $\mathcal{X} \in \mathbb{R}^{I_1 \times \ldots \times I_d}$).

We make use of the multiindex notation $\mathbf{i} := (i_1, \ldots, i_d) \in \mathcal{I}$.

Definition 2.2. *We introduce the following rank-$(r_1, \ldots, r_d)$ tensor-product matrix format:*

$$\mathcal{A} = \sum_{k_1=1}^{r_1} \ldots \sum_{k_d=1}^{r_d} b_{k_1 \ldots k_d} V_{k_1}^{(1)} \otimes \ldots \otimes V_{k_d}^{(d)} \in \mathbb{R}^{I_1^2 \times \ldots \times I_d^2}, \tag{2.4}$$

where the Kronecker factors $V_{k_\ell}^{(\ell)} \in \mathbb{R}^{I_\ell \times I_\ell}$, $k_\ell = 1, \ldots, r_\ell$, $\ell = 1, \ldots, d$, are matrices of a certain structure (say, $\mathcal{H}$-matrix, wavelet based format, Toeplitz/circulant, low-rank, etc.). Here $\mathbf{r} = (r_1, \ldots, r_d)$ is again called the Kronecker rank.

The matrix representation by the format (2.4) is a model reduction which is a generalisation of the low-rank approximation of matrices, corresponding to the case $d = 2$.

Remark 2.1. The matrix representation (2.4) is a reminiscence of the Tucker decomposition of multidimensional tensors (see (1.1)), while (2.1) comply with the CP model (see (1.3)).

With the help of the so-called n-mode tensor-matrix product (see [5]), we introduce the short notation

$$\mathcal{A} = \mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \ldots \times_d \mathbf{V}^{(d)} \equiv \mathcal{B} \times_{\mathbf{r}} \{\mathbf{V}\}, \tag{2.5}$$

with tensors $\mathbf{V}^{(\ell)} \in \mathbb{R}^{I_\ell \times I_\ell \times r_\ell}$ and $\mathcal{B} = \{b_{\mathbf{k}}\} \in \mathbb{R}^{r_1 \times \ldots \times r_d}$, where the latter is the core tensor. We denote matrices by uppercase letters, e.g., A, and tensors by calligraphic letters, e.g., $\mathcal{B}$. Besides, we set $\mathbf{A} = [A_1 A_2 \ldots]$, where A_i is the ith column matrix/vector of $\mathbf{A}$, e.g., $\mathbf{V}^{(\ell)} = [V_1^{(\ell)} V_2^{(\ell)} \ldots V_{r_\ell}^{(\ell)}]$. Notice that in the case $d = 2$, representation (2.5) is a multilinear equivalent of the matrix factorisation, i.e., we have

$$A = \mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \equiv \mathbf{V}^{(1)} \cdot \mathcal{B} \cdot {\mathbf{V}^{(2)}}^\top, \quad \mathcal{B} \in \mathbb{R}^{r_1 \times r_2}.$$

Similarly to the class of tensors $\mathcal{T}_{\mathbf{r}}$ in the Tucker model, i.e., if the components $V_{k_\ell}^{(\ell)} \in \mathbb{R}^{I_\ell}$ in (1.1) are mutually orthogonal vectors of arbitrary structure ($\mathbf{V}^{(\ell)} \in \mathbb{R}^{I_\ell \times r_\ell}$ are orthogonal matrices), we introduce the notation $A \in \mathcal{T}_{\mathcal{H},\mathbf{r}}$ for the multilinear matrix class with the canonical components having hierarchical structure.

Remark 2.2. The orthonormality assumption in the Tucker model (1.1) is not constraint. But even with this assumption, the core tensor is still not uniquely defined since each transformation

$$\mathbf{V}^{(\ell)} \to \widetilde{\mathbf{V}}^{(\ell)} := S^{(\ell)} \mathbf{V}^{(\ell)}, \quad \mathcal{B} \to \widetilde{\mathcal{B}} := \mathcal{B} \times_1 {S^{(1)}}^\top \times \ldots \times_d {S^{(d)}}^\top$$

represents the same tensor $\mathcal{A}$ for any choice of orthogonal matrices $S^{(\ell)} \in \mathbb{R}^{r_\ell \times r_\ell}$, $\ell = 1, \ldots, d$. In turn, due to the orthogonality requirement, the CP model (1.3) can be retrieved from (1.1) only in the special case of orthogonally decomposable tensors.

Clearly, we have

$$\mathcal{C}_r = \mathcal{T}_{\mathbf{r}} \ \text{ if } r = 1; \quad \mathcal{C}_r^\perp \subset \mathcal{T}_{\mathbf{r}}, \ \ \mathcal{C}_r \not\subset \mathcal{T}_{\mathbf{r}} \ \text{ if } r \geqslant 2.$$

Hence, in general, the CP decomposition (1.3) cannot be retrieved by rotation and "diagonal" truncation of the Tucker model.

We simplify the complexity analysis and set $r_\ell = r$, $n_\ell = n$ ($\ell = 1, \ldots, d$); the general case can be treated completely similarly.

Lemma 2.2. *The storage cost for $A \in \mathfrak{T}_{\mathcal{H},\mathbf{r}}$ is estimated by $O(drsn\log n)+r^d$, while for the supersymmetric tensor we arrive at the memory consumption $O(rsn\log n)+r^d/d$.*

Multiplication by a rank-r_0 vector requires $O(drr_0sn\log n)$ operations. Let $A_1, A_2 \in \mathfrak{T}_{\mathcal{H},\mathbf{r}}$. Then both the A_1A_2 and the Hadamard matrix product $A_1 \odot A_2$ can be computed and stored in $O(dr^2s^2n\log n)+r^{2d}$ operations.

Proof. The storage requirement for A is trivial. Let $\mathcal{X} = x_1 \times_2 \ldots \times_d x_d$ with $x_\ell \in \mathbb{R}^{I_\ell}$. Then $\mathcal{A}\mathcal{X} \equiv \mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \ldots \times_d \mathbf{V}^{(d)}\mathcal{X} = \mathcal{B} \times_1 \mathbf{V}^{(1)}x_1 \times_2 \mathbf{V}^{(2)}x_2 \ldots \times_d \mathbf{V}^{(d)}x_d$ implies the second assertion. Now we set $\mathcal{A}_1 = \mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \ldots \times_d \mathbf{V}^{(d)}$, $\mathcal{A}_2 = \mathcal{C} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_d \mathbf{U}^{(d)}$, to obtain the representation

$$\mathcal{A}_1\mathcal{A}_2 = \Big(\sum_{\mathbf{k}=1}^{\mathbf{r}} b_{\mathbf{k}} \bigotimes_{\ell=1}^{d} V_{k_\ell}^{(\ell)}\Big)\Big(\sum_{\mathbf{m}=1}^{\mathbf{r}} c_{\mathbf{m}} \bigotimes_{\ell=1}^{d} U_{m_\ell}^{(\ell)}\Big) = \sum_{\mathbf{k}=1}^{\mathbf{r}}\sum_{\mathbf{m}=1}^{\mathbf{r}} b_{\mathbf{k}}c_{\mathbf{m}} \bigotimes_{\ell=1}^{d} V_{k_\ell}^{(\ell)}U_{m_\ell}^{(\ell)},$$

which includes dr^2 canonical components and where the core tensor $\{b_{\mathbf{k}}c_{\mathbf{m}}\}$ has $(r^2)^d = r^{2d}$ entries. Analogously, for the Hadamard product we have

$$\mathcal{A}_1 \odot \mathcal{A}_2 = \Big(\sum_{\mathbf{k}=1}^{\mathbf{r}} b_{\mathbf{k}} \bigotimes_{\ell=1}^{d} V_{k_\ell}^{(\ell)}\Big) \odot \Big(\sum_{\mathbf{m}=1}^{\mathbf{r}} c_{\mathbf{m}} \bigotimes_{\ell=1}^{d} U_{m_\ell}^{(\ell)}\Big) = \sum_{\mathbf{k}=1}^{\mathbf{r}}\sum_{\mathbf{m}=1}^{\mathbf{r}} b_{\mathbf{k}}c_{\mathbf{m}} \bigotimes_{\ell=1}^{d} V_{k_\ell}^{(\ell)} \odot U_{m_\ell}^{(\ell)},$$

and take into account that the Hadamard product of two $\mathcal{H}$-matrices has a linear-logarithmic cost (see [26]). This completes our proof. □

As mentioned above, the CP decomposition (2.1) can be viewed as a special case of the generalised Tucker decomposition (2.4), where $r = r_1 = \ldots = r_d$ and only the superdiagonal of $\mathcal{B}$ is nonzero. In this case, we introduce the notation $A \in \mathfrak{C}_{\mathcal{I}\times\mathcal{J},r}$ or more specifically $A \in \mathfrak{C}_{\mathcal{H},r}$ if the canonical components are matrices of the generic or $\mathcal{H}$-matrix structure, respectively (in particular, we can now identify $HKT_{(r,s)} \equiv \mathfrak{C}_{\mathcal{H},r}$). Notice that from the numerical point of view, a CP decomposition can generally not be retrieved by "diagonal" truncation of the Tucker model (see Remark 2.2) since

(a) both formats might have rather different sets of canonical components $\{\mathbf{V}\}$,

(b) the corresponding decompositions usually realise distinct local minima of the cost functional (see Section 3).

In the following we introduce more economical formats by imposing a certain data-sparse structure on the core tensor $\mathcal{B}$ on the one hand and by specifying a priori the set of structured matrices $V_k^{(\ell)}$ in the format $\mathfrak{C}_{\mathcal{H},r}$ on the other.

2.4. Two-level rank-$(r_1,\ldots,r_d)$ decomposition. As soon as the tensor-product representation of function-related operators is concerned, both decompositions (2.1) and (2.4) rely on the deep connection to the separable approximation multivariate analytic/asymptotically smooth functions with point singularities (see [20,26] concerning (2.1)). In particular, (2.4) can be derived by using *tensor-product interpolation* with respect to some fixed system of approximating functions (say, *sinc*-functions, plain waves, interpolating wavelets) which indicates that one can adapt some fixed basis to represent the canonical components. In this way, the entries $b_{\mathbf{k}}$ are specified by the trace of the approximating function on an interpolation grid, hence, in turn, the core tensor can again be represented in a certain data-sparse format. Based on the above observation, we come to a new concept of *multilevel (telescopic) decomposition* of the type of (2.4).

First, we introduce the two-level matrix/tensor decomposition of the type of (1.1), (2.4).

Definition 2.3. *Let $\mathcal{A}$ have the form* (1.1) *with canonical components $\mathbf{V}^{(\ell)} \in \mathbb{R}^{I_\ell \times r_\ell} \in \{\mathcal{U}\}$, where the matrices $\{\mathcal{U}\}$ span some fixed basis which does not depend on $\mathcal{A}$ (uniform basis), while the core tensor $\mathcal{B} = \{b_{\mathbf{k}}\} \in \mathbb{R}^{r_1 \times \dots \times r_d}$ is represented by the rank-q CP decomposition* (2.1). *Then we say that $\mathcal{A}$ allows a two-level rank-$(\mathbf{r}, q)$ decomposition and denote $\mathcal{A} \in \mathfrak{T}_{(\mathcal{U}(\mathcal{I}),\mathbf{r},q)} \equiv \mathfrak{T}_{(\mathcal{U},\mathbf{r},q)}$. Respectively, introducing a similar definition for matrices in* (2.4), *we denote the new formats as $A \in \mathfrak{T}_{(\mathcal{U}(\mathcal{I}\times\mathcal{J}),\mathbf{r},q)}$.*

The example of a uniform basis is given by structured matrices generated by a fixed set of basis functions (say, uniform $\mathcal{H}$-matrices/low-rank matrices).

Lemma 2.3. *The storage for $A \in \mathfrak{T}_{(\mathcal{U}(\mathcal{I}\times\mathcal{J}),\mathbf{r},q)}$ is estimated by $O(drq)$ with $r = \max\{r_1, \dots, r_d\}$. The multiplication by a rank-r_0 vector of size n^d requires $O(drr_0qn^2)$ operations.*

Let $A_1, A_2 \in \mathfrak{T}_{(\mathcal{U}(\mathcal{I}\times\mathcal{J}),\mathbf{r},q)}$ then the matrix-product A_1A_2 can be stored in $O(dq^2r^2)$ memory units provided that a certain "product" basis $\{\widetilde{\mathcal{U}}\}$ is already precomputed. The inner product of two matrices (tensors), $\langle \mathcal{A}_1, \mathcal{A}_2 \rangle$, can be computed in $q^2n\prod_{\ell=1}^d(2r_\ell+1)$ operations.

In the case of supersymmetric tensors the factor d can be removed from the complexity estimates.

Proof. If $A \in \mathfrak{T}_{(\mathcal{U}(\mathcal{I}\times\mathcal{J}),\mathbf{r},q)}$ then only the core tensor need to be stored. Since, by definition the latter has the CP tensor-format in $\mathbb{R}^{\mathbf{r}}$, i.e.,

$$\mathcal{B} = \sum_{k=1}^{q} \lambda_k \times_1 B_k^{(1)} \times_2 \dots \times_d B_k^{(d)}, \quad \lambda_k \in \mathbb{R}, \tag{2.6}$$

where $B_k^{(\ell)} \in \mathbb{R}^r$, the desired storage requirements follow. The analysis for the matrix-vector product is straightforward due to the relation

$$\mathcal{B} \times_1 \mathbf{V}^{(1)}x_1 \times_2 \mathbf{V}^{(2)}x_2 \dots \times_d \mathbf{V}^{(d)}x_d = \sum_{k=1}^{q} \lambda_k \times_1 B_k^{(1)}\mathbf{V}^{(1)}x_1 \times_2 \dots \times_d B_k^{(d)}\mathbf{V}^{(d)}x_d.$$

To prove the matrix-matrix complexity, we set

$$\mathcal{A}_1 = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_d \mathbf{U}^{(d)}, \quad \mathcal{A}_2 = \mathcal{C} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \dots \times_d \mathbf{V}^{(d)}.$$

The representation (2.6) now implies (same for A_2)

$$\mathcal{A}_1 = \sum_{k=1}^{q} \lambda_k \bigotimes_{\ell=1}^{d} \mathbf{U}^{(\ell)} B_k^{(\ell)}. \tag{2.7}$$

In fact, substituting (2.6) in (2.4) and setting $B_k^{(\ell)} = \{b_{k,k_\ell}\}_{k_\ell=1}^{r_\ell}$ lead to

$$\mathcal{A}_1 = \sum_{k=1}^{q} \lambda_k \left(\sum_{k_1=1}^{r_1} \dots \sum_{k_d=1}^{r_d} b_{k,k_1} \dots b_{k,k_d} \times_1 U_{k_1}^{(1)} \times_2 \dots \times_d U_{k_d}^{(d)} \right) =$$

$$\sum_{k=1}^{q} \lambda_k \times_1 \left(\sum_{k_1=1}^{r_1} b_{k,k_1} U_{k_1}^{(1)} \right) \times_2 \dots \times_d \left(\sum_{k_d=1}^{r_d} b_{k,k_d} U_{k_d}^{(d)} \right) = \sum_{k=1}^{q} \lambda_k \bigotimes_{\ell=1}^{d} \mathbf{U}^{(\ell)} B_k^{(\ell)}.$$

Imposing (2.7), we finally obtain

$$\mathcal{A}_1\mathcal{A}_2 = \left(\sum_{k=1}^{q} \lambda_k \bigotimes_{\ell=1}^{d} \mathbf{U}^{(\ell)} B_k^{(\ell)} \right) \left(\sum_{m=1}^{q} \mu_m \bigotimes_{\ell=1}^{d} \mathbf{V}^{(\ell)} C_m^{(\ell)} \right) = \sum_{k=1}^{q} \sum_{m=1}^{q} \lambda_k\mu_m \left(\bigotimes_{\ell=1}^{d} \mathbf{U}^{(\ell)} B_k^{(\ell)} \mathbf{V}^{(\ell)} C_m^{(\ell)} \right) =$$

$$\sum_{k=1}^{q}\sum_{m=1}^{q}\lambda_k\mu_m\bigotimes_{\ell=1}^{d}\Big(\sum_{k_\ell=1}^{r}\sum_{m_\ell=1}^{r}b_{k,k_\ell}c_{k,m_\ell}U_{k_\ell}^{(\ell)}V_{m_\ell}^{(\ell)}\Big),$$

where $\{\lambda_k\mu_m\}$ has q^2 entries. In turn, assuming that the matrices $U_{k_\ell}^{(\ell)}V_{m_\ell}^{(\ell)}$ can be stored *a priori*, we need only r^2 coefficients to represent each canonical component in the product matrix. To prove the inner product complexity (say, in the matrix case) we note that

$$\Big\langle\bigotimes_{\ell=1}^{d}U^{(\ell)},\bigotimes_{\ell=1}^{d}V^{(\ell)}\Big\rangle:=\prod_{\ell=1}^{d}\langle U^{(\ell)},V^{(\ell)}\rangle.$$

Then the result follows from the representation

$$\langle\mathcal{A}_1,\mathcal{A}_2\rangle:=\sum_{k=1}^{q}\sum_{m=1}^{q}\lambda_k\mu_m\Big\langle\bigotimes_{\ell=1}^{d}\mathbf{U}^{(\ell)}B_k^\ell,\bigotimes_{\ell=1}^{d}\mathbf{V}^{(\ell)}C_m^\ell\Big\rangle=\sum_{k=1}^{q}\sum_{m=1}^{q}\lambda_k\mu_m\prod_{\ell=1}^{d}\langle\mathbf{U}^{(\ell)}B_k^\ell,\mathbf{V}^{(\ell)}C_m^\ell\rangle.$$

In the case of supersymmetric tensors we just take into account that the canonical components coincide for different $\ell=1,\dots,d$ and the same holds for the core tensor. □

We can observe that the proposed combination of a CP decomposition (to represent the core tensor) with a fixed basis $\{\mathcal{U}\}$ (to represent the canonical components) reduces dramatically the memory demands. Even more important, it also improves the computational complexity of the numerical approximation (see Section 4).

If the dimension $\mathbf{r}$ of $\mathcal{B}$ is still large enough, one can introduce a next-level decomposition of the core tensor which leads to the *multilevel ("telescopic") version of the method*.

Example 2.2. The familiar sparse grid methods in $\mathbb{R}^d$ can be interpreted as the Tucker decomposition with the fixed tensor-product hierarchical basis of size n^d to represent the canonical components endowed with the "hyperbolic cross" selection of nonzero elements in the core tensor $\mathcal{B}$ of the full size n^d, where $r_\ell=n$ $(\ell=1,\dots,d)$. The approximability features in this format rely on the regularity with respect to square integrable mixed weak derivatives leading to an $O(n\log^{d-1}n)$ estimate on the number of nonzero elements in the "compressed" tensor $\mathcal{B}$.

Example 2.3. The tensor-product *sinc* interpolation method (see [20]) corresponds to the Tucker model, where the canonical components are represented in the tensor-product *sinc*-basis, while the fully populated core tensor $\mathcal{B}$ has a reduced size, e.g., $r_\ell=|\log\varepsilon|\log n$ $(\ell=1,\dots,d)$. In this way, a sparsification strategy is based on the application of the classical Kotelnikov-Whittaker-Shannon *sampling theorem* (originally applied to band limited signals) to analytic, exponentially decaying multivariate functions.

The tensor decompositions described in Sections 2.2 – 2.4 can be applied to the following classes of operators/tensors:

- tensors generated by a class of analytic functions;
- integral operators with analytic/asymptotically smooth kernels;
- matrix-valued functions, e.g., A^{-1}, A^α, $\exp(A)$.

In this paper we analyse approximation methods for function-generated tensors. Further applications will be discussed elsewhere.

3. Computational aspects of CP and Tucker models

In this section we give a brief survey of the existing iterative methods to approximately compute the CP or Tucker decompositions of a given tensor $\mathcal{A}_0 \in \mathbb{R}^{\mathcal{I}}$. The latter can already be given in the CP format but with a rather large Kronecker rank r_0 that can be reduced via certain algebraic/analytic/combined approximations.

3.1. CP decomposition. The most general approach is to derive CP-components by straightforward minimisation of the quadratic cost functional

$$f(\mathcal{A}) := \|\mathcal{A} - \mathcal{A}_0\|^2 \to \min \tag{3.1}$$

over all rank-r tensors $\mathcal{A} \in \mathfrak{C}_r$, which will be parametrised as in (1.3) and with the constraints $\|V_k^{(\ell)}\| = 1$ $(k = 1, \ldots, r;\ \ell = 1, \ldots, d)$.

To find the local minima of (3.1), Newton-type algorithms can be applied to the Lagrange equation corresponding to the unconstrained minimisation problem: Find $\mathcal{A} \in \mathfrak{C}_r$ and the Lagrange multipliers $\lambda^{(k,\ell)} \in \mathbb{R}$ such that

$$\langle \mathcal{A} - \mathcal{A}_0, \mathcal{A} - \mathcal{A}_0 \rangle + \sum_{k=1}^{r} \sum_{\ell=1}^{d} \lambda^{(k,\ell)} (\|V_k^{(\ell)}\|^2 - 1) \to \min . \tag{3.2}$$

In the case of a satisfactory convergence behaviour (in the presence of a good initial guess) the complexity of one Newton's iteration may be estimated at least by $O(dr^2n + r^d)$.

As a second common approach, one can resort to an alternating least-squares (ALS) algorithm, which is as follows: let $B = \mathrm{diag}\{b_1, \ldots, b_d\}$ in (1.3) and assume that all matrices $\mathbf{V}^{(\ell)}$, $\ell \neq m$, are fixed. Then (3.1) is a quadratic expression in the components of the matrix $\mathbf{V}^{(m)} \cdot B$, hence we obtain a classical least-squares problem. To drive the solution toward the local minima, an ALS iteration repeats this procedure for each component $m = 1, \ldots, d$ until convergence (or termination) is attained. The components of $B = \mathrm{diag}\{b_1, \ldots, b_d\}$ are obtained by normalisation of the columns $V_k^{(m)}$ $(k = 1, \ldots, r)$ to the unitlength.

Remark 3.1. In general, the convergence analysis of both Newton's and ALS schemes is still an open question. The intrinsic difficulty to achieve the robust convergence of such nonlinear iterations is due to the well-known effect of degeneracy of the minimising sequence (if $d > 2$, the corresponding set $\mathfrak{C}_r$ of structured tensors is no longer closed). For example, the finite difeerence (FD) matrix representation $A \in \mathbb{R}^{n^3 \times n^3}$ for the 3D Laplace operator is given by a tensor of Kronecker rank three,

$$A := V \otimes I \otimes I + I \otimes V \otimes I + I \otimes I \otimes V, \quad V, I \in \mathbb{R}^{n \times n},$$

where I is the $n \times n$ identity matrix and $V = \mathrm{tridiag}\,\{-1, 2, -1\}$. It can be approximated with any tolerance $\varepsilon > 0$ by a tensor of rank two,

$$A_k := I \otimes I \otimes (V - kI) + k\left(I + \frac{1}{k}V\right) \otimes \left(I + \frac{1}{k}V\right) \otimes I, \quad A - A_k = \frac{1}{k} V \otimes V \otimes I \to 0, \quad \text{as } k \to \infty.$$

However, the minimising sequence A_k does not converge as $k \to \infty$. Surprisingly, the Tucker rank of the FD "d-dimensional Laplacian"

$$A := V \otimes I \otimes \ldots \times I + I \otimes V \otimes \ldots \otimes I + \ldots + I \otimes I \otimes \ldots \otimes V \in \mathbb{R}^{n^d \times n^d}, \quad V, I \in \mathbb{R}^{n \times n},$$

is equal to $\mathbf{r} = (2, 2, \ldots, 2) \in \mathbb{N}^d$ independent of d (the proof is the direct application of a higher-order SVD). On the other hand, the Kronecker rank of A is $r = d$.

Notice that in the case of orthogonally decomposable tensors in $\mathfrak{C}_r^{\perp}$ the incremental rank-1 approximation algorithm correctly computes its CP representation (see [28, 42] concerning the convergence theory). However, in spite of their attractive computational features, the orthogonally decomposable tensors usually do not provide a low-rank approximation in the considered applications.

Now consider in more detail the simple special case of a CP model that is the best rank-1 approximation, since it is an important ingredient in typical multilinear algebra algorithms. To derive the corresponding Lagrange equations, we notice that due to the normalisation $\|\mathcal{A}\|^2 = b_1^2$, the minimisation problem (3.2) appears to be equivalent to the dual problem of maximising the *generalised Rayleigh quotient* over the unit-norm vectors (eliminates b_1),

$$|\mathcal{A}_0 \times_1 V^{(1)^\top} \times_2 \ldots \times_d V^{(d)^\top}|^2 - \sum_{\ell=1}^{d} \lambda^{(\ell)} \left(\|V^{(\ell)}\|^2 - 1\right) \to \max. \tag{3.3}$$

For any solution of this problem, the corresponding scalar b_1 can be chosen as a minimiser of $t^2 - 2ta_0$ with $a_0 = \mathcal{A}_0 \times_1 V^{(1)^\top} \times \ldots \times_d V^{(d)^\top}$ that is $b_1 = a_0$. Differentiating (3.3) with respect to $V^{(m)}$ $(1 \leqslant m \leqslant d)$ leads to the equations

$$a_0\, \mathcal{A}_0 \times_1 V^{(1)^\top} \ldots \times_{m-1} V^{(m-1)^\top} \times_{m+1} V^{(m+1)^\top} \ldots \times_d V^{(d)^\top} = \lambda^{(m)}\, V^{(m)},$$

which imply $\lambda^{(m)} = b_1^2$. Finally, the Lagrange equations read as (see [5])

$$\mathcal{A}_0 \times_1 V^{(1)^\top} \ldots \times_{m-1} V^{(m-1)^\top} \times_{m+1} V^{(m+1)^\top} \ldots \times_d V^{(d)^\top} = b_1\, V^{(m)},$$

$$\mathcal{A}_0 \times_1 V^{(1)^\top} \times_2 \ldots \times_d V^{(d)^\top} = b_1, \quad \|V^{(m)}\| = 1 \quad (1 \leqslant m \leqslant d).$$

The above system of Lagrange equations can be solved by an ALS algorithm, so that in each step the approximant to the scalar b_1 and the estimate of the vectors $V^{(m)}$ $(m = 1, \ldots, d)$ are optimised, while the rest vector-components with $\ell \neq m$ are kept constant. This is a higher-order generalisation of the *power method* for matrices [14]. The ALS method for the best rank-1 approximation is proved to have a locally linear convergence rate (see [42]). Alternatively, one can apply a Newton's type method that provides locally quadratic convergence.

3.2. Orthogonal rank-$(r_1, \ldots, r_d)$ decomposition. In general, the numerical complexity of the Tucker model strongly depends on the size of the data-array $\mathcal{A}$ which may require substantial computational resources (see [35] for related discussion).

In the case of an orthogonal rank-$(r_1, \ldots, r_d)$ decomposition (i.e., all $\mathbf{V}^{(\ell)} = [V_1^{(\ell)} V_2^{(\ell)} \ldots \ldots V_{r_\ell}^{(\ell)}]$ $(\ell = 1, \ldots, d)$ are orthogonal matrices), the minimisation problem (3.1) is constrained to all rank-$\mathbf{r}$ tensors $\mathcal{A} \in \mathfrak{T}_{\mathbf{r}}$. Introducing the Stiefel manifold

$$\mathcal{V}_{n,r} := \{Y \in \mathbb{R}^{n \times r} : Y^\top Y = I \in \mathbb{R}^{r \times r}\},$$

we can impose componentwise constraints as $\mathbf{V}^{(\ell)} \in \mathcal{V}_\ell := \mathcal{V}_{n_\ell, r_\ell}$ $(\ell = 1, \ldots, d)$. As in the case of rank-1 approximation, the core tensor can be eliminated from (3.2). For given components $\mathbf{V}^{(\ell)}$, we denote

$$\mathcal{B}^{(\neg m)} = \mathcal{A}_0 \times_1 \mathbf{V}^{(1)^\top} \ldots \times_{m-1} \mathbf{V}^{(m-1)^\top} \times_{m+1} \mathbf{V}^{(m+1)^\top} \ldots \times_d \mathbf{V}^{(d)^\top}$$

and let $\mathbf{B}^{(\neg m)} \in \mathbb{R}^{n_m \times \overline{r}_m}$ be the corresponding matrix representation, where $\overline{r}_m = \prod_{\ell=1, \ell \neq m}^{d} r_\ell$.

Lemma 3.1. *For given $\mathcal{A}_0 \in \mathbb{R}^{I_1 \times \ldots \times I_d}$, the minimisation problem* (3.2) *on $\mathfrak{T}_{\mathbf{r}}$ is equivalent to the dual maximisation problem*

$$g(\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(d)}) := \|\mathcal{A}_0 \times_1 \mathbf{V}^{(1)^\top} \times_2 \ldots \times_d \mathbf{V}^{(d)^\top}\|^2 \to \max \tag{3.4}$$

over a set of canonical components $\mathbf{V}^{(\ell)} \in \mathbb{R}^{|I_\ell| \times r_\ell}$ from the Stiefel manifold, i.e., $\mathbf{V}^{(\ell)} \in \mathcal{V}_\ell$ $(\ell = 1, \ldots, d)$.

Problem (3.4) *has at least one global maximum. At each extremal point the corresponding Lagrange equations read as* $(1 \leqslant m \leqslant d)$

$$2(\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top}) \cdot \mathbf{B}^{(\neg m)} \cdot \mathbf{B}^{(\neg m)^\top} \cdot \mathbf{V}^{(m)} = 0. \tag{3.5}$$

Under the compatibility condition

$$r_m \leqslant \prod_{\ell=1, \ell \neq m}^{d} r_\ell \quad (1 \leqslant m \leqslant d) \tag{3.6}$$

equation (3.5) *is solvable for any $m = 1, \ldots, d$. For given matrices $\mathbf{V}^{(m)}$ $(m = 1, \ldots, d)$, the tensor $\mathcal{B}$ that minimises* (3.2) *is represented by*

$$\mathcal{B} = \mathcal{A}_0 \times_1 \mathbf{V}^{(1)^\top} \times_2 \ldots \times_d \mathbf{V}^{(d)^\top} \in \mathbb{R}^{r_1 \times \ldots \times r_d},$$

providing the equation $f(\mathcal{A}) + g(\mathcal{A}) = \|\mathcal{A}\|^2$.

Proof. The first and last assertions are direct consequences of Theorem 4.2 [5].

The existence of the global maximum follows from the compactness of the Stiefel manifold $\mathcal{V}_\ell$ with respect to the Frobenius norm. The justification of the Lagrange equation is based on the standard variational arguments. Notice that in the matrix notations the function g takes the form

$$g(\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(d)}) = \langle \mathbf{B}^{(\neg m)^\top} \cdot \mathbf{V}^{(m)}, \mathbf{B}^{(\neg m)^\top} \cdot \mathbf{V}^{(m)} \rangle.$$

To calculate the derivative of g over the Stiefel manifold $\mathcal{V}_m$, first we notice that the tangent space at $V \in \mathcal{V}_m$ is

$$\mathcal{T}_V \mathcal{V}_m = \{\delta \in \mathbb{R}^{n_m \times r_m} : \delta^\top V + V^\top \delta = 0\}.$$

Second, making use of the orthogonal projections $\mathbf{V}^{(m)}\mathbf{V}^{(m)^\top}$ and $\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top}$ onto the space spanned by columns of $\mathbf{V}^{(m)}$ and onto its orthogonal complement, respectively, we are able to represent any (constraint) variation in the tangent space $\delta_{\mathcal{T}} \in \mathcal{T}_V \mathcal{V}_m$ over the arbitrary variation $\delta \in \mathbb{R}^{n_m \times r_m}$ by

$$\delta_{\mathcal{T}} = (\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top})\delta. \tag{3.7}$$

In fact, since the orthogonality $\mathbf{V}^{(m)^\top}\mathbf{V}^{(m)} = \mathbf{I}$, it is readily seen that

$$\delta_{\mathcal{T}}{}^\top \mathbf{V}^{(m)} + \mathbf{V}^{(m)^\top}\delta_{\mathcal{T}} = \delta^\top(\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top})\mathbf{V}^{(m)} + (\mathbf{V}^{(m)^\top}\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top})\delta =$$

$$\delta^\top(\mathbf{V}^{(m)} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top}\mathbf{V}^{(m)}) + (\mathbf{V}^{(m)^\top} - \mathbf{V}^{(m)^\top}\mathbf{V}^{(m)}\mathbf{V}^{(m)^\top})\delta = 0.$$

With $\delta_{\mathcal{T}} \in \mathcal{T}_{\mathbf{V}^{(m)}} \mathcal{V}_m$ being an arbitrary variation in the tangent space, with the properties of a scalar product and using relation (3.7), we derive equations for the Frechét derivative in $\mathbf{V}^{(m)}$,

$$g(\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(m)} + \delta_{\mathcal{T}}, \ldots, \mathbf{V}^{(d)}) - g(\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(m)}, \ldots \mathbf{V}^{(d)}) \approx 2\langle \mathbf{B}^{(\neg m)^\top} \cdot \mathbf{V}^{(m)}, \mathbf{B}^{(\neg m)^\top} \cdot \delta_{\mathcal{T}} \rangle =$$

$$2\langle \mathbf{B}^{(\neg m)} \cdot \mathbf{B}^{(\neg m)^\top} \cdot \mathbf{V}^{(m)}, (\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top})\delta \rangle =$$
$$2\langle (\mathbf{I} - \mathbf{V}^{(m)}\mathbf{V}^{(m)^\top}) \cdot \mathbf{B}^{(\neg m)} \cdot \mathbf{B}^{(\neg m)^\top} \cdot \mathbf{V}^{(m)}, \delta \rangle \quad \forall \delta \in \mathbb{R}^{n_m \times r_m}.$$

Now equation (3.5) follows.

Furthermore, since we are looking for the solution of (3.5) which is given by an orthonomal basis for the dominant subspace of the m-mode space spanned by columns of $\mathbf{B}^{(\neg m)}$, the compatibility relation (3.6) appears as the necessary and sufficient condition for the solvability of the arising SVD. □

It is readily seen that in the case of the rank-1 approximation (i.e., $\mathbf{r} = 1$) the system of the Lagrange equations (3.5) is equivalet to that presented in Section 3.

Based on Lemma 3.1 the rank-$(r_1, \ldots, r_d)$ approximation can be calculated by the orthogonal ALS iteration (see, e.g., [5] for the particular scheme). The ALS algorithm can be implemented in three steps.

ALS$(\mathcal{T}_\mathbf{r})$.

Step I. Compute the initial guess (analytic approximation, truncation of the so-called higher-order SVD, approximation with a smaller Tucker rank).

Step II. For each $(m = 1, \ldots, d)$ the ALS algorithm optimises the canonical component $\mathbf{V}^{(m)}$ by solving equation (3.5), while the other matrix-components are kept constant. Termination criterion: fixed number of iterations or control of the current increment.

Step III. Compute the core tensor.

This scheme is a higher-order extension of the *orthogonal iteration* for matrices [14].

The Newton-type algorithm can be also applied to the system of Lagrange equations (3.5). For successful convergence of a nonlinear iteration, the initial guess should belong to the attraction region of the global optimum (see Step I above).

Remark 3.2. The complexity of the minimisation algorithm can be reduced dramatically if the target tensor $\mathcal{A}_0$ is already represented in the format with moderate Kronecker ranks R_ℓ such that $Cn \geqslant R_\ell > r_\ell$ $(\ell = 1, \ldots, d)$.

Further complexity reduction is possible if one looks for the canonical components represented in the fixed basis (the latter may depend on the input-tensor) and if, at the same time, one can sparsify the core tensor. In the following, we focus on the analysis of the two-level structured rank-$(r_1, \ldots, r_d)$ decomposition which can be applied to certain function-related matrices/tensors and to the corresponding discrete nonlocal operators.

3.4 Multilevel rank-$(r_1, \ldots, r_L)$ and two-level rank-$(\mathbf{r}, q)$ models. *A multilevel rank-*$(r_1, \ldots, r_L)$ *CP decomposition* can be computed by successive application of SVD (if there is no a priori information on the Kronecker rank) or by orthogonal iteration with some fixed rank r_ℓ on each level. The corresponding simplifications are standard if the target tensor is already represented by moderate Kronecker ranks R_ℓ, $n \leqslant R_\ell > r_\ell$.

Problem independent basis. A two-level rank-$(\mathbf{r}, q)$ model can be computed using an interpolation process with respect to the fixed basis on the first level (say, *Sinc*- or wavelet-interpolation) and then applying an analytic-algebraic method to compute the CP decomposition to the low-dimensional core tensor of size $r_1 \times \ldots \times r_d$. For moderate problems, the first-level decomposition can also be computed by the orthogonal ALS iteration as described in Section 3.2.

Problem dependent basis. The "fixed basis" can be adapted to the problem by truncation of the higher-order SVD (see [4]) applied to the target tensor. Again, one can benefit if the latter is already represented in the tensorform with rather small Kronecker ranks $R_\ell > r_\ell$, the

QR-decomposition of the corresponding unfolding matrices can be applied for each dimension $\ell = 1, \ldots, d$. The data-sparse representation of the core tensor is along the line as above.

4. Approximating function generated tensors

In this section we discuss the low Kronecker rank approximation of a special class of higher-order tensors arising as certain "discretisations" of multivariate functions. They will be called function-generated tensors (FGTs). FGTs directly appear from

(a) the representation of multivariate functions (say, solutions or righthand sides in the FEM/BEM discretisations in $\mathbb{R}^d$);

(b) the representation of "local potentials" acting as a multiplication operator;

(c) Nyström, collocation or Galerkin discretisations of integral operators;

(d) the approximation to some analytic matrix-valued functions.

4.1. Basic definitions. In the following we define the FGTs corresponding to the Nyström/collocation and Galerkin discretisations.

In the case of the interpolation method, we let $\mathcal{I}_\ell = I_{\ell,1} \times \ldots \times I_{\ell,p}$ be the product index set, where we use multiindices $\mathbf{i}_\ell = (i_{\ell,1}, \ldots, i_{\ell,p}) \in \mathcal{I}_\ell$ ($\ell = 1, \ldots, d$) with the components $i_{\ell,m} \in \{1, \ldots, n\}$ ($m = 1, \ldots, p$). Furthermore, let ω_ℓ be a uniform rectangular grid on $\Pi := [a_0, b_0]^p$, $a_0, b_0 > 0$, indexed by $\mathcal{I}_\ell$, and let $\{\zeta^{(1)}_{\mathbf{i}_1}, \ldots, \zeta^{(d)}_{\mathbf{i}_d}\}$ with $\mathbf{i}_\ell \in \mathcal{I}_\ell$ ($\ell = 1, \ldots, d$), be a set of collocation points living on the tensor-product lattice $\omega_{\mathbf{d}} := \omega_1 \times \ldots \times \omega_d$ in a hypercube $\Omega := \Pi^d \in \mathbb{R}^{\mathbf{d}}$ with $\mathbf{d} = dp$ so that $(\zeta^{(1)}_{\mathbf{i}_1}, \ldots, \zeta^{(d)}_{\mathbf{i}_d}) \in \mathcal{I}^{(\mathbf{d})} := \mathcal{I}_1 \times \ldots \times \mathcal{I}_d$. We also define $|\mathbf{i}_\ell| = \max\limits_{m \leqslant p} |i_{\ell,m}|$ and similarly for $|\mathbf{i}|$, $\mathbf{i} = (i_1, \ldots, i_d)$. In our applications we have $d \geqslant 2$ with some fixed $p = 1, 2, 3$.

Definition 4.1 (Collocation case, FGT(C)). Given a multivariate function $g : \Omega \to \mathbb{R}$, we introduce the collocation-type FGT of order d by

$$\mathcal{A} \equiv \mathcal{A}(g) := [a_{\mathbf{i}_1 \ldots \mathbf{i}_d}] \in \mathbb{R}^{\mathcal{I}_1 \times \ldots \times \mathcal{I}_d} \quad \text{with} \quad a_{\mathbf{i}_1 \ldots \mathbf{i}_d} := g(\zeta^{(1)}_{\mathbf{i}_1}, \ldots, \zeta^{(d)}_{\mathbf{i}_d}). \tag{4.1}$$

In the case of Galerkin schemes we make use of the tensor-product test functions

$$\phi^{\mathbf{i}}(x_1, \ldots, x_d) = \prod_{\ell=1}^{d} \phi^{i_\ell}_\ell(x_\ell), \quad \mathbf{i} = (i_1, \ldots, i_d) \in \mathbb{R}^{\mathcal{I}_1 \times \ldots \times \mathcal{I}_d}, \quad i_\ell \in I_n := \{1, \ldots, n\}, \tag{4.2}$$

and $\psi^{\mathbf{j}}$ with $\mathbf{j} = (j_1, \ldots, j_d) \in \mathbb{R}^{\mathcal{J}_1 \times \ldots \times \mathcal{J}_d}$, $j_\ell \in I_n$, of a similar product form.

Definition 4.2 (Galerkin case, FGT(G)). Given a multivariate function $g : \Omega \times \Omega \to \mathbb{R}$ with $\Omega \in \mathbb{R}^d$ and the tensor-product basis set (4.2), we let $p = 2$, $\zeta^{(\ell)} = (x_\ell, y_\ell)$, $\mathbf{m}_\ell = (i_\ell, j_\ell) \in \mathcal{M}_\ell$ and introduce the Galerkin-type dth order FGT by $\mathcal{A} \equiv \mathcal{A}(g) := [a_{\mathbf{m}_1 \ldots \mathbf{m}_d}] \in \mathbb{R}^{\mathcal{M}_1 \times \ldots \times \mathcal{M}_d}$ with

$$a_{\mathbf{m}_1 \ldots \mathbf{m}_d} := \int\limits_{\Omega \times \Omega} g(\zeta^{(1)}, \ldots, \zeta^{(d)}) \phi^{\mathbf{i}}(x_1, \ldots, x_d) \psi^{\mathbf{j}}(y_1, \ldots, y_d)\, dx\, dy. \tag{4.3}$$

In various applications, the function g is analytic in all variables except for the "small" set of singularity points given either by a hyper-plane $\mathcal{S}(g) := \{\zeta \in \Omega : \zeta^{(1)} = \zeta^{(2)} = \ldots = \zeta^{(d)}\}$ or by a single point $\mathcal{S}(g) := \{\zeta \in \Omega : \zeta^{(1)} = \zeta^{(2)} = \ldots = \zeta^{(d)} = 0\}$.

In numerical calculations involving nonlocal/integral operators (e.g., those that arise from the Hartree-Fock and Boltzmann equations), n may vary from several hundreds to several

thousands, therefore, for $d \geqslant 3$, the naive "entry-wise" representation to the tensor $\mathcal{A}$ in (4.1) amounts to substantial computer resources (at least of order $O(n^{dp})$).

Some examples of multivariate functions are given in Section 4.4.

4.2. Kronecker rank in CP decomposition. We recall that the CP-type decompositions like (1.3) (or (2.1) in the matrix case) can be derived by using the corresponding separable expansion of the generating function g (see [20, 22] for more details). Assume that we are given a set of approximating functions $\{\Phi_k^\ell : \mathbb{R}^p \to \mathbb{R}\}$ $(\ell = 1, \dots, d)$.

Proposition 4.1. *Suppose that a multivariate function $g : \Omega \to \mathbb{R}$ can be approximated by a separable expansion with respect to $\{\Phi_k^\ell\}$,*

$$g_r(\boldsymbol{\zeta}) := \sum_{k=1}^{r} \mu_k \Phi_k^{(1)}(\zeta^{(1)}) \cdots \Phi_k^{(d)}(\zeta^{(d)}) \approx g(\boldsymbol{\zeta}), \quad \zeta^{(\ell)} \in \mathbb{R}^p, \quad \ell = 1, \dots, d, \tag{4.4}$$

where $\mu_k \in \mathbb{R}$. Then the FGT(C) defined by the CP decomposition (1.3) *via $\mathcal{A}_r := \mathcal{A}(g_r)$ as in Def.* 4.1 *with $b_k = \mu_k$ and with*

$$V_k^{(\ell)} = \{\Phi_k^{(\ell)}(\zeta_{\mathbf{i}_\ell}^{(\ell)})\}_{\mathbf{i}_\ell \in \mathcal{I}_\ell} \in \mathbb{R}^{\mathcal{I}_\ell}, \tag{4.5}$$

and the FGT(G) corresponding to the choice

$$V_k^{(\ell)} = \int \Phi_k^{(\ell)}(\zeta_{\mathbf{i}_\ell}^{(\ell)}) \phi_\ell^{i_\ell}(x_\ell) \psi_\ell^{j_\ell}(y_\ell)\, dx_\ell\, dy_\ell \in \mathbb{R}^{\mathcal{I}_\ell \times \mathcal{J}_\ell}, \quad \ell = 1, \dots, d, \quad k = 1, \dots, r, \tag{4.6}$$

both provide the error estimate $\|\mathcal{A}(g) - \mathcal{A}_{(r)}(g_r)\|_\infty \leqslant C\|g - g_r\|_{L^\infty(\Omega)}$, where $C = 1$ in the FGT(C) case.

Proof. Analysis for the FGT(C) is presented in [20]. In the Galerkin case, we easily obtain

$$|a_{\mathbf{m}_1 \dots \mathbf{m}_d} - a^{(r)}_{\mathbf{m}_1 \dots \mathbf{m}_d}| = \left| \int_{\Omega \times \Omega} (g(x,y) - g_r(x,y)) \phi^{\mathbf{i}}(x) \psi^{\mathbf{j}}(y) dx dy \right| \leqslant \|g - g_r\|_{L^\infty(\Omega)} \int_{\Omega \times \Omega} |\phi^{\mathbf{i}}(x) \psi^{\mathbf{j}}(y)| dx dy,$$

then the result follows. □

In computationally efficient algorithms the separation rank r is supposed to be as small as possible, while the set of functions $\{\Phi_k^{(\ell)} : \mathbb{R}^p \to \mathbb{R}\}$ can be fixed *a priori* or chosen adaptively to the problem.

Though in general the construction of decomposition (4.4) with a small separation rank r is a complicated numerical task, in many interesting applications efficient approximation methods are available. In particular, for a class of multivariate functions (say, certain shift-invariant *Green's kernels* in $\mathbb{R}^d$) it is possible to obtain a dimensionally-independent Kronecker rank estimate $r = O(\log n |\log \varepsilon|)$ based on *sinc*-quadrature methods or the approximation by exponential sums (see the case-study examples in [2, 20]).

In this section we discuss the constructive CP decomposition of FGTs applied to the *general class of generating functions* characterised in terms of their Laplace transform. The approximation results are based on the *sinc-quadrature methods*.

We consider the class of multivariate functions $g : \mathbb{R}^d \to \mathbb{R}$ parametrised by $g(\boldsymbol{\zeta}) = G(\rho)$ with $\rho = \rho_1(\zeta_1) + \ldots + \rho_d(\zeta_d) > 0$, $\rho_\ell : \mathbb{R}^p \to \mathbb{R}_+$ (with small $p \in \mathbb{N}_{\geqslant 1}$), where the univariate function $G : \mathbb{R}_+ \to \mathbb{R}$ can be expressed in terms of the Laplace transform $G(\rho) = \int_{\mathbb{R}_+} \mathcal{G}(\tau) e^{-\rho\tau}\, d\tau$. Now the FGT(G) approximation corresponds to $p = 2$, $\zeta_\ell = (x_\ell, y_\ell)$.

Without loss of generality, we suppose that $\phi_\ell^{i_\ell}(\cdot)=\phi(\cdot+(i_\ell-1)h)$ $(\ell=1,\dots,d)$ with a single scaling function ϕ, where $h>0$ is the mesh parameter, and the same for $\psi_\ell^{j_\ell}(\cdot)$. We also simplify and set $\rho_\ell=\rho_0(x_\ell,y_\ell)$ $(\ell=1,\dots,d)$ and, moreover, $\rho_0:[a,b]^2\to[a_1,b_1]\in\mathbb{R}_{>0}$, while $\rho\in[a_d,b_d]\in\mathbb{R}_{>0}$. The more general multilevel setting (say, corresponding to the wavelet basis) can be analysed in a completely similar way. For each $i,j\in I_n$, we introduce the parameter dependent function

$$\Psi_{i,j}(\tau):=\int\limits_{\mathbb{R}^2} e^{-\tau\rho_0(x,y)}\phi(x+(i-1)h)\psi(y+(j-1)h)\,dx\,dy,\quad \tau\geqslant 0,$$

as well as the auxiliary function $f_I(\tau):=\mathcal{G}(\tau)e^{-\rho\tau}$ and assume that

$$a_{\mathbf{ij}}=\int\limits_{\mathbb{R}_+}\mathcal{G}(\tau)\prod_{\ell=1}^{d}\Psi_{i_\ell j_\ell}(\tau)\,d\tau\quad\text{for}\quad(\mathbf{i},\mathbf{j})\in\mathcal{I}\times\mathcal{J}. \tag{4.7}$$

Theorem 4.1 (FGT(C) approximation). *Assume that*

(a) $\mathcal{G}(\tau)$ has an analytic extension $\mathcal{G}(w)$, $w\in\Omega_{\mathcal{G}}$, to a certain domain $\Omega_{\mathcal{G}}\subset\mathbb{C}$ which can be mapped conformally onto the strip D_δ (see Appendix), such that $w=\phi(z)$, $z\in D_\delta$ and $\phi^{-1}:\Omega_{\mathcal{G}}\to D_\delta$;

(b) for each $\rho\in[a_d,b_d]$ with $a_d>0$, the function $f(z):=\phi'(z)f_I(\phi(z))$ belongs to the Hardy space $H^1(D_\delta)$ with $N(f,D_\delta)<\infty$ uniformly in ρ;

(c) $f(t)$, $t\in\mathbb{R}$, has (c1) exponential or (c2) hyperexponential decay as $t\to\pm\infty$.

Then, for each $M\in\mathbb{N}_+$, the FGT(C), $\mathcal{A}(g)$ defined on $[a,b]^d$ allows an exponentially convergent supersymmetric CP decomposition $\mathcal{A}_{(r)}\in\mathfrak{C}_r$ with $V_k^{(\ell)}$ as in (4.5)*, where $\Phi_k^{(\ell)}(\zeta^{(\ell)})=e^{-a_k\zeta^{(\ell)}}$ $(\ell=1,\dots,d)$, and where μ_k, a_k are explicitly given by the substitution of $f(t)$ into the* sinc*-quadrature* (5.1)*, such that we have*

$$\|\mathcal{A}(g)-\mathcal{A}_{(r)}\|_\infty\leqslant Ce^{-\alpha M^\nu}\quad\textit{with}\ \ r=2M+1, \tag{4.8}$$

where $\nu=1/2$, $\alpha=\sqrt{2\pi\delta b}$ in the case of (c1) and with $\nu=1$, $\alpha=2\pi\delta b/\log(2\pi aM/b)$ in the case of (c2).

(FGT(G) approximation). *Assume that (a) holds and for each $\rho\in[a_d,b_d]$ and $(\mathbf{i},\mathbf{j})\in\mathcal{I}\times\mathcal{J}$:*

(b′) The transformed integrand $f(z):=\phi'(z)\mathcal{G}(\phi(z))\prod_{\ell=1}^d\Psi_{i_\ell j_\ell}(\phi(z))$ belongs to the Hardy space $H^1(D_\delta)$ with $N(f,D_\delta)<\infty$ uniformly in ρ; item (c) holds.

Then, for each $M\in\mathbb{N}$ the FGT(C), $\mathcal{A}(g)$ defined on $[a,b]^d$ allows a supersymmetric CP decomposition $\mathcal{A}_{(r)}\in\mathfrak{C}_r$ with $V_k^{(\ell)}$ as in (4.6) *that yields the error estimate* (4.8)*.*

Proof. In the FGT(C) case, we directly apply the *sinc*-quadrature theory to the transformed integrand $f(z)$ to obtain $T_M(f,\mathfrak{h}):=\mathfrak{h}\sum_{k=-M}^M f(k\mathfrak{h})\approx G(\rho)$ (see Appendix) with

$$|G(\rho)-T_M(\rho)|\leqslant Ce^{-\alpha M^\nu},\quad \rho\in[a_d,b_d],$$

and with the respective α,ν. Combining this estimate with Proposition 4.1 and taking into account the separability property of the exponential proves the first assertion.

To prove the FGT(G) case, we recall (4.7). Again, we apply the *sinc*-quadrature to the transformed integrand $f(z)$ to obtain the exponential convergence as in the case of the FGT(C) approximation. Since our quadrature does not depend on the index $\mathbf{i}$, $\mathbf{j}$, this completes the proof. □

Theorem 4.1 proves the existence of a CP decomposition to the FGT $\mathcal{A}(g)$ with the Kronecker rank $r=O(|\log\varepsilon|\log n)$ (in the (c2) case) or $r=O(\log^2\varepsilon)$ (in the (c1) case).

Remark 4.1. The sinc-quadrature requires pointwise evaluation of the Laplace transform which can be costly (say, in the case of the matrix-valued function $\mathcal{G}(\cdot)$ or if $\mathcal{G}(\cdot)$ is not given explicitly). In such cases one can apply fast numerical methods to compute $\mathcal{G}$ (see [31]) based on the inversion formula $\mathcal{G}(\tau) = (2\pi i)^{-1} \int_\Gamma e^{\tau z} G(z)\, dz$, $\tau > 0$, where Γ is a suitable path connecting $-i\infty$ to $+i\infty$.

Notice that in some applications, instead of the Laplace transform, it is more convenient to apply the Gaussian transform $G(\rho) = \int_{\mathbb{R}_+} \widetilde{\mathcal{G}}(\tau) e^{-\rho^2\tau^2}\, d\tau$ which can be analysed either directly or by reduction to the Laplace transform via substitutions $\rho^2 = \nu$, $\tau^2 = t$.

4.3. Rank estimates for the two-level rank-$(r_1, \ldots, r_d)$ model. For the class of applications based on the separation via tensor-product interpolation, the CP model typically leads to the Kronecker rank estimate $r_{CP} = r^d$ with $r = O(\log n |\log \varepsilon|)$, where the dimensional parameter d gets into the exponent. In such cases one can apply the two-level rank-$(r_1, \ldots, r_d)$ Tucker decomposition instead of the rank-r_{CP} canonical model.

The next lemma shows that the error of the Tucker decomposition is directly related to the error of the separable approximation of the generating function.

Lemma 4.1. *Let $g : \Omega \to \mathbb{R}$ be approximated by the separable expansion*

$$g_{\mathbf{r}}(\boldsymbol{\zeta}) := \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} b_{k_1\ldots k_d} \Phi^{(1)}_{k_1}(\zeta^{(1)}) \cdots \Phi^{(d)}_{k_d}(\zeta^{(d)}) \approx g, \quad \zeta^{(\ell)} \in \mathbb{R}^p, \quad 1 \leqslant \ell \leqslant d, \tag{4.9}$$

where $b_{k_1\ldots k_d} \in \mathbb{R}$. Then the FGT(C) of the form $\mathcal{A}_{(\mathbf{r})} := \mathcal{A}(g_{\mathbf{r}}) \in \mathfrak{T}_{\mathbf{r}}$ generated by $g_{\mathbf{r}}$ with

$$V^{(\ell)}_{k_\ell} = \{\Phi^{(\ell)}_{k_\ell}(\zeta^{(\ell)}_{\mathbf{i}_\ell})\}_{\mathbf{i}_\ell \in \mathcal{I}_\ell} \in \mathbb{R}^{\mathcal{I}_\ell}, \tag{4.10}$$

and the FGT(G), corresponding to the choice,

$$V^{(\ell)}_{k_\ell} = \int \Phi^{(\ell)}_{k_\ell}(\zeta^{(\ell)}_{\mathbf{i}_\ell}) \phi^{i_\ell}_\ell(x_\ell) \psi^{j_\ell}_\ell(y_\ell) dx_\ell dy_\ell \in \mathbb{R}^{\mathcal{I}_\ell \times \mathcal{J}_\ell}, \quad \ell = 1, \ldots, d, \quad k_\ell = 1, \ldots, r_\ell, \tag{4.11}$$

both provide the error estimate $\|\mathcal{A}(g) - \mathcal{A}_{(r)}(g_r)\|_\infty \leqslant C \|g - g_r\|_{L^\infty(\Omega)}$, where $C = 1$ in the FGT(C) case.

Proof. By the construction of $\mathcal{A}_{(\mathbf{r})}$ we have

$$\|\mathcal{A} - \mathcal{A}_{(\mathbf{r})}\|_\infty = \max_{(\mathbf{i}_1, \ldots, \mathbf{i}_d) \in \mathcal{I}^{\mathbf{d}}} \left\{ \left| g(\zeta^{(1)}_{\mathbf{i}_1}, \ldots, \zeta^{(d)}_{\mathbf{i}_d}) - \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} b_{k_1\ldots k_d} \Phi^{(1)}_{k_1}(\zeta^{(1)}_{\mathbf{i}_1}) \cdots \Phi^{(d)}_{k_d}(\zeta^{(d)}_{\mathbf{i}_d}) \right| \right\},$$

which proves the assertion in the FGT(C) case. The Galerkin-type approximation can be analysed as in Proposition 4.1. □

For the class of analytic functions with point singularities expansion (4.9) can be derived via tensor-product *Sinc*-interpolation. As an alternative, a tensor-product wavelet (or some other hierarchical basis) as well as polynomial approximations can be applied. However, our choice is motivated by the following favourable features of the *Sinc*-basis in $L^2(\mathbb{R})$:

(a) complete $L^2(\mathbb{R})$-orthogonal basis in the space of band-limited functions;

(b) interpolation basis; exponential convergence of the *Sinc*-interpolation/quadratures in the Hardy space $H^1(D_\delta)$;

(c) perfect approximation to the Gaussian $\exp(-\alpha x^2)$, $x \in \mathbb{R}$ (when the Gaussian basis works well the *Sinc*-basis is also good);

(d) can be appied to a multiresolution analysis (*Sinc*-wavelet [26]).

Theorem 4.2. *Assume that $g(\zeta)$ satisfies the requirements for the tensor-product sinc-interpolation (see Appendix). Then the FGT(C), $\mathcal{A}(g)$, allows an exponentially convergent rank-$(r, \dots, r)$ decomposition $\mathcal{A}_{(\mathbf{r})} \in \mathfrak{T}_{\mathbf{r}}$ with $V_{k_\ell}^{(\ell)}$ as in (4.10), where $\Phi_{k_\ell}^{(\ell)}(\zeta^{(\ell)}) = \mathrm{sinc}(-a_{k_\ell}\zeta^{(\ell)})$ $(\ell = 1, \dots, d)$, and where $b_{\mathbf{k}}$ are explicitly represented via* sinc*-interpolation* (5.5)*, such that*

$$\|\mathcal{A}(g) - \mathcal{A}_{(\mathbf{r})}\|_\infty \leqslant C(1 + \log M)^d e^{-\alpha M^\nu} \quad \text{with } r = 2M + 1, \tag{4.12}$$

with $\nu = 1/2$, $\alpha = \sqrt{2\pi\delta b}$ in the (c1) case and with $\nu = 1$, $\alpha = 2\pi\delta b/\log(2\pi a M/b)$ in the (c2) case as in Theorem 4.1*.*

Suppose there is a separable scaling $g(\zeta) \to \widetilde{g}(\zeta) := f_1(\zeta_1) \dots f_d(\zeta_d) g(\zeta)$ such that $\widetilde{g}$ satisfies the conditions of Theorem 4.1*. Then there is a decomposition $\mathcal{A}_{(\mathbf{r})} \in \mathfrak{T}_{(\mathcal{U},\mathbf{r},q)}$ that converges exponentially in both r and q.*

Proof. We apply Lemma 4.1 yielding an exponential error bound for the tensor-product sinc-interpolation (see Appendix), which proves the first assertion. Since $\mathcal{B}$ can be represented by the tensor generated by $\widetilde{g}$ with respect to the *sinc*-collocation grid, the application of Theorem 4.1 ensures that $\mathcal{B}$ allows a rank-q CP decomposition that converges exponentially in q. Choosing the uniform basis spanned by *Sinc* functions completes the proof. □

The error estimate (4.12) yields $\max\limits_{\ell} r_\ell = O(|\log \varepsilon|\delta^{-1})$ and similarly for q. In turn, in some cases we get the estimate $\delta^{-1} = O(\log n)$ (see [20]).

4.4. Some examples. The following examples of multivariate functions arise in large-scale applications.

Example 4.1. Let $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_d) \in \mathbb{R}^d$ and choose $p = 1$. The function $g : \mathbb{R}^d \to \mathbb{R}$ defined by $g(\rho) := 1/\rho$ with $\rho = \zeta_1 + \ldots + \zeta_d$, $\zeta_i > 0$ arises in numerical PDEs (elliptic operator inverse in $\mathbb{R}^d$) and in quantum chemistry simulations, where d may vary up to one hundred. The analysis is based on the Laplace transform $1/\rho = \int_{\mathbb{R}_+} e^{-\rho t}\, dt$. The details can be found in [21].

Example 4.2. Let $x, y \in \mathbb{R}^d$, $p = 2$ and define $\rho = |x - y|^2 = \zeta_1^2 + \ldots + \zeta_d^2$ with $\zeta_\ell = x_\ell - y_\ell : \mathbb{R}^2 \to \mathbb{R}$, $\boldsymbol{\zeta} \in \mathbb{R}^d$. The family of functions $g(x, y) \equiv g(\boldsymbol{\zeta}) := 1/\rho^\lambda$ with $\lambda \in \mathbb{R}_{>0}$, arises in potential theory, in quantum chemistry, and in computational gas dynamics (see [25]). The choice of $\lambda = 1/2$ corresponds to the classical Newton potential, while $\lambda = -1/2$ refers to the distance function.

Low separation rank decomposition into the multivariate functions $1/\rho$, $1/\sqrt{\rho}$ and the related Galerkin approximations were discussed in [2, 20–22, 37], while the kernel function ρ^μ, $\mu \in \mathbb{R}$, was considered in [25].

Let us take a closer look to the FGT(G) for the Newton potential $1/\sqrt{\rho}$ in the hypercube $[-R, R]^d \in \mathbb{R}^d$. As a basic example, we consider piecewise linear finite elements defined by the scaling functions $\phi(x)$, $\psi(x)$ associated with a tensor-product grid with a step-size $h > 0$. The extension to other types of FEs including those with nonlocal support is straightforward.

Lemma 4.2. *The FGT(G) for the Newton potential $1/\sqrt{\rho}$ in the hypercube $[-R, R]^d \in \mathbb{R}^d$ allows a CP approximation with the exponential convergence rate (independent of d) as in* (4.12) *with $\nu = 1/2$.*

Proof. We apply the FGT(G)-version of Theorem 4.1. Assumption (4.7) can be easily proven. Obviously, we have $\rho_0(x, y) = (x - y)^2$ $(x, y \in \mathbb{R})$, hence making use of the Gaussian transform $\rho^{-1/2} = 2\pi^{-1/2} \int_{\mathbb{R}_+} e^{-\rho\tau^2}\, d\tau$, we obtain $\Psi_{i,j}(\tau) = \Psi_{|i-j|}(\tau) := \int_{\mathbb{R}^2} e^{-\tau^2(x-y)^2} \phi(x) \times \psi(y + |i - j|h)\, dx\, dy$, $\tau \geqslant 0$. Following [36], we choose the analyticity domain as a sector

$\Omega_{\mathcal{G}} := \{w \in \mathbb{C} : |arg(w)| < \delta\}$ with the apex angle $0 < 2\delta < \pi/2$ (here $\mathcal{G} = 1$), and then apply the conformal map $\varphi^{-1} : \Omega_{\mathcal{G}} \to D_\delta$ with $w = \varphi(z) = e^z$, $\varphi^{-1}(w) = \log(w)$ (see Theorem 4.1(a)).

To check condition (b') of Theorem 4.1, first, we notice that the transformed integrand $f(z) := \exp(z)\prod_{\ell=1}^{d}\Psi_{i_\ell j_\ell}(\phi(z))$ belongs to the Hardy space $H^1(D_\delta)$. Let $H = (H_1, \ldots, H_d) \in \mathbb{R}^d$ with $H_\ell = |i_\ell - j_\ell|h \leqslant R$ and let

$$f * g(u) = \int\limits_{\mathbb{R}} f(x)g(u-x)\,dx$$

be the convolution product in $\mathbb{R}^d$ provided that $q(x) = |f| * |g|$ is locally integrable. Now, using the shift property of convolution, $f(\cdot + C) * g(\cdot) = f * g(\cdot + C)$ and applying the Fubini theorem in the form

$$(f * g, \mu)_{L^2} = \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} f(x)g(y)\mu(x+y)\,dx\,dy, \quad \mu \in \mathcal{D}(\mathbb{R}^d),$$

we obtain (note that ψ is an even function)

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} e^{-w^2|x-y|^2}\phi(x)\psi(y-H)\,dx\,dy = \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} e^{-w^2|u|^2}\phi(x)\psi(u-x+H)\,dx\,du =$$

$$\int\limits_{\mathbb{R}^d} e^{-w^2|u|^2}\phi(\cdot) * \psi(\cdot + H)(u)\,du = \int\limits_{\mathbb{R}^d} e^{-w^2|u|^2}(\phi * \psi)(u+H)\,du = \int\limits_{\mathbb{R}^d} e^{-w^2|v-H|^2}(\phi * \psi)(v)\,dv,$$

taking into account that the functions ϕ and ψ have a compact support $[-h, h]^d$.

Notice that the support of $\phi * \psi$ is $[-2h, 2h]^d$. Now we estimate the constant $N(f, D_\delta)$ by

$$N(f, D_\delta) = \int\limits_{\partial\Omega_{\mathcal{G}}} |f(w)|\,|dw| = \int\limits_{\partial\Omega_{\mathcal{G}}}\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} |e^{-w^2|x-y|^2}\phi(x)\psi(y-H)\,dx\,dy|\,|dw| =$$

$$2\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}^d} |e^{-\zeta^2\exp(2i\delta)|u-H|^2}(\phi * \psi)(u)\,du|\,d\zeta \leqslant 2\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}_+} |e^{-\zeta^2\exp(2i\delta)|u-H|^2}|\,d\zeta\,|\phi * \psi|(u)\,du =$$

$$2\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}_+} e^{-\zeta^2\cos(2\delta)|u-H|^2}\,d\zeta\,|\phi * \psi|(u)\,du.$$

Calculating the integral in the variable ζ analytically, we now proceed

$$2\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}_+} e^{-\zeta^2\cos(2\delta)|u-H|^2}\,d\zeta|\phi * \psi|(u)\,du = \frac{2}{\sqrt{\cos(2\delta)}}\int\limits_{\mathbb{R}^d}\frac{|\phi * \psi|(u)}{|u-H|}\,du \leqslant$$

$$\frac{2\max|\phi * \psi|}{\sqrt{\cos(2\delta)}}\int\limits_{\mathrm{supp}(\phi*\psi)}\frac{du}{|u-H|} \leqslant \frac{C\mathrm{meas}(\mathrm{supp}(\phi * \psi))}{\mathrm{diam}(\mathrm{supp}(\phi * \psi))\sqrt{\cos(2\delta)}}.$$

This bound is uniform in H since the Newton potential

$$V(H) = \frac{1}{|u|} * |\phi * \psi|(H)$$

is a harmonic function outside of $supp(\phi * \psi)$, $\Delta V = -4\pi|\phi * \psi|$, and $V(H) \to 0$ as $|H| \to \infty$. Finally, we check that condition ($c1$) is also valid, which completes the proof. □

Example 4.3. Let ρ be defined as in Example 4.2. The function

$$g(x,y) \equiv g(\boldsymbol{\zeta}) := \exp(-\rho^{\lambda}) \quad \text{with } \lambda \in \mathbb{R}_{>0},$$

arises in quantum chemistry ($\lambda = 1/2$) and in stochastic PDEs. The following Lemma shows the existence of a low-rank CP and two-level rank-$(r_1,\dots,r_d)$ decompositions into the FGT generated by $e^{-2\sqrt{\alpha}|x-y|} = e^{-2\sqrt{\alpha\rho}}$.

Lemma 4.3. *(I) Let $\mathcal{A}$ be the FGT(C) generated by $e^{-2\sqrt{\alpha}|x-y|}$ in Ω^d (see Def. 4.1). Then there is an exponentially convergent CP decomposition into the off-diagonal part of $\mathcal{A}$ ($|\mathbf{i}-\mathbf{j}| \geqslant 1$) with the Kronecker rank $r = O(|\log\varepsilon|(|\log\varepsilon| + |\log h|))$. Besides,* $\operatorname{diag}\mathcal{A}(G)$ *is a rank-1 tensor.*

(II) The corresponding FGT(G) with respect to the FE basis in $[-R,R]^d$ as above allows an exponentially convergent rank r CP decomposition into the target tensor $\mathcal{A}(G)$.

(III) There is a two-level rank-$(r,\dots,r)$ decomposition into the FGT(G), $\mathcal{A}(G)$ with $q = r$ and with r as above. On any hyper-cube $[R_0,R]^d$ with $R_0 > 0$, there is the two-level rank-$(r,\dots,r)$ decomposition into the FGT(C) with $q = r$ and with r as above.

Proof. (I). We apply the Laplace transform to represent $G(\rho) = e^{-2\sqrt{\alpha\rho}}$ ($\rho > 0$),

$$e^{-2\sqrt{\alpha\rho}} = \frac{\sqrt{\alpha}}{\sqrt{\pi}} \int\limits_{\mathbb{R}_+} \tau^{-3/2} \exp(-\alpha/\tau - \tau\rho)\, d\tau,$$

which corresponds to the choice $\mathcal{G}(\tau) = \sqrt{(\alpha/\pi)}\tau^{-3/2}e^{\alpha/\tau}$. By the substitution $\tau = e^t$ (i.e., $\varphi(z) = e^z$), we obtain

$$e^{-2\sqrt{\alpha\rho}} = \frac{\sqrt{\alpha}}{\sqrt{\pi}} \int\limits_{\mathbb{R}} f(t;\alpha,\rho)\, dt \quad \text{with } f(t;\alpha,\rho) = \exp(-t/2 - \alpha e^{-t} - \rho e^t).$$

The decay of the integrand $f(t;\alpha,\rho)$ on the real axis is

$$f(t;\alpha,\rho) \approx e^{-t/2-\rho e^t} \quad \text{as } t \to \infty, \quad f(t;\alpha,\rho) \approx e^{|t|/2-\alpha e^{|t|}} \quad \text{as } t \to -\infty,$$

corresponding to $a = 1$, $b = \min\{\alpha,\rho\}$, $C = 1$ in (5.3). Moreover, it is easy to check that $f(z) \in H^1(D_\delta)$, $\delta < \pi/2$ with a uniformly bounded constant $N(f,D_\delta)$ in both α and ρ. Hence (5.4) holds implying the bound $M = O(|\log\varepsilon|(|\log\varepsilon| + \log 1/b))$. Using Theorem 4.1, we prove the existence of a rank-$(2M+1)$ CP decomposition applied to the corresponding FGT(C) by the interpolation grid ω that fits to the restriction $\rho > h$, with $h > 0$ being the mesh parameter. Furthermore, the diagonal part of $\mathcal{A}(G)$ corresponds to the choice $\rho = 0$, which proves part (I).

In the case of FGT(G), we have

$$\Psi_{|i-j|}(\tau) := \int\limits_{\mathbb{R}^2} e^{-\tau(x-y)^2}\phi(x)\psi(y + |i-j|h)\, dx\, dy, \quad \tau \geqslant 0.$$

Clearly, $f(z;\alpha,\rho) \in H^1(D_\delta)$ with $\delta < \pi/2$. To prove condition (b′) of Theorem 4.1, we note that $|\mathcal{G}(\zeta\exp(i\delta))| \leqslant C_0 < \infty$ for $\zeta \in [0,\infty)$. The following arguments are slight modifications of those in the previous lemma,

$$N(f,D_\delta) = \int\limits_{\partial\Omega_{\mathcal{G}}} |f(w)|\,|dw| = \int\limits_{\partial\Omega_{\mathcal{G}}} \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} |\mathcal{G}(w)||e^{-w|x-y|^2}\phi(x)\psi(y-H)\, dx\, dy|\,|dw| \leqslant$$

$$2\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}^d}|\mathcal{G}(\zeta\exp(i\delta))||e^{-\zeta\exp(i\delta)|u-H|^2}(\phi*\psi)(u)du|d\zeta\leqslant 2C_0\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}_+}|e^{-\zeta\exp(i\delta)|u-H|^2}|\,d\zeta\,|\phi*\psi|(u)\,du,$$

$$\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}_+}|e^{-\zeta\exp(i\delta)|u-H|^2}|\,d\zeta\,|\phi*\psi|(u)\,du=\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}_+}e^{-\zeta\cos(\delta)|u-H|^2}\,d\zeta\,|\phi*\psi|(u)\,du=$$

$$\frac{1}{\cos(\delta)}\int\limits_{\mathbb{R}^d}\frac{|\phi*\psi|(u)}{|u-H|^2}du\leqslant\frac{\max|\phi*\psi|}{\cos(\delta)}\int\limits_{\mathrm{supp}\,(\phi*\psi)}\frac{du}{|u-H|^2}\leqslant\frac{C\cdot\mathrm{meas}\,(\,\mathrm{supp}\,(\phi*\psi))}{\cos(\delta)\cdot\mathrm{diam}\,(\,\mathrm{supp}\,(\phi*\psi))}.$$

This bound is uniform in H. Finally, it is easily seen that condition (c) of Theorem 4.3 is satisfied with $b=1/2$, which completes the proof of part (II). The first statement in item (III) is a direct consequence of part (II).

To prove the last assertion, first, we construct the tensor-product *Sinc*-interpolant to the scaled function $G_1(\rho)=\prod_{\ell=1}^d|\zeta_\ell|^aG(\rho)$ with some $0<a<1$, which leads to a rank-$(r,\ldots,r)$ Tucker decomposition of $\mathcal{A}(G_1)$, where $V_k^{(\ell)}$ are computed with the modified basis functions $|x_\ell|^{-a}\phi_\ell(x_\ell)\psi_\ell(y_\ell)$ (see Lemma 4.5).

The core tensor has now the form $\mathcal{B}=\mathcal{A}(b_1\times_2\ldots\times_d b_1)\odot INT_{sinc}(G)$, where $INT_{sinc}(G)$ represents the trace of G at the *Sinc*-interpolation points, and b_1 is the corresponding trace of the weight function $|\zeta_\ell|^a$. Applying part (I), we approximate $INT_{sinc}(G)$ by the CP-model with the Kronecker rank r. In this way, we obtain $\mathcal{A}(G)=\mathcal{A}(b_1^{-1}\times_2\ldots\times_d b_1^{-1})\odot\mathcal{A}(G_1)$, hence the desired representation is given by rescaling the corresponding representation constructed for $\mathcal{A}(G_1)$. □

The FGT(C) approximation (with a dimensionally independent rank) is well suited for the collocation or FEM-Galerkin type schemes in the range of arguments $|x-y|\geqslant h>0$. The FGT(G) approximation is applicable to a complete tensor. The *Sinc*-interpolation is useful for constructing a two-level rank-$(r_1,\ldots,r_d)$ decomposition.

The numerical illustrations below (see Fig. 4.1) show the exponential convergence of the two-level approximation. On the first level, we apply the *Sinc*-interpolation with respect to $x\in[0,B]$ to the scaled function $x^\alpha\exp(-\sqrt{x^2+y^2})$ $(\alpha=0.5)$, which already satisfies the required conditions. The core tensor is represented via the CP model using the truncated SVD. The first and the second pair of pictures correspond to the choice $B=1.0$ and $B=10.$, respectively.

Remark 4.2. It is easily seen that Lemmata 4.2, 4.3 actually apply to the general basis that ensures the local integrability of $|\phi|*|\psi|$ and the decay property of $|\phi*\psi|$ in the form

$$\int\limits_{\mathbb{R}}|\mathcal{F}(1/|x|^\beta)\mathcal{F}(|\phi*\psi|)|\,d\omega<\infty,\quad\mathcal{F}-\text{the Fourier transform}$$

with $\beta=1,2$, respectively (say, with $|\phi*\psi(u)|\leqslant C\exp(-a|u|)$). In particular, the latter is the case for both compactly supported and "Mexican hat"-type wavelets.

Example 4.4. Further examples of Green's kernels are given by the Yukawa and Helmholtz potentials (the latter is considered for a low frequency)

$$g_1(x,y)\equiv g(\boldsymbol{\zeta}):=\frac{e^{-\mu\sqrt{\rho}}}{\sqrt{\rho}},\quad\mu\in\mathbb{R}_{\geqslant 0};\quad g_2(x,y)\equiv g(\boldsymbol{\zeta}):=\frac{e^{-\kappa\sqrt{\rho}}}{\sqrt{\rho}},\quad\kappa\in\mathbb{C}.$$

Here we sketch the analysis for the Yukawa potential. Applying the Laplace transform

$$\frac{e^{-\mu\sqrt{\rho}}}{\sqrt{\rho}} = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}_+} \tau^{-1/2} \exp\left(-\rho\tau - \mu^2/\tau\right) d\tau,$$

we find that the analytic properties of the integrand are completely similar to those in Example 4.3. Hence Lemma 4.3 with the corresponding modifications can be applied.

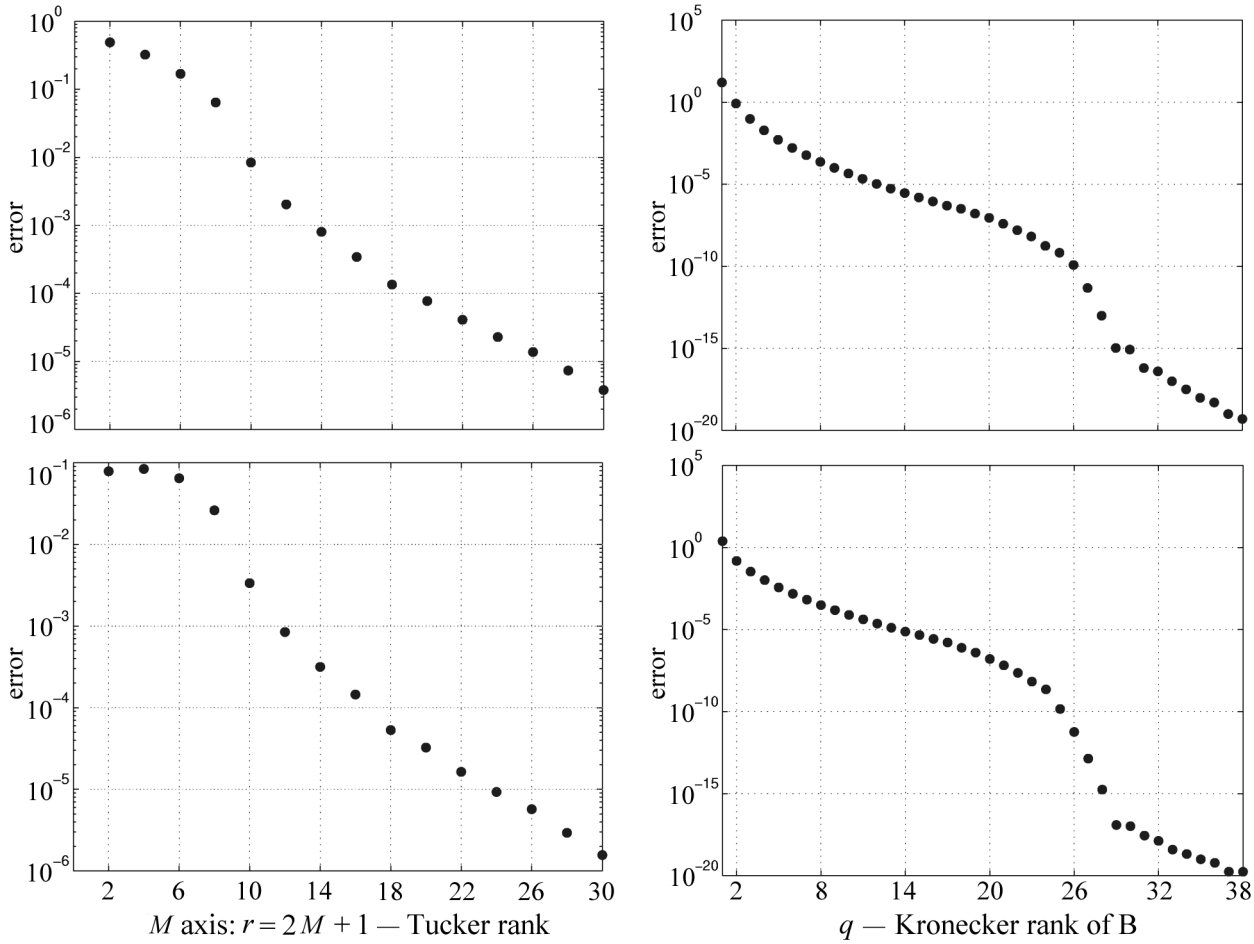


Fig. 4.1. Two-level rank-$(r_1, \dots, r_d)$ approximation to $\exp(-|x-y|)$ in the format $\mathfrak{T}_{(\mathfrak{U},\mathbf{r},q)}$ (see Def. 2.3) with $d = 2$: *Sinc*-interpolation error ($x^\alpha \exp[-(x^2+y^2)^{1/2}]$, $x \in [0,1]$, $y = 0.1$) (left), SVD for the core tensor (right)

Example 4.5. The so-called *variable hard spheres* model (related to the deterministic Boltzmann equation) is specified by the function

$$g_{1,\lambda}(x,y) := |x|^\lambda \operatorname{sinc}\left(\frac{|x||y|}{\pi}\right), \quad x, y \in \mathbb{R}^p, \quad \lambda \in (-3, 1],$$

where the sinc-function is defined in Appendix. A more general kernel function is given by $g_{2,\lambda}(x,y) := |x-y|^\lambda (|x|^2 + |y|^2 + 2|\langle x, y\rangle|)^{1/2}$. These examples correspond to the choice $d = 2$, $p = 1, 2, 3$. It is worth to note that $g_{1,\lambda}$ depends solely on two scalar variables $|x|$, $|y|$ instead of $2p$ variables in the general case, while $g_{2,\lambda}$ includes also the scalar product $\langle x, y\rangle$ (see [25] for the separable approximation to $g_{1,\lambda}$ and $g_{2,\lambda}$).

Appendix: sinc methods

Following the standard tools in the *sinc*-methods (see [36]), we introduce the Hardy space $H^1(D_\delta)$ as the set of all complex-valued functions f, which are analytic in the strip $D_\delta := \{z \in \mathbb{C} : |\Im\mathfrak{m}\, z| \leqslant \delta\}$ with some $\delta < \pi/2$, such that

$$N(f, D_\delta) := \int\limits_{\partial D_\delta} |f(z)|\,|dz| = \int\limits_{\mathbb{R}} \left(|f(x+i\delta)| + |f(x-i\delta)|\right)\, dx < \infty.$$

Let

$$S(k, \mathfrak{h})(x) = \frac{\sin\left[\pi(x-k\mathfrak{h})/\mathfrak{h}\right]}{\pi(x-k\mathfrak{h})/\mathfrak{h}} \equiv \operatorname{sinc}\left(\frac{x}{\mathfrak{h}} - k\right) \quad (k \in \mathbb{Z},\ \mathfrak{h} > 0,\ x \in \mathbb{R})$$

be the k-th Sinc function with step size $\mathfrak{h}$, evaluated at x, where the sinc-function is defined by $\operatorname{sinc}(z) = (\pi z)^{-1}\sin(\pi z)$, $z \in \mathbb{C}$. Given $f \in H^1(D_\delta)$, $\mathfrak{h} > 0$, and $M \in \mathbb{N}_0$, the corresponding Sinc quadrature reads as

$$T_M(f, \mathfrak{h}) := \mathfrak{h} \sum_{k=-M}^{M} f(k\mathfrak{h}) \approx \int\limits_{\mathbb{R}} f(\xi)\, d\xi. \tag{5.1}$$

If

$$|f(\xi)| \leqslant C \exp(-b|\xi|) \quad \text{for all } \xi \in \mathbb{R} \text{ with } b, C > 0, \tag{5.2}$$

then the quadrature error satisfies

$$\left| \int\limits_{\mathbb{R}} f(\xi)\, d\xi - T_M(f, \mathfrak{h}) \right| \leqslant C e^{-\sqrt{2\pi\delta b M}} \quad \text{with } \mathfrak{h} = \sqrt{2\pi\delta/bM},$$

and with a positive constant C depending only on f, δ, b (see [36]). If f possesses the hyper-exponential decay

$$|f(\xi)| \leqslant C \exp(-b\mathrm{e}^{a|\xi|}) \quad \text{for all } \xi \in \mathbb{R} \text{ with } a, b, C > 0, \tag{5.3}$$

then the choice $\mathfrak{h} = (aM)^{-1}\log(2\pi aM/b)$ leads to (see [10])

$$\left| \int\limits_{\mathbb{R}} f(\xi)\, d\xi - T_M(f, \mathfrak{h}) \right| \leqslant C\, N(f, D_\delta)\, e^{-2\pi\delta aM/\log(2\pi aM/b)}. \tag{5.4}$$

Note that $2M+1$ is the number of quadrature/interpolation points. If f is an even function, this number reduces to $M+1$.

The classical *Sinc* interpolant (cardinal series representation) reads as

$$C_M(f, \mathfrak{h}) = \sum_{\nu=-M}^{M} S(\nu, \mathfrak{h}) f(\nu\mathfrak{h}) \approx f. \tag{5.5}$$

If (5.2) holds, then the interpolation error satisfies $\|f - C_M(f, \mathfrak{h})\|_\infty \leqslant C M^{1/2} e^{-\sqrt{\pi\delta b M}}$, with $\mathfrak{h} = \sqrt{\pi\delta/bM}$. Assuming the hyper-exponential decay of f, we obtain

$$\|f - C_M(f, \mathfrak{h})\|_\infty \leqslant (2\pi\delta)^{-1} C\, N(f, D_\delta) e^{-\pi\delta aM/\log(\pi aM/b)} \quad \text{with } \mathfrak{h} = \log(\pi aM/b)/\,(aM)\,.$$

All results can be reformulated for an arbitrary subinterval in $\mathbb{R}$, in particular, for $\mathbb{R}_+$.

The *Sinc* interpolation method can be extended to the multidimensional case. Let a function $g_\ell(\cdot) : I_\ell \to \mathbb{R}$ be a univariate parameter-dependent function, which is the restriction of a multivariate function g onto I_ℓ with fixed remaining variables. Suppose that $g_\ell(\cdot)$ satisfies all regularity and decay conditions above, uniformly in $\ell = 1, \ldots, d$. It is shown in [20] that the *tensor-product Sinc interpolation* $\mathbf{C}_M$ with respect to d variables,

$$\mathbf{C}_M g := C_M^{(1)} \ldots C_M^{(d)} g,$$

provides the exponential error estimate

$$|g(\zeta) - \mathbf{C}_M(g, \mathfrak{h})(\zeta)| \leqslant (2\pi\delta)^{-1} C \Lambda_M^d \max_{\ell=1,\ldots,d} N(g_\ell(\cdot), D_\delta)\, e^{\frac{-\pi\delta M}{\log M}}$$

with the Lebesgue constant $\Lambda_M = O(\log M)$, and where $C_M^{(\ell)} g = C_M^{(\ell)}(g, \mathfrak{h})$ denotes the univariate *Sinc* interpolation from (5.5) applied to the variable $\zeta_\ell \in I_\ell$.

Acknowledgements. This work is based in part on the results of the collaboration with W. Hackbusch (MPI MIS, Leipzig). Helpful and stimulating discussions with W. Hackbusch, Ch. Lubich (University of Tuebingen) and with E. Tyrtyshnikov (University of Moscow) are gratefully acknowledged.

References

1. G. Beylkin, M. M. Mohlenkamp, *Numerical operator calculus in higher dimensions*, PNAS, **99**(2002), no. 16, pp. 10246–10251.

2. G. Braess and W. Hackbusch, *Approximation of $1/x$ by exponential sums in $[1, \infty)$*, IMA J. Numer. Anal. **25** (2005), no. 4, pp. 685–697.

3. J. D. Carrol and J. Chang, *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart — Young" decomposition*, Psychometrika, **35** (1970), pp. 283–319.

4. L. De Lathauwer, B. De Moor, J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., **21** (2000), no. 4, pp. 1253–1278.

5. L. De Lathauwer, B. De Moor, J. Vandewalle, *On the best rank-1 and rank-$(R_1, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., **21** (2000), pp. 1324–1342.

6. L. De Lathauwer, B. De Moor, J. Vandewalle, *Computation of the canoninical decomposition by means of a simultaneous generalized Schur decomposition*, SIAM J. Matrix Anal. Appl., **26** (2004), pp. 295–327.

7. H.-J. Flad, W. Hackbusch, B. N. Khoromskij, and R. Schneider, *Concept of data-sparse tensor-product approximation in many-particle models* (in preparation).

8. H.-J. Flad, W. Hackbusch, D. Kolb, and R. Schneider, *Wavelet approximation of correlated wavefunctions. I. Basics*, J. Chem. Phys., **116** (2002), pp. 9641–9657.

9. I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij, *Data-sparse approximation to operator-valued functions of elliptic operators*. Math. Comp., 73 (2003), pp. 1297–1324.

10. I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij, *Data-sparse approximation to a class of operator-valued functions*. Math. Comp., **74** (2005), pp. 681–708.

11. I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij, *Tensor-product approximation to elliptic and parabolic solution operators in higher dimensions*, Computing, **74** (2005), pp. 131–157.

12. I. P. Gavrilyuk and V. L. Makarov, *Exponentially convergent algorithms for the operator exponential with applications to inhomegeneous problems in Banach spaces*, SIAM J. on Numer. Anal., **43** (2005), no. 5, pp. 2144–2171.

13. I. P. Gavrilyuk and V. L. Makarov, *Algorithms without accuracy saturation for evolution equations in Hilbert and Banach spaces*, Math. Comp., **74** (2005), pp. 555–583.

14. G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1996.

15. L. Grasedyck and W. Hackbusch, *Construction and arithmetics of $\mathcal{H}$-matrices*, Computing, **70** (2003), pp. 295–334.

16. M. Griebel and S. Knapek, *Optimized tensor-product approximation spaces*, Constr. Approx., **16** (2000), pp. 303–332.

17. W. Hackbusch, *A sparse matrix arithmetic based on $\mathcal{H}$-matrices. I. Introduction to $\mathcal{H}$-matrices*, Computing, **62** (1999), pp. 89–108.

18. W. Hackbusch and B. N. Khoromskij, *A sparse $\mathcal{H}$-matrix arithmetic. II. Application to multidimensional problems*, Computing, **64** (2000), pp. 21–47.

19. W. Hackbusch and B. N. Khoromskij, *A sparse $\mathcal{H}$-matrix arithmetic: General complexity estimates*, J. Comp. Appl. Math., **125** (2000), pp. 479–501.

20. W. Hackbusch and B. N. Khoromskij, *Low-rank Kronecker product approximation to multidimensional nonlocal operators. Part I. Separable approximation of multivariate functions*, Computing, **76** (2006), pp. 177–202.

21. W. Hackbusch and B. N. Khoromskij, *Low-rank Kronecker product approximation to multidimensional nonlocal operators. Part II. HKT representations of certain operators*, Computing, **76** (2006), pp. 203–225.

22. W. Hackbusch, B. N. Khoromskij and E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product approximations*, J. Numer. Math., **13** (2005), pp. 119–156.

23. W. Hackbusch, B. N. Khoromskij and E. E. Tyrtyshnikov, *Approximate Iterations for Structured Matrices*, Preprint 112, MPI MIS, Leipzig, 2005, submitted.

24. R. Harshman, *Foundation of the PARAFAC procedure: Model and conditions for an "explanatory" multimode factor analysis*, UCLA Working Papers in Phonetics, **16** (1970), pp. 1–84.

25. B. N. Khoromskij, *Structured data-sparse approximation to high order tensors arising from the deterministic Boltzmann equation*, Preprint 4, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, 2005 (to appear in Math. Comp.).

26. B. N. Khoromskij, *An Introduction to Structured Tensor-Product Representation of Discrete Nonlocal Operators*, Lecture Notes 27, MPI MIS, Leipzig, 2005.

27. O. Koch and Ch. Lubich, *Dynamical Low Rank Approximation,* Preprint Uni. Tuebingen, 2005.

28. T. Kolda, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., **23** (2001), pp. 243–255.

29. J. B. Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics*, Linear Algebra Appl., **18** (1977), pp. 95–138.

30. J. B. Kruskal, *Rank, decomposition, and uniqueness for 3-way and N-way arrays*, In Multiway Data Analysis, R. Coppi and S. Bolasco, eds., North-Holland, Amsterdam, 1989, pp. 7–18.

31. M. Lopez-Fernandez, C. Palencia and A. Schädle, emphA spectral order method for inverting sectorial Laplace transform, SINUM 2006, to appear.

32. Ch. Lubich, *On Variational Approximations in Quantum Molecular Dynamics*, Math. Comp., **74** (2005), pp. 765–779.

33. T. von Petersdorf and Ch. Schwab, *Numerical solution of parabolic equations in high dimensions*, M2AN, **38** (2004), no. 1, pp. 93–127.

34. Ch. Schwab and R. A. Todor, Sparse Finite Elements for Stochastic Elliptic Problems — higher order moments, Computing, **71** (2003), pp. 43–63.

35. A. Smilde, R. Bro, P. Geladi, *Muti-way Analysis*, Wiley, 2004.

36. F. Stenger, *Numerical methods based on Sinc and analytic functions*, Springer – Verlag, Heidelberg, 1993.

37. S. Schwinger, *Separierbare Approximation der Coulomb — Wechselwirkung*, Diplomarbeit, TU Chemnitz/MPI MIS Leipzig, 2005.

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

39. E. E. Tyrtyshnikov, *Tensor approximations of matrices generated by asymptotically smooth functions*, Sbornik: Mathematics **194** (2003), no. 5-6, pp. 941–954 (translated from Mat. Sb. **194** (2003), no. 6, pp. 146–160).

40. E. E. Tyrtyshnikov, *Kronecker-product approximations for some function-related matrices*, Linear Algebra Appl., **379** (2004), pp. 423–437.

41. H. Yserentant, *On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives*, Numer. Math., **98** (2004), pp. 731–759.

42. T. Zang and G. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., **23** (2001), pp. 534–550.