Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Tensor-Product Approximation to Operators and Functions in High Dimensions

(revised version: March 2007)

by

Boris N. Khoromskij, and Wolfgang Hackbusch

Preprint no.: 139 2006

Tensor-Product Approximation to Operators and Functions in High Dimensions

Wolfgang Hackbusch and Boris N. Khoromskij
Max-Planck-Institute for Mathematics in the Sciences,
Inselstr. 22-26, D-04103 Leipzig, Germany.
{wh, bokh}@mis.mpg.de

March 27, 2007

Abstract

In recent papers tensor-product structured Nyström and Galerkin type approximations of certain multi-dimensional integral operators have been introduced and analysed. In the present paper we focus on the analysis of the collocation type schemes with respect to the tensor-product basis in a high spatial dimension d. Approximations up to an accuracy $\mathcal{O}(N^{-\alpha/d})$ are proven to have the storage complexity $\mathcal{O}(dN^{1/d}\log^q N)$ with q independent of d, where N is the discrete problem size. In particular, we apply the theory to a collocation discretisation of the Newton potential with the kernel $\frac{1}{|x-y|}$, $x, y \in \mathbb{R}^d$, $d \geq 3$. Numerical illustrations are given in the case of $d = 3$.

AMS Subject Classification: 65F50, 65F30, 46B28, 47A80

1 Introduction

The construction of efficient representations to multi-variate functions and related operators plays a crucial role in the numerical analysis of higher dimensional problems arising in a wide range of modern applications. For example we mention multi-dimensional integral equations, elliptic and parabolic boundary value problems posed in $\mathbb{R}^d$, $d \geq 2$.

In multi-dimensional applications, standard numerical methods usually fail due to the so-called "curse of dimensionality" (Bellman). This effect can be relaxed or completely avoided by a systematic application of Kronecker-type tensor-product representations of the arising high-order tensors. Algebraic methods for tensor-product approximations to high-order tensors have been extensively discussed in the literature (see [25, 4, 5, 16, 21, 27] and related references).

In recent papers modern methods of structured tensor-product approximations to some classes of multi-dimensional integral operators and operator-valued functions have been applied successfully (see [1, 14, 10, 2, 12, 13, 17, 19, 22] and references therein). Approximations via the Nyström and Galerkin methods have been considered in [14, 13, 19]. Applications to nonlocal operators associated with the density matrix ansatz for solving the Hartree-Fock equation [7, 2], computation of molecular density functions by the Ornstein-Zernicke equation [6], as well as collision integrals of the deterministic Boltzmann equation [18] have demonstrated the efficiency of low-rank tensor product decompositions.

In the present paper we discuss analytic methods for tensor-product approximations to multi-dimensional integral operators. For the case of collocation schemes we focus on the construction of tensor decompositions which are exponentially convergent in the separation rank. It is worthwhile to note that on the one hand, collocation schemes can be applied to much more general class of integral operators than the Nyström methods (including kernels with the diagonal singularity), on the other hand, they are much simpler than the Galerkin methods (requiring only a one-fold integration).

Approximations up to the accuracy $\mathcal{O}(n^{-\alpha})$ are proven to have the storage complexity $\mathcal{O}(dn\log^q n)$ with q independent of d, where $N = n^d$ is the discrete problem size (compare with the linear complexity $\mathcal{O}(n^d)$). For example, such methods can be applied to the classical Newton, Yukawa and Helmholtz kernels $\frac{1}{|x-y|}$, $\frac{e^{-\lambda|x-y|}}{|x-y|}$ and $\frac{\cos(\lambda|x-y|)}{|x-y|}$ with $x, y \in \mathbb{R}^d$.

The rest of the paper is organised as follows. In Section 2 analytic methods for the separable approximation via collocation schemes of multi-variate functions and related tensors are presented and analysed. We describe constructive schemes via *Sinc*-quadrature and *Sinc*-interpolation methods. In Section 3 we apply the results of Section 2 to integral operators in $\mathbb{R}^d$ in the collocation case. We complete the article with some numerical examples illustrating the efficiency of the low tensor-rank approximation of Newton's potential via optimised *Sinc*-quadratures.

2 Separable Approximation of Functions and Tensors

2.1 Approximation of Functions with Low Separation Rank

We start the discussion on the level of functions. In many applications we are interested in approximating a multi-variate function $f = f(x_1, ..., x_d)$ (from a certain class $\mathcal{H}$) in the set of separable functions

$$\mathcal{M}_1 = \{u : u(x) = \phi_1(x_1) \cdot ... \cdot \phi_d(x_d), \quad \phi_k \in H\}, \tag{2.1}$$

where H is a real, separable Hilbert space of functions defined on $\mathbb{R}$ (say, $H = L^2(\mathbb{R})$). A better approximation can be obtained by allowing for a linear combination of separable products in the approximation set,

$$\mathcal{M}_{\mathbf{r}} = \{u : u(x) = \sum_{\mathbf{k}} b_{\mathbf{k}} \phi^{(1)}_{k_1}(x_1) \cdot ... \cdot \phi^{(d)}_{k_d}(x_d), \quad b_{\mathbf{k}} \in \mathbb{R},\ \phi^{(\ell)}_k \in H\}, \tag{2.2}$$

where the sum is taken over multi-indices $\mathbf{k} = (k_1, ..., k_d)$ with $1 \le k_\ell \le r_\ell$, $r_\ell \in \mathbb{N}$, and $\mathbf{r} = (r_1, ..., r_d)$. We call the coefficients

$$\mathcal{B} = \{b_{\mathbf{k}}\} \in \mathbb{R}^{r_1 \times ... \times r_d} \tag{2.3}$$

the core tensor. Without loss of generality we can assume that the components $\phi^{(\ell)}_{k_\ell}$ $(\ell = 1, ..., d)$ are orthonormal, i.e.,

$$(\phi^{(\ell)}_{k_\ell}, \phi^{(\ell)}_{m_\ell}) = \delta_{k_\ell, m_\ell}, \qquad k_\ell, m_\ell = 1, ..., r_\ell,$$

where δ_{k_ℓ, m_ℓ} is Kronecker's delta.

Approximations in the set

$$\mathcal{M}_r = \{u : u(x) = \sum_{k=1}^{r} b_k \phi^{(1)}_k(x_1) \cdot .... \cdot \phi^{(d)}_k(x_d), \quad b_k \in \mathbb{R},\ \phi^{(\ell)}_k \in H\} \subset \mathcal{M}_{\mathbf{r}}\,, \tag{2.4}$$

with normalised components $\|\phi^{(\ell)}_k\| = 1$ can be considered. This is the special case of the approximation problem in $\mathcal{M}_{\mathbf{r}}$ with $\mathbf{r} = (r, ..., r)$, under the constraint that all off-diagonal elements of the coefficient tensor $\mathcal{B} = \{b_{\mathbf{k}}\}$ are zero. Since $\mathcal{M}_{\mathbf{r}}$ is not a linear space, we obtain a difficult nonlinear approximation problem when we want to estimate

$$\sigma(f, \mathcal{S}) := \inf_{s \in \mathcal{S}} \|f - s\| \tag{2.5}$$

for $f \in \mathcal{H}$, where either $\mathcal{S} = \mathcal{M}_{\mathbf{r}}$ or $\mathcal{S} = \mathcal{M}_r$.

2.1.1 Approximation in $\mathcal{S} = \mathcal{M}_r$

For $\mathcal{S} = \mathcal{M}_r$, the approximation problem (2.5) can be considered in the framework of best r-term approximation with regard to a redundant dictionary (cf. [24]).

A system $\mathcal{D}$ of functions from $\mathcal{H}$ is called a dictionary, if each $g \in \mathcal{D}$ has norm one and its linear span is dense in $\mathcal{H}$. We denote by $\Sigma_r(\mathcal{D})$ the collection of all functions in $\mathcal{H}$ which can be written in the form

$$s = \sum_{g \in \Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad \#\Lambda \le r$$

with $c_g \in \mathbb{R}$ and $r \in \mathbb{N}$. For $f \in \mathcal{H}$, the best r-term approximation error is defined by

$$\sigma_r(f, \mathcal{D}) := \inf_{s \in \Sigma_r(\mathcal{D})} \|f - s\|.$$

Let $\mathcal{H}$ be a real separable Hilbert space. A simple algorithm that inductively computes an estimate to the best r-term approximation is known as the so-called *Pure Greedy Algorithm* (see [24] and respective references). Let $g = g(f) \in \mathcal{D}$ be an element from $\mathcal{D}$ maximising $|(f, g)|$. We define

$$G(f) := (f, g)g, \quad R(f) := f - G(f).$$

Now the Pure Greedy Algorithm reads as follows: Define $R_0(f) := f$ and $G_0(f) := 0$. Then, for all $1 \le m \le r$, define

$$G_m(f) := G_{m-1}(f) + G(R_{m-1}(f)), R_m(f) := f - G_m(f) = R(R_{m-1}(f))$$

inductively. The output $G_r(f, \mathcal{D})$ of this algorithm is proven to realise the best r-term approximation in the particular case when $\mathcal{D}$ is an orthogonal basis of $\mathcal{H}$.

For the approximation problem on $\mathcal{M}_r$ we set

$$\mathcal{D} := \{g \in \mathcal{H} \cap \mathcal{M}_1 : \|g\| = 1\}, \quad \text{and hence} \quad \Sigma_r(\mathcal{D}) = \mathcal{M}_r.$$

The Pure Greedy Algorithm can be applied to functions characterised via the approximation property

$$\sigma_r(f, \mathcal{D}) \le r^{-q}, \quad r = 1, 2, ...,$$

with some $q \in (0, 1/2]$, and leads to the error bound (cf. [24])

$$\|f - G_r(f, \mathcal{D})\| \le C(q, \mathcal{D})r^{-q}, \quad r = 1, 2, ...,$$

which is "too pessimistic" in our applications. More precisely, we are interested in an efficient r-term approximation on a class of analytic functions with point singularities. In this case, under certain assumptions, we are able to prove exponential convergence

$$\sigma_r(f, \mathcal{D}) \le C \exp(-r^q), \quad r = 1, 2, ...,$$

with $q = 1$ or $q = 1/2$. Since, in general, the Pure Greedy Algorithm fails to recover exponential convergence, we will discuss more special numerical methods to estimate $\sigma_r(f, \mathcal{D})$ for this special class of analytic functions. Specifically, we consider quadrature- and interpolation-based approaches.

2.1.2 Approximation in $\mathcal{S} = \mathcal{M}_\mathbf{r}$

Notice that the coefficients $b_\mathbf{k}$ and the "single-component" functions $\phi_{k_\ell}^{(\ell)}$ in (2.2) are not uniquely defined (up to orthogonal transforms). However, this does not pose any problems from the computational point of view since the minimisation problem (2.5) is equivalent to the *dual maximisation problem* on $\mathcal{V}_\ell$, $\ell = 1, ..., d$, which does not include $b_\mathbf{k}$.

Assume that there exists a minimiser of the problem (2.5). Then, for given orthonormal components $\Phi^{(\ell)} = (\phi_1^{(\ell)}, ..., \phi_{r_\ell}^{(\ell)})$ $(\ell = 1, ..., d)$, the coefficient tensor $b_\mathbf{k}$ minimising (2.5) is represented by

$$b_\mathbf{k} = \left(f, \phi_{k_1}^{(1)}(\cdot) \cdot ... \cdot \phi_{k_d}^{(d)}(\cdot)\right), \quad \mathbf{k} = (k_1, ..., k_d). \tag{2.6}$$

For given $f \in \mathcal{H}$, the minimisation problem (2.5) with $\mathcal{S} = \mathcal{M}_\mathbf{r}$ is equivalent to the maximisation problem

$$\sigma(f; \mathcal{M}_\mathbf{r}) := \sup_{\Phi^{(\ell)}} \left\| \sum_\mathbf{k} f(x_1, ..., x_d)\phi_{k_1}^{(1)}(x_1) \cdot ... \cdot \phi_{k_d}^{(d)}(x_d) \right\|^2,$$

where $\Phi^{(\ell)}$, $\ell = 1, ..., d$, is taken from the set of r_ℓ-tuples $\Phi^{(\ell)} = (\phi_1^{(\ell)}, ..., \phi_{r_\ell}^{(\ell)})$ with orthonormal components.

In fact, let $f_{(\mathbf{r})} = \sum_\mathbf{k} b_\mathbf{k}\phi_{k_1}^{(1)}(x_1) \cdot ... \cdot \phi_{k_d}^{(d)}(x_d)$ be the solution of problem (2.5). Then we obtain the identity

$$\|f_{(\mathbf{r})}\| = \|\mathcal{B}\|_F,$$

since orthonormal components do not effect the L^2-norm. Now, with fixed components $\Phi^{(\ell)}$ $(\ell = 1, ..., d)$, relation (2.5) is actually a linear least-squares problem with respect to $b_{\mathbf{k}}$,

$$(f, f) - 2(f, \sum_{\mathbf{k}} b_{\mathbf{k}} \phi^{(1)}_{k_1}(x_1) \cdot ... \cdot \phi^{(d)}_{k_d}(x_d)) + (\mathcal{B}, \mathcal{B}) \to \min .$$

Solving the corresponding Lagrange equation

$$-(f, \sum_{\mathbf{k}} \delta b_{\mathbf{k}} \phi^{(1)}_{k_1}(x_1) \cdot ... \cdot \phi^{(d)}_{k_d}(x_d)) + (\mathcal{B}, \delta\mathcal{B}) = 0 \quad \text{for all } \delta\mathcal{B} \in \mathbb{R}^{r_1 \times ... \times r_d},$$

implies (2.6). Now we obtain

$$\|f - f_{(\mathbf{r})}\|^2 = \|f\|^2 - \|\mathcal{B}\|_F^2,$$

and substitution of (2.6) proves the assertion.

2.2 Tucker and Canonical Tensor Decompositions

Higher-order tensors (multi-dimensional arrays) appear in numerical computations as the discrete analogue of multi-variate functions. We consider d-th order tensors $\mathcal{A} = [a_{i_1 ... i_d}]_{(i_1 ... i_d) \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$ defined on the product index set $\mathcal{I} = I_1 \times ... \times I_d$. It is a generalisation of vectors (tensors of order 1) and matrices (tensors of order 2). We use the Frobenius norm $\|\mathcal{A}\| := \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$ induced by the *inner product*

$$\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{(i_1, ..., i_d) \in \mathcal{I}} a_{i_1 ... i_d} b_{i_1 ... i_d} \quad \text{with } \mathcal{A}, \mathcal{B} \in \mathbb{R}^{\mathcal{I}}, \tag{2.7}$$

which corresponds to the Euclidean norm of a vector. Below we will discuss tensor-product approximations which can be viewed as an analogue to low-rank approximations of matrices, where a large system matrix is replaced by a low rank matrix (compare the classical approximation of integral operators using degenerate kernels).

The class of rank-1 tensors is a discrete analogue of the class of separable functions $\mathcal{M}_1$. In the following we use the notation $\otimes$ to represent the canonical (rank-1) tensor

$$\mathcal{U} \equiv \{u_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{I}} = b \cdot U^{(1)} \otimes ... \otimes U^{(d)} \in \mathbb{R}^{\mathcal{I}},$$

defined by $u_{i_1 ... i_d} = b \cdot u^{(1)}_{i_1} \cdots u^{(d)}_{i_d}$ with $U^{(\ell)} \equiv \{u^{(\ell)}_{i_\ell}\}_{i_\ell \in I_\ell} \in \mathbb{R}^{I_\ell}$ and with a multi-index $\mathbf{i} := (i_1, ..., i_d) \in \mathcal{I}$.

The discrete analogue of the approximation in $\mathcal{M}_{\mathbf{r}}$ given by (2.2) is called the Tucker representation which deals with the approximation

$$\mathcal{A}_{(\mathbf{r})} = \sum_{k_1=1}^{r_1} ... \sum_{k_d=1}^{r_d} b_{k_1 ... k_d} \cdot V^{(1)}_{k_1} \otimes ... \otimes V^{(d)}_{k_d} \approx \mathcal{A}, \tag{2.8}$$

where the Kronecker factors $V^{(\ell)}_{k_\ell} \in \mathbb{R}^{I_\ell}$ $(k_\ell = 1, ..., r_\ell,\ \ell = 1, ..., d)$ are real vectors of the respective size $n_\ell = |I_\ell|$. Without loss of generality, we assume that for all ℓ the vectors $\{V^{(\ell)}_{k_\ell} : k_\ell = 1, ..., r_\ell\}$ are orthonormal. In the following, we denote by $\boldsymbol{\mathcal{T}}_{\mathbf{r}}$ the set of tensors represented by (2.8). Conventionally, we use the short notations $\mathbf{r} = (r_1, ..., r_d)$ (Tucker rank) and $\mathcal{B} = \{b_{\mathbf{k}}\} \in \mathbb{R}^{r_1 \times ... \times r_d}$ (core tensor). Notice that the representation of elements $\mathcal{A} \in \boldsymbol{\mathcal{T}}_{\mathbf{r}}$ even with orthogonal $\mathbf{V}^{(\ell)}$ is not unique due to the rotational uncertainty in the core tensor $\mathcal{B}$.

The canonical representation is defined by

$$\mathcal{A}_{(r)} = \sum_{k=1}^{r} b_k \cdot V^{(1)}_k \otimes ... \otimes V^{(d)}_k, \quad b_k \in \mathbb{R}, \tag{2.9}$$

where the Kronecker factors $V^{(\ell)}_k \in \mathbb{R}^{I_\ell}$ are normalised vectors (in chemometrics literature it is often called CANDECOMP/PARAFAC, or shortly CP model). The minimal number r in the representation (2.9) is called the Kronecker rank of $\mathcal{A}_{(r)}$. We denote by $\boldsymbol{\mathcal{C}}_r$ the set of tensors represented by (2.9). If we let $r = r_\ell$,

$n = n_\ell$ ($\ell = 1, ..., d$), then both the CP and Tucker representations require only drn numbers to represent the canonical components plus r (resp. r^d) memory units for the core tensor $\mathcal{B}$.

The main computational problem is the approximation of a given higher-order tensor $\mathcal{A}_0$ in a certain set of structured low-rank tensors $\boldsymbol{\mathcal{S}}$. In particular, $\boldsymbol{\mathcal{S}}$ may be one of the classes $\boldsymbol{\mathcal{T}}_{\mathbf{r}}$ or $\boldsymbol{\mathcal{C}}_r$.

There are *algebraic, analytically-based and combined* strategies for computing a Kronecker tensor-product decomposition of a higher-order tensor.

In this paper we apply *analytically-based representation* methods, which are efficient for a special class of *function-related* operators/tensors (see definitions and examples in §3).

In the context of integral operators, we consider the representation problem for a class of real-valued square matrices related to discrete multi-dimensional operators posed in $\mathbb{R}^d$, such that $A \in \mathbb{R}^{N\times N}$, $N = n^d$. More precisely, let $A \in \mathbb{R}^{\mathcal{I}\times\mathcal{I}}$ with $\#\mathcal{I} = N$ be a real-valued matrix defined on the index set $\mathcal{I} := I_n \times ... \times I_n$ (d factors) with

$$I_n = \{1, ..., n\}.$$

A matrix A (resp. a vector X) can also be regarded as a d-th order tensor $\mathcal{A} \in \mathbb{R}^{I_1^2\times...\times I_d^2}$ (resp. $X \in \mathbb{R}^{I_1\times...\times I_d}$). Hence one needs numerically tractable data-sparse representations of the arising high-dimensional tensors. 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 multi-factor version as in (2.11).

The general rank-$(r_1, ..., r_d)$ Tucker-type matrix decomposition uses the tensor-product matrix format

$$A = \sum_{k_1=1}^{r_1} ... \sum_{k_d=1}^{r_d} b_{k_1...k_d} V_{k_1}^{(1)} \otimes ... \otimes V_{k_d}^{(d)} \in \mathbb{R}^{I_1^2\times...\times I_d^2}, \quad b_{k_1...k_d} \in \mathbb{R}, \tag{2.10}$$

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

The matrix representation by the format (2.10) is a generalisation of the low-rank approximation of matrices, corresponding to the case $d = 2$. Note that (2.10) is identical to (2.8) except that now $V_{k_\ell}^{(\ell)}$ are matrices and not vectors.

The canonical Kronecker tensor-product format as proposed in [14, 12] reads

$$A = \sum_{k=1}^{r} b_k V_k^{(1)} \otimes ... \otimes V_k^{(d)}, \quad b_k \in \mathbb{R}, \tag{2.11}$$

where the Kronecker factors $V_k^{(\ell)} \in \mathbb{R}^{n\times n}$ may be matrices of a certain structure (say, hierarchical matrices). Again, (2.11) is identical to (2.9), but with vectors $V_k^{(\ell)}$ replaced by matrices.

Approximations of function-related matrices by matrices of the form (2.11) were, e.g., studied in [14], [26]. The main result of these papers are estimates of the form $r = \mathcal{O}(\log^2 \varepsilon)$ and $r = \mathcal{O}(|\log\varepsilon| \log n)$, where ε is the prescribed approximation accuracy. If there is no structure in the Kronecker factors then the storage is $\mathcal{O}(drn^2)$, while the matrix-times-matrix complexity is $\mathcal{O}(dr^2n^3)$. Introducing the hierarchical ($\mathcal{H}$-matrix) approximation to the Kronecker factors (HKT-approximations) leads to estimates of the form $\mathcal{O}(dr^2 n \log^q n)$ (under certain assumptions on the origin of the matrices [14]).

2.3 Collocation-type Approximation of Function Related Tensors

Here we discuss the low Kronecker rank approximation of a special class of higher-order tensors related to certain "discretisations" of multi-variate functions, which will be called function-generated tensors (FGTs). They directly arise from

(a) a separable approximation of multi-variate functions;

(b) Nyström/collocation/Galerkin discretisations of integral operators;

(c) the tensor-product approximation of some analytic matrix-valued functions.

In the following we define FGTs corresponding to collocation-type discretisation.

2.3.1 General Error Estimate

Let ω_ℓ^p $(\ell = 1, ..., d)$ be a uniform tensor-product grid of intervals on a rectangle $\Pi := [a_0, b_0]^p$, $a_0, b_0 > 0$, indexed by $\mathcal{I}_\ell = I_{\ell,1} \times ... \times I_{\ell,p}$ with $\mathcal{I}_\ell$ being the product index set such that for $\mathbf{i}_\ell = (i_{\ell,1}, ..., i_{\ell,p}) \in \mathcal{I}_\ell$ we have $i_{\ell,m} \in I_n := \{1, ..., n\}$ $(m = 1, ..., p)$. Furthermore, let $\omega_{\mathbf{d}}^p := \omega_1^p \times ... \times \omega_d^p$ be the corresponding tensor-product lattice in a hypercube $\Omega := \Pi^d \subset \mathbb{R}^{\mathbf{d}}$ with $\mathbf{d} = dp$.

We denote by $\{x_{\mathbf{i}_1}^{(1)}, ..., x_{\mathbf{i}_d}^{(d)}\}$ with $\mathbf{i}_\ell \in \mathcal{I}_\ell$ $(\ell = 1, ..., d)$ a set of collocation points living on the tensor-product lattice $\Omega_d := \omega_1 \times ... \times \omega_d$.

In our applications we have $d \geq 2$ with some fixed $p \in \{1, 2, 3\}$. In particular, matrix decompositions correspond to the choice $p = 2$. In this case we introduce the reordered index set of pairs $\mathcal{M}_\ell := \{\mathbf{m}_\ell : \mathbf{m}_\ell = (i_\ell, j_\ell),\ i_\ell, j_\ell \in I_n\}$ $(\ell = 1, ..., d)$, so that $\mathcal{I} = \mathcal{M}_1 \times ... \times \mathcal{M}_d$ with $\mathcal{M}_\ell = I_n \times I_n$.

The Nyström and Galerkin approximations to function related tensors were discussed in [12, 19]. In the following we focus on the collocation-type schemes, which are based on tensor-product ansatz functions

$$\psi^{\mathbf{i}}(y_1, ..., y_d) = \prod_{\ell=1}^{d} \psi_\ell^{i_\ell}(y_\ell), \quad \mathbf{i} = (i_1, ..., i_d) \in \mathcal{I}_1 \times ... \times \mathcal{I}_d. \tag{2.12}$$

In the following definition, g is a given function defined on $\Omega \times \Omega$.

Definition 2.1 *(Collocation, FGT(C)). Given the tensor-product basis set (2.12), we introduce the variable* $\zeta_{i_\ell}^{(\ell)} := (x_{i_\ell}^{(\ell)}, y_\ell)$ *with the collocation point* $x_{i_\ell}^{(\ell)}$ *and* $y_\ell \in \Pi$*, the pair* $\mathbf{m}_\ell := (i_\ell, j_\ell) \in \mathcal{M}_\ell$ *and define the collocation-type d-th order FGT by* $\mathcal{A} \equiv \mathcal{A}(g) := [a_{\mathbf{m}_1 ... \mathbf{m}_d}] \in \mathbb{R}^{\mathcal{M}_1 \times ... \times \mathcal{M}_d}$ *with*

$$a_{\mathbf{m}_1 ... \mathbf{m}_d} := \int_\Omega g(\zeta_{i_1}^{(1)}, ..., \zeta_{i_d}^{(d)}) \psi^{\mathbf{j}}(y_1, ..., y_d) dy, \quad \mathbf{m}_\ell \in \mathcal{M}_\ell. \tag{2.13}$$

In numerical calculations involving integral operators (e.g., arising in classical potential theory or from the Hartree-Fock, Ornstein-Zernicke and Boltzmann equations), n may vary from several hundreds to several thousands, therefore, for $d \geq 3$, a naive "entry-wise" representation to the fully-populated tensor $\mathcal{A}$ in (2.13) amounts to substantial computer resources, at least of the order $\mathcal{O}(n^{dp})$.

The key observation is that there is a natural duality between separable approximation of the multivariate generating function and the tensor-product decomposition of the related multi-dimensional array. Hence, the CP-type decompositions like (2.9) (or (2.11) in the matrix case) can be derived by using a corresponding separable expansion of the generating function g (see [12, 14] for more details).

Lemma 2.2 *Suppose that a multi-variate function* $g : \Omega \subset \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$ *can be approximated by a separable expansion*

$$g_r(\boldsymbol{\zeta}) := \sum_{k=1}^{r} \mu_k \Phi_k^{(1)}(\zeta^{(1)}) \cdots \Phi_k^{(d)}(\zeta^{(d)}) \approx g(\boldsymbol{\zeta}), \quad \boldsymbol{\zeta} = (\zeta^{(1)}, ..., \zeta^{(d)}) \in \mathbb{R}^{\mathbf{d}}, \tag{2.14}$$

where $\mu_k \in \mathbb{R}$ *and* $\Phi_k^\ell : \Pi \subset \mathbb{R}^2 \to \mathbb{R}$*. Define the CP decomposition (2.9) via* $\mathcal{A}_{(r)} := \mathcal{A}(g_r)$ *(cf. Definition 2.1) with the choice,*

$$V_k^{(\ell)} = \left\{ \int \Phi_k^{(\ell)}(\zeta_i^{(\ell)}) \psi_\ell^j(y_\ell) dy_\ell \right\}_{(i,j) \in \mathcal{M}_\ell} \in \mathbb{R}^{\mathcal{I}_\ell \times \mathcal{J}_\ell}, \quad \ell = 1, ..., d, \quad k = 1, ..., r, \tag{2.15}$$

and with $\zeta_i^{(\ell)} = (x_i^{(\ell)}, y_\ell)$*,* $i \in \mathcal{I}_\ell$*. Then the FGT(C)* $\mathcal{A}_{(r)}$ *provides the error estimate*

$$\|\mathcal{A}(g) - \mathcal{A}_{(r)}(g_r)\|_\infty \leq C \|g - g_r\|_{L^\infty(\Omega)}.$$

Proof. Using (2.13) we readily obtain

$$|a_{\mathbf{m}_1 ... \mathbf{m}_d} - a_{\mathbf{m}_1 ... \mathbf{m}_d}^{(r)}| \leq \max_{x \in \omega_{\mathbf{d}}} \left| \int_\Omega (g(x, y) - g_r(x, y)) \psi^{\mathbf{j}}(y) dy \right| \leq \|g - g_r\|_{L^\infty(\Omega)} \int_\Omega \left| \psi^{\mathbf{j}}(y) \right| dy,$$

and the result follows with $C = \max_{\mathbf{j}} \int_{\operatorname{supp} \psi^{\mathbf{j}}} |\psi^{\mathbf{j}}(y)| dy$. ∎

Though in general a decomposition (2.14) with small separation rank r is a complicated numerical task, in many interesting applications efficient approximation methods are available. In particular, for a class of multi-variate functions (say, for certain shift-invariant Green's kernels in $\mathbb{R}^d$) it is possible to obtain a dimensionally-independent Kronecker rank $r = \mathcal{O}(\log n |\log \varepsilon|)$, e.g., based on *Sinc*-quadrature methods or an approximation by exponential sums (see case-study examples in [12, 3, 18]).

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

Lemma 2.3 *Let $g : \Omega \to \mathbb{R}$ be approximated by a separable expansion*

$$g_{\mathbf{r}}(\boldsymbol{\zeta}) := \sum_{k_1=1}^{r_1} ... \sum_{k_d=1}^{r_d} b_{k_1...k_d} \Phi_{k_1}^{(1)}(\zeta^{(1)}) \cdots \Phi_{k_d}^{(d)}(\zeta^{(d)}) \approx g, \quad \zeta^{(\ell)} \in \mathbb{R}^2, \quad 1 \le \ell \le d, \tag{2.16}$$

where $b_{k_1...k_d} \in \mathbb{R}$. Then the FGT(C), corresponding to the choice

$$V_{k_\ell}^{(\ell)} = \left\{ \int \Phi_{k_\ell}^{(\ell)}(\zeta_i^{(\ell)}) \psi_\ell^j(y_\ell) dy_\ell \right\}_{(i,j)\in\mathcal{M}_\ell} \in \mathbb{R}^{\ I_\ell \times \ J_\ell}, \quad \ell = 1, ..., d, \quad k_\ell = 1, ..., r_\ell \tag{2.17}$$

with $\zeta_i^{(\ell)} = (x_i^{(\ell)}, y_\ell)$ provides the error estimate

$$\|\mathcal{A}(g) - \mathcal{A}_{(r)}(g_r)\|_\infty \le C \|g - g_r\|_{L^\infty(\Omega)}.$$

Proof. In the FGT(C) case, by the construction of $\mathcal{A}_{(\mathbf{r})}$, we have

$$\begin{aligned} \|\mathcal{A} - \mathcal{A}_{(\mathbf{r})}\|_\infty &\le \max_{x\in\omega_{\mathbf{d}}} \left| \int_\Omega \left(g(x,y) - \sum_{k_1=1}^{r_1} ... \sum_{k_d=1}^{r_d} b_{k_1...k_d} \Phi_{k_1}^{(1)}(\zeta^{(1)}) \cdots \Phi_{k_d}^{(d)}(\zeta^{(d)}) \right) \psi^{\mathbf{j}}(y) dy \right| \\ &\le \|g - g_r\|_{L^\infty(\Omega)} \max_{\mathbf{j}} \int_{supp\ \psi^{\mathbf{j}}} |\psi^{\mathbf{j}}| \left|\psi^{\mathbf{j}}(y)\right| dy, \end{aligned}$$

which proves the assertion. ∎

Next we discuss the constructive CP and Tucker decomposition of FGTs applied to a general class of *analytic generating functions* characterised in terms of their Laplace transform. The construction is based on *Sinc-approximation methods*.

2.3.2 Error Bounds for Canonical Decomposition of FGTs

We use constructive approximation based on the *Sinc*-quadrature and *Sinc*-interpolation methods. For the readers convenience we recall the standard approximation results by the *Sinc*-methods (cf. [23, 9]). First, we introduce the Hardy space $H^1(D_\delta)$ as the set of all complex-valued functions f, which are analytic in the strip

$$D_\delta := \{z \in \mathbb{C} : |\Im\mathfrak{m}\, z| < \delta\}, \tag{2.18}$$

such that

$$N(f, D_\delta) := \int_{\partial D_\delta} |f(z)|\, |\mathrm{d}z| = \int_{\mathbb{R}} \left(|f(x + i\delta)| + |f(x - i\delta)|\right) \mathrm{d}x < \infty.$$

Given $f \in H^1(D_\delta)$, $\mathfrak{h} > 0$, and $M \in \mathbb{N}_0$, the corresponding sinc-quadrature reads as

$$T_M(f, \mathfrak{h}) := \mathfrak{h} \sum_{k=-M}^{M} f(k\mathfrak{h}) \approx \int_{\mathbb{R}} f(\xi) \mathrm{d}\xi. \tag{2.19}$$

Proposition 2.4 *Let $f \in H^1(D_\delta)$, $\mathfrak{h} > 0$, and $M \in \mathbb{N}_0$ be given. If*

$$|f(\xi)| \le C \exp(-b|\xi|) \qquad \text{for all } \xi \in \mathbb{R} \text{ with } b, C > 0, \tag{2.20}$$

then the quadrature error satisfies

$$\left|\int_{\mathbb{R}} f(\xi)\mathrm{d}\xi - T_M(f,\mathfrak{h})\right| \le Ce^{-\sqrt{2\pi\delta bM}} \quad \textit{with } \mathfrak{h} = \sqrt{2\pi\delta/bM}$$

and with a positive constant C depending only on f, δ, b (cf. [23]). If f possesses the hyper-exponential decay

$$|f(\xi)| \le C\exp(-b\mathrm{e}^{a|\xi|}) \qquad \textit{for all } \xi \in \mathbb{R} \quad \textit{with } a, b, C > 0, \tag{2.21}$$

then the choice $\mathfrak{h} = \log(\frac{2\pi aM}{b})/\,(aM)$ leads to (cf. [9])

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

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

We consider a class of multi-variate functions $g : \mathbb{R}^d \to \mathbb{R}$ parametrised by $g(\boldsymbol{\zeta}) = G(\rho(\boldsymbol{\zeta})) \equiv G(\rho)$ with $\rho \equiv \rho(\boldsymbol{\zeta}) = \rho_1(\zeta^{(1)}) + ... + \rho_d(\zeta^{(d)}) > 0$, $\rho_\ell : \mathbb{R}^2 \to \mathbb{R}_+$, where the univariate function $G : \mathbb{R}_+ \to \mathbb{R}$ can be represented via the Laplace transform

$$G(\rho) = \int_{\mathbb{R}_+} \mathcal{G}(\tau)e^{-\rho\tau}d\tau.$$

The FGT(C) approximation corresponds to $p = 2$, $\zeta^{(\ell)} = (x_\ell, y_\ell)$ (cf. Definition 2.5). Without loss of generality, we introduce one and the same scaling function

$$\phi^i(\cdot) = \phi(\cdot + (i-1)h), \quad i \in I_n, \tag{2.22}$$

for all spatial dimensions $\ell = 1, ..., d$, where $h > 0$ is the mesh parameter. We simplify further and set $\rho \equiv \rho(\boldsymbol{\zeta}) = \sum_{\ell=1}^d \rho_0(\zeta^{(\ell)})$, i.e.,

$$\rho_\ell = \rho_0(x_\ell, y_\ell) \quad (\ell = 1, ..., d) \quad \text{with } \rho_0 : [a,b]^2 \to \mathbb{R}_+. \tag{2.23}$$

For $i \in I_n$, let $\{\bar{x}_i\}$ be the set of cell-centred collocation points on $[a,b]$. For each $i, j \in I_n$, we introduce the parameter dependent integral

$$\Psi_{i,j}(\tau) := \int_{\mathbb{R}^2} e^{-\rho_0(\bar{x}_i,y)\tau}\phi(y + (j-1)h)dy, \quad \tau \ge 0. \tag{2.24}$$

Theorem 2.5 *(FGT(C) approximation). Assume (**a**)-(**c**) below:*

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

*(**b**) for all $(\mathbf{i}, \mathbf{j}) \in \mathcal{I} \times \mathcal{J}$ the transformed integrand*

$$f(z) := \varphi'(z)\mathcal{G}(\varphi(z)) \prod_{\ell=1}^d \Psi_{i_\ell j_\ell}(\varphi(z)) \tag{2.25}$$

belongs to the Hardy space $H^1(D_\delta)$ with $N(f, D_\delta) < \infty$ uniformly in $(\mathbf{i}, \mathbf{j})$;

*(**c**) the function $f(t)$, $t \in \mathbb{R}$, in (2.25) has either exponential (c1) or hyper-exponential (c2) decay as $t \to \pm\infty$.*

*Under the assumptions (**a**)-(**c**), we have that, for each $M \in \mathbb{N}_+$, the FGT(C), $\mathcal{A}(g)$, defined on $[a,b]^d$ allows an exponentially convergent super-symmetric[1] CP decomposition $\mathcal{A}_{(r)} \in \boldsymbol{\mathcal{C}}_r$ with $V_k^{(\ell)}$ as in (2.15), where the expansion (2.14) is obtained by the substitution of f from (2.25) into the* sinc*-quadrature (2.19), such that we have*

$$\|\mathcal{A}(g) - \mathcal{A}_{(r)}\|_\infty \le Ce^{-\alpha M^\nu} \quad \textit{with } r = 2M+1, \tag{2.26}$$

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

[1] A d-th order tensor is called super-symmetric if it is invariant under arbitrary permutations of indices in $\{1, ..., d\}$

Proof. First, we notice that by definition

$$a_{\mathbf{ij}} = \int_{\mathbb{R}_+} \mathcal{G}(\tau) \prod_{\ell=1}^{d} \Psi_{i_\ell j_\ell}(\tau) d\tau = \int_{\mathbb{R}} f(t) dt \quad \text{for } (\mathbf{i},\mathbf{j}) \in \mathcal{I} \times \mathcal{J}. \tag{2.27}$$

We now apply the *Sinc*-quadrature to the transformed integrand f to obtain

$$T_M(f,\mathfrak{h}) := \mathfrak{h} \sum_{k=-M}^{M} f(k\mathfrak{h}) \approx \int_{\mathbb{R}} f(t) dt, \quad (\mathbf{i},\mathbf{j}) \in \mathcal{I} \times \mathcal{J}$$

with

$$\left| \int_{\mathbb{R}} f(t) dt - T_M(f,\mathfrak{h}) \right| \le C e^{-\alpha M^\nu},$$

and with the respective α, ν (see Proposition 2.4). Combining this estimate with (2.27) and taking into account the separability property of the exponential prove the assertion for all $(\mathbf{i},\mathbf{j}) \in \mathcal{I} \times \mathcal{J}$. Noticing that our quadrature does not depend on the index $(\mathbf{i},\mathbf{j})$ completes the proof. ∎

Theorem 2.5 proves the existence of a CP decomposition to the FGT $\mathcal{A}(g)$ with the Kronecker rank $r = \mathcal{O}(|\log\varepsilon| \log 1/h)$ (in the case (c2)) or $r = \mathcal{O}(\log^2 \varepsilon)$ (in the case (c1)), which provide an approximation of order $\mathcal{O}(\varepsilon)$. In our applications we usually have $1/h = \mathcal{O}(n)$, where n is the number of grid-points in one spacial direction. Theorem 2.5 typically applies to translation invariant or spherically symmetric functions (see examples in §3).

2.3.3 Error Bounds for Tucker Decomposition of FGTs

For the class of applications with more general than translation invariant functions the analytic separation methods are based on tensor-product interpolation. This leads to the rank-$(r_1, ..., r_d)$ Tucker decomposition with small rank parameters r_ℓ. Again we recall the related results on the *Sinc*-interpolation method. Let

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

be the k-th sinc-function with step size $\mathfrak{h}$, evaluated at x with the sinc-function given by

$$\mathrm{sinc}(z) = \frac{\sin(\pi z)}{\pi z}, \quad z \in \mathbb{C}.$$

The classical *Sinc* interpolant (cardinal series representation) is given by

$$C_M(f,\mathfrak{h}) = \sum_{\nu=-M}^{M} S(\nu,\mathfrak{h}) f(\nu\mathfrak{h}) \approx f. \tag{2.28}$$

If (2.20) holds then the interpolation error satisfies (cf. [23])

$$\|f - C_M(f,\mathfrak{h})\|_\infty \le C M^{1/2} e^{-\sqrt{\pi\delta b M}} \quad \text{with } \mathfrak{h} = \sqrt{\pi\delta/bM}, \tag{2.29}$$

where δ specifies the width of the strip D_δ in (2.18). Assuming the hyper-exponential decay of f as in (2.21), we obtain (cf. [9])

$$\|f - C_M(f,\mathfrak{h})\|_\infty \le C \frac{N(f,D_\delta)}{2\pi\delta} \mathrm{e}^{-\pi\delta aM/\log(\pi aM/b)} \quad \text{with } \mathfrak{h} = \log(\frac{\pi aM}{b})/(aM). \tag{2.30}$$

The *Sinc*-interpolation method can be extended to the multi-dimensional case. For each $\ell = 1, ..., d$, let $g_\ell(\cdot) : \delta_\ell = [a_0, b_0] \to \mathbb{R}$ be a univariate parameter-dependent function in variable $\zeta^{(\ell)}$, which is the restriction of a multi-variate function $g(\zeta^{(1)}, ..., \zeta^{(d)})$ onto δ_ℓ with fixed remaining variables $\zeta^{(1)}, ..., \zeta^{(\ell-1)}, \zeta^{(\ell+1)}, ..., \zeta^{(d)}$. Suppose that $g_\ell(\cdot)$ satisfies all the regularity and decay conditions above, uniformly in $\ell = 1, ..., d$. It is shown

in [12] that the *tensor-product Sinc interpolation* $\mathbf{C}_M g := C_M^{(1)} ... C_M^{(d)} g$ with respect to d variables, provides the exponential error estimate

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

with the stability (Lebesgue) constant $\Lambda_M = \mathcal{O}(\log M)$, and where $C_M^{(\ell)} g = C_M^{(\ell)}(g, \mathfrak{h})$ denotes the univariate *Sinc* interpolation from (2.28) applied to the variable $\zeta_\ell \in I_\ell$.

For a class of analytic functions with point singularities the expansion (2.16) can be derived via tensor-product *Sinc*-interpolation applied with respect to variables $\rho_1, ..., \rho_d$.

Theorem 2.6 *Assume that all conditions in Theorem 2.5 are satisfied. Then the FGT(C), $\mathcal{A}(g)$, allows an exponentially convergent rank-$(r, ..., r)$ Tucker decomposition $\mathcal{A}_{(\mathbf{r})} \in \boldsymbol{\mathcal{T}}_{\mathbf{r}}$ with $V_{k_\ell}^{(\ell)}$ as in (2.17), where $\Phi_{k_\ell}^{(\ell)}(\zeta^{(\ell)}) = \mathrm{sinc}(-a_{k_\ell}\rho_0(\zeta^{(\ell)}))$ with ρ_0 from (2.23) ($\ell = 1, ..., d$), and where $b_{\mathbf{k}}$ are explicitly represented via the* sinc*-interpolation (2.28), such that*

$$\|\mathcal{A}(g) - \mathcal{A}_{(\mathbf{r})}\|_\infty \le C(1 + \log M)^d e^{-\alpha M^\nu} \quad \text{with } r = 2M + 1, \tag{2.31}$$

with $\nu = \frac{1}{2}$, $\alpha = \sqrt{2\pi\delta b}$ in the case (c1) and with $\nu = 1$, $\alpha = \frac{2\pi\delta b}{\log(2\pi a M/b)}$ in the case (c2) as in Theorem 2.5.

Proof. Modifying the proof of Theorem 2.5, we now apply the *Sinc*-interpolation. In particular, the error bounds (2.29) and (2.30) show exponential convergence in M for the tensor-product sinc-interpolant $\mathbf{C}_M g$, which proves the assertion. ∎

The error estimate (2.31) yields $\max_\ell r_\ell = \mathcal{O}(|\log\varepsilon|\delta^{-1})$. In some cases we get the estimate $\delta^{-1} = \mathcal{O}(\log 1/h)$ (cf. [12]).

3 Tensor Approximation of Integral Operators

3.1 Canonical and Tucker Decompositions in $\mathbb{R}^d$

The principal ingredient in the structured tensor-product representation of integral operators in many spatial dimensions is a separable approximation of the multi-variate function representing the kernel of the operator. Given the integral operator $\mathbb{G} : L^2(\Omega) \to L^2(\Omega)$ in $\Omega := [0, 1]^d \in \mathbb{R}^d$, $d \ge 2$,

$$(\mathbb{G}u)(x) := \int_\Omega g(x, y)u(y)\mathrm{d}y, \quad x, y \in \Omega,$$

with some shift-invariant kernel function $g(x, y) = g(|x - y|)$, which can be represented in the form

$$g(x, y) = g(\zeta_1, ..., \zeta_d) \equiv g\left(\sqrt{\zeta_1^2 + ... + \zeta_d^2}\right),$$

where $\zeta_\ell = |x_\ell - y_\ell| \in [0, 1]$, $\ell = 1, ..., d$.

To approximate the operator $\mathbb{G}$, we consider a collocation scheme with tensor-product test functions $\psi^{\mathbf{i}}(x_1, ..., x_d)$ as in (2.12).

If the kernel function g allows a global separable approximation, cf. Lemma 2.6, we approximate the collocation stiffness matrix

$$A = \{(\mathcal{A}\phi^{\mathbf{j}})_{|\bar{x}_{\mathbf{i}}}\}_{\mathbf{i},\mathbf{j} \in I_n^d} \in \mathbb{R}^{N\times N}, \quad N = n^d, \quad \bar{x}_{\mathbf{i}} \in \Omega_d,$$

by a matrix $A_{(r)}$ of the form (2.11), where the V_k^ℓ are $n \times n$ matrices given by

$$V_k^\ell = \left\{\int_0^1 \Phi_k^\ell(|\bar{x}_i - y_\ell|)\psi_\ell^j(y_\ell)\mathrm{d}y_\ell\right\}_{i,j=1}^n, \quad \ell = 1, ..., d, \tag{3.1}$$

providing the corresponding error estimate in l_∞ matrix norm. For standard singular kernels (say, Green's kernels) the direct separable approximation is usually not possible. In this case one can apply Theorem 2.9. In both cases we are able to prove the existence of a low Kronecker rank CP approximation for the class of multi-dimensional integral operators.

Note that $\|A - A_{(r)}\|$ can be easily estimated in, say, the Frobenius matrix norm.

When using the tensor-product *Sinc* interpolation, the function $\Phi_k^\ell(|u-v|)$ can be proved to be asymptotically smooth. For the class of kernel functions approximated by exponential sums, the factor $\Phi_k^\ell(|u-v|)$ even appears to be globally smooth (indeed, it is the entire function). Hence, the canonical components V_k^ℓ can be further approximated in the $\mathcal{H}$-matrix format (cf. [13]). In the case of uniform grids also the Toeplitz-type structure can be used to represent $n \times n$ matrices V_k^ℓ.

For the class of translation-invariant kernels (see [12] and examples below), we obtain a dimensionally independent bound

$$r = \mathcal{O}(\log\left(h^{-1}\right)\log\left(\varepsilon^{-1}\right)\log\left(\log\varepsilon^{-1}\right)).$$

Following Definition 2.1, we introduce the d-th order FGT(C) representing the integral operator $\mathbb{G}$,

$$\mathcal{A} \equiv \mathcal{A}(g) := [a_{\mathbf{m}_1\ldots\mathbf{m}_d}] \in \mathbb{R}^{\mathcal{M}_1\times\ldots\times\mathcal{M}_d}.$$

Assume that the kernel function $g(x,y) \equiv g(\zeta^{(1)},...,\zeta^{(d)})$ allows a separable approximation (2.16) via the sinc-interpolation, so that the approximation converges exponentially in $r = \max_\ell r_\ell$ (see Theorem 2.6). Then the associated rank-$(r_1,...,r_d)$ Tucker decomposition (2.10) in $\boldsymbol{\mathcal{T}}_{\mathbf{r}}$ (cf. (2.10)) is specified by the Kronecker factors $V_{k\ell}^{(\ell)} \in \mathbb{R}^{\mathcal{M}_\ell}$, explicitly defined by (2.17). Let $\mathbf{r} = (r,...,r)$. Theorem 2.6 now yields the error estimate

$$\|\mathcal{A}(g) - \mathcal{A}_{(\mathbf{r})}\|_\infty \le Ce^{-\alpha M^\nu} \quad \text{with } r = 2M+1, \tag{3.2}$$

and with constants α, ν from (2.31).

As it was already mentioned, (3.2) yields $\max_\ell r_\ell = \mathcal{O}(|\log\varepsilon|\delta^{-1})$ with δ from (2.18). In turn, for a class of shift-invariant kernels we get the estimate $\delta^{-1} = \mathcal{O}(\log n)$. In general, given a tolerance $\varepsilon > 0$, we have the bound

$$r = \mathcal{O}\left(\left[\log(n)\log\left(\varepsilon^{-1}\right)\log\left(\log\varepsilon^{-1}\right)\right]^{d-1}\right).$$

The numerical complexity of the Tucker decomposition is estimated by $drn^2 + r^d$. The storage cost for the corresponding Tucker approximation combined with hierarchical matrices has the complexity $drn\log^q n + r^d$. Notice that the Tucker approximation can be applied to more general kernel functions compared with the canonical representation (as it was already mentioned, the latter is usually restricted to the class of translation-invariant kernels).

3.2 Application to the Newton Potential

Let $x, y \in \mathbb{R}^d$, $p = 2$, and define $\rho = |x-y|^2 = \zeta_1^2 + ... + \zeta_d^2$ with $\zeta_\ell = x_\ell - y_\ell : \mathbb{R}^2 \to \mathbb{R}$, $\boldsymbol{\zeta} \in \mathbb{R}^{2d}$. The family of functions

$$g(x,y) \equiv g(\boldsymbol{\zeta}) := 1/\rho^\lambda \quad \text{with } \lambda \in \mathbb{R}_{>0},$$

arises in potential theory, in quantum chemistry and in computational gas dynamics (cf. [18]). The choice $\lambda = 1/2$ corresponds to the classical Newton potential, while $\lambda = -1/2$ refers to the Euclidean distance function. Low separation rank decomposition to the multi-variate functions $1/\rho$, $1/\sqrt{\rho}$ and to the related Galerkin approximations were discussed in [12, 13, 14, 19], while the kernel function ρ^μ, $\mu \in \mathbb{R}$, was considered in [18].

Let us take a closer look to the collocation-type FGT corresponding to the Newton potential $1/\sqrt{\rho}$ in the hypercube $[-R,R]^d \in \mathbb{R}^d$. As a basic example, we consider piecewise constant finite elements on the uniform grid with step-size $h > 0$, defined by scaling functions $\phi(x) = \psi(x)$ associated with a tensor-product grid. Again, we let $\{\bar{x}_i\}$ be the set of cell-centred collocation points.

In our case, for the function in (2.24) we have $\rho_0(x,y) = (x-y)^2$ $(x, y \in \mathbb{R})$, hence making use of the Gaussian transform

$$\frac{1}{\sqrt{\rho}} = \frac{2}{\sqrt{\pi}}\int_{\mathbb{R}_+} e^{-\rho\tau^2}d\tau,$$

we obtain

$$\Psi_{i,j}(\tau) = \Psi_{|i-j|}(\tau) := \int_{\mathbb{R}^2} e^{-\tau^2(\bar{x}_i - y)^2} \phi^j(y)\, dy, \quad \tau \geq 0, \quad i,j \in I_n$$

(see (2.24), (2.22) for the definition of $\Psi_{i,j}$, ϕ^j).

Lemma 3.1 *The FGT(G) for the Newton potential $1/\sqrt{\rho}$ allows a CP approximation in the hypercube $[-R,R]^d \in \mathbb{R}^d$ with exponential convergence rate (independent of d) as in (2.31), where $\nu = 1/2$.*

Proof. We apply Theorem 2.5. To check the condition (a), let us choose the analyticity domain as a sector $\Omega_{\mathcal{G}} := \{w \in \mathbb{C} : |arg(w)| < \delta\}$ with apex angle $0 < 2\delta < \pi/2$ (here $\mathcal{G} = 1$), and then apply the conformal map

$$\varphi^{-1} : \Omega_{\mathcal{G}} \to D_\delta \quad \text{with} \quad w = \varphi(z) = e^z,\ \varphi^{-1}(w) = \log(w)$$

(cf. Theorem 2.5(a)).

To check condition (b) of Theorem 2.5, first, we notice that the transformed integrand

$$f(z) := \exp(z) \prod_{\ell=1}^{d} \Psi_{i_\ell j_\ell}(\varphi(z))$$

belongs to the Hardy space $H^1(D_\delta)$. In fact, introducing the error function erf by

$$\mathrm{erf}(t) := \frac{2}{\sqrt{\pi}} \int_0^t e^{-\tau^2}\, \mathrm{d}\tau, \tag{3.3}$$

we calculate the explicit representation

$$\Psi_{i_\ell, j_\ell}(\tau) = \Psi_i(\tau) = \frac{\pi^{\frac{d-1}{2d}}}{2\,\tau} \big\{ \mathrm{erf}(\tau\, ih) - \mathrm{erf}(\tau\,(i-1)h) \big\}, \tag{3.4}$$

with $x_{i_\ell}^\ell = x_i = (i-1)h$, $n_\ell = n$, $h = b/n$ (uniform grid spacing) for $i = i_\ell - j_\ell + 1 = 1, ..., n$, $\ell = 1, ..., d$. Since $\mathrm{erf}(z)/z$ is an entire function it proves the required analyticity of f.

Now we estimate the constant $N(f, D_\delta)$ applying arguments similar to those in [19] (cf. Lemma 4.7).

Finally, we check condition (c1). Using properties of the erf-function as $t \to \pm\infty$, we obtain the required asymptotical behaviour of $f(t)$, $t \to \pm\infty$, with $d \geq 2$. This completes our proof. ∎

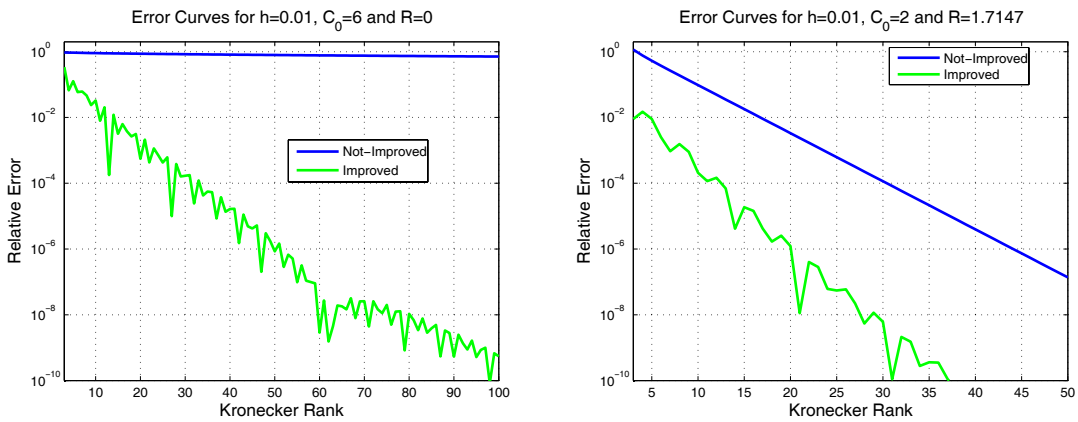

Figure 3.1: Comparison between the improved and not-improved sinc-quadratures for $d = 3$, $h = 0.01$, $R = 0$ (left) and $R = \sqrt{3}$ (right).

Lemma 3.1 proves the exponential convergence of the canonical decomposition with $\nu = 1/2$. However, it is also possible to apply the improved quadrature with hyper-exponential decay of the integrand which leads

to the true exponential convergence with $\nu = 1$. Using a variable transformation $t = \sinh(u)$ and taking advantage of the symmetry of the integrand we obtain the quadrature formula

$$I = \int_{\mathbb{R}} f(t)\,\mathrm{d}t = \int_{\mathbb{R}^+} 2\,\cosh(u)\,f(\sinh(u))\,\mathrm{d}u \approx \sum_{k=0}^{M} w_k^{(M)}\,f(t_k^{(M)}) =: I_M \tag{3.5}$$

with

$$t_k^{(M)} := \sinh(k\mathfrak{h}_M) \tag{3.6}$$

and

$$w_k^{(M)} := \begin{cases} \mathfrak{h}_M & \text{for } k = 0 \\ 2\,\mathfrak{h}_M \cosh(k\mathfrak{h}_M) & \text{for } k > 0 \end{cases} \tag{3.7}$$

with the choice $\mathfrak{h}_M = C_0 \frac{\log(M)}{M}$ for some C_0 (see Lemma 5.1 in [12]).

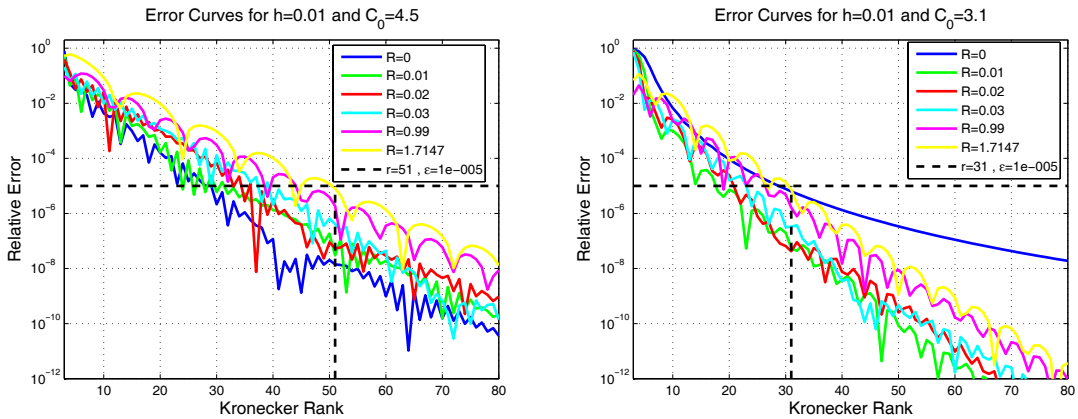

Figure 3.2: Non-optimised (left) and optimised (right) errors for $h = 0.01$, $\varepsilon = 10^{-5}$.

In the numerical illustrations we consider the case $d = 3$. Due to the Toeplitz structure of the $n \times n$ matrices V_k^ℓ, in the numerical experiments below we control the accuracy of our quadrature-based decompositions only for a fixed index $i_\ell = 1$ and vary the index $j_\ell = 1, ..., n$ ($\ell = 1, ..., 3$). Hence, in our notation we distinguish the distance R from the observation point to the origin: for example, $R = 0$ corresponds to $j_\ell = 1$, while $R = \sqrt{3}$ corresponds to $j_\ell = n$ ($\ell = 1, ..., 3$).

First we demonstrate the advantage of the improved quadrature (3.5), see Figure 3.1.

For a fixed number of quadrature terms M, in order to obtain uniform error control for all indices $j = j_\ell = 1, ..., n$, we optimise the quadrature with respect to the factor C_0 in $\mathfrak{h}_M = C_0 \frac{\ln(M)}{M}$, such that the quadrature errors are approximately equalised for two limiting cases $R = 1$ and $R = \sqrt{3}$. Then the error for all intermediate values of R lie in the "corridor" between the above mentioned error bounds. Figure 3.2 presents non-optimised (left) and optimised (right) errors considered for the limiting values of R (top) and other representative data (bottom), for $h = 0.01$ and $\varepsilon = 10^{-5}$. For our quadrature-based decompositions we observe the exponential convergence in the Kronecker rank.

Further reduction of the Kronecker rank can be achieved by applying the so-called near-far field decomposition. It is based on the observation that the quadrature optimisation for the off-diagonal part of the target matrix (i.e., without the diagonal elements corresponding to $j_\ell \geq 2$) leads to a much smaller Kronecker rank compared with an approximation of the whole matrix. In this case the low Kronecker rank representation of the complete matrix is obtained by adding a rank-1 term representing the diagonal part ($j_\ell = 1$). The numerical results are depicted in Figure 3.3 (indicate the rank reduction from 30 to 20).

References

[1] G. Beylkin and M.J. Mohlenkamp: *Numerical operator calculus in higher dimensions.* Proc. Natl. Acad. Sci. USA **99** (2002), 10246-10251.

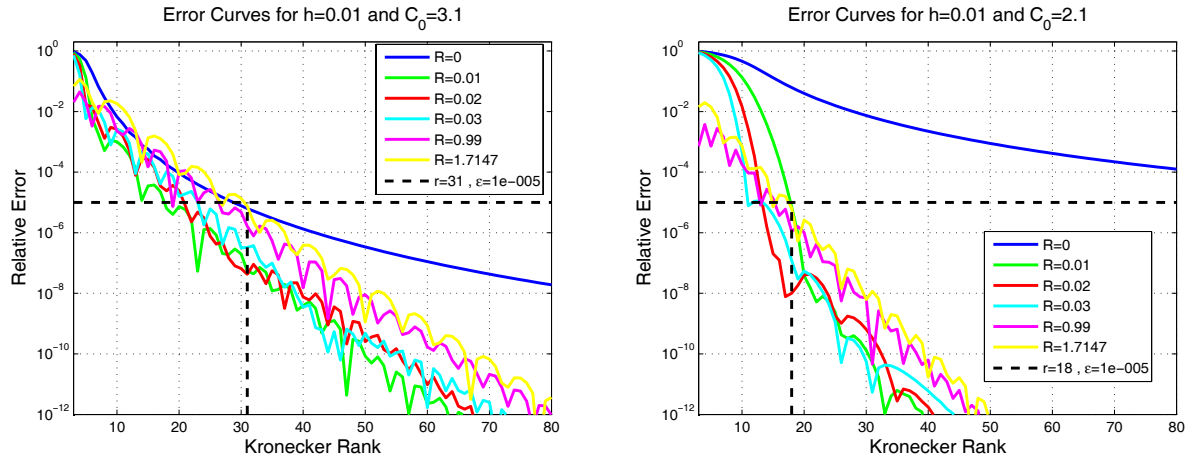

Figure 3.3: Optimal quadratures without (left) and with near-far field decomposition (right) for $h = 10^{-2}$ and $\varepsilon = 10^{-5}$.

[2] G. Beylkin and M.J. Mohlenkamp: *Algorithms for numerical analysis in higher dimensions.* SIAM J. Sci. Comput. **26** (2005) 2133-2159.

[3] G. Beylkin and L. Morzón: *On approximation of functions by exponential sums.* Applied and Comp. Harmonic Analysis **19** (2005) 17-48.

[4] L. De Lathauwer, B. De Moor, J. Vandewalle: *On the best rank-*1 *and rank-*$(R_1, ..., R_N)$ *approximation of higher-order tensors.* SIAM J. Matrix Anal. Appl. **21** (2000) 1324-1342.

[5] L. De Lathauwer, B. De Moor, J. Vandewalle: *Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition.* SIAM J. Matrix Anal. Appl. **26** (2004) 295-327.

[6] M.V. Fedorov, G. Chuev, H.-J. Flad, L. Grasedyck, and B.N. Khoromskij: *Low-rank wavelet solver for the Ornstein-Zernike integral equation.* To appear in Computing 2007.

[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] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *Data-sparse approximation to operator-valued functions of elliptic operators.* Math. Comp. **73** (2004), 1297–1324.

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

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

[11] W. Hackbusch: *Approximation of 1/———x——— by exponentials for wavelet applications.* Computing **76** (2006) 359-366.

[12] W. Hackbusch and B.N. Khoromskij: *Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part I. Separable approximation of multi-variate functions;* Computing **76** (2006), 177-202.

[13] W. Hackbusch and B.N. Khoromskij: *Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part II. HKT representations of certain operators.* Computing **76** (2006), 203-225.

[14] W. Hackbusch, B.N. Khoromskij, and E. Tyrtyshnikov: *Hierarchical Kronecker tensor-product approximation.* J. Numer. Math. **13** (2005), 119-156.

[15] W. Hackbusch, B.N. Khoromskij, and E.E. Tyrtyshnikov: *Approximate iterations for structured matrices.* Preprint 112, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig 2005.

[16] R. Harshman: *Foundation of the PARAFAC procedure: Model and conditions for an "explanatory" multi-mode factor analysis.* UCLA Working Papers in Phonetics **16** (1970), 1-84.

[17] B.N. Khoromskij: *An introduction to structured tensor-product representation of discrete nonlocal operators.* Lecture Notes **27**, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, Leipzig 2005.

[18] 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.).

[19] B.N. Khoromskij: *Structured rank-$(r_1, ..., r_d)$ decomposition of function-related tensors in $\mathbb{R}^d$.* Comp. Meth. Appl. Math. **6** (2006), 2, 194-220.

[20] B.N. Khoromskij and V. Khoromskaia: *Low rank Tucker-type approximation to classical potentials.* Preprint 105, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig 2006.

[21] T. Kolda: *Orthogonal tensor decompositions.* SIAM J. Matrix Anal. Appl. **23** (2001) 243-255.

[22] Ch. Lubich: *On variational approximations in quantum molecular dynamics.* Math. Comp. **74** (2005), 765-779.

[23] F. Stenger: *Numerical methods based on Sinc and analytic functions.* Springer-Verlag, 1993.

[24] V.N. Temlyakov: *Greedy algorithms and M-term approximation with regard to redundant dictionaries.* J. of Approx. Theory **98** (1999), 117-145.

[25] L.R. Tucker: *Some mathematical notes on three-mode factor analysis.* Psychometrika **31** (1966) 279-311.

[26] E.E. Tyrtyshnikov: *Tensor approximations of matrices generated by asymptotically smooth functions.* (Russian) Mat. Sb. **194**, No. 6 (2003), 147–160; translation in Sb. Math. **194** (2003), 941–954.

[27] T. Zhang and G.H. Golub: *Rank-one approximation to high order tensors.* SIAM J. Matrix Anal. Appl. **23** (2001), 534-550.