Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Multigrid Accelerated Tensor Approximation of Function Related Multi-dimensional Arrays

(revised version: November 2009)

by

Boris N. Khoromskij, and Venera Khoromskaia

Preprint no.: 40 2008

SIAM J. SCI. COMPUT.
Vol. 31, No. 4, pp. 3002–3026

MULTIGRID ACCELERATED TENSOR APPROXIMATION OF FUNCTION RELATED MULTIDIMENSIONAL ARRAYS*

B. N. KHOROMSKIJ† AND V. KHOROMSKAIA†

Abstract. In this paper, we describe and analyze a novel tensor approximation method for discretized multidimensional functions and operators in $\mathbb{R}^d$, based on the idea of multigrid acceleration. The approach stands on successive reiterations of the orthogonal Tucker tensor approximation on a sequence of nested refined grids. On the one hand, it provides a good initial guess for the nonlinear iterations to find the approximating subspaces on finer grids; on the other hand, it allows us to transfer from the coarse-to-fine grids the important data structure information on the location of the so-called most important fibers in directional unfolding matrices. The method indicates linear complexity with respect to the size of data representing the input tensor. In particular, if the target tensor is given by using the rank-R canonical model, then our approximation method is proved to have linear scaling in the univariate grid size n and in the input rank R. The method is tested by three-dimensional (3D) electronic structure calculations. For the multigrid accelerated low Tucker-rank approximation of the all electron densities having strong nuclear cusps, we obtain high resolution of their 3D convolution product with the Newton potential. The accuracy of order 10^{-6} in max-norm is achieved on large $n \times n \times n$ grids up to $n = 1.6 \cdot 10^4$, with the time scale in several minutes.

Key words. orthogonal Tucker tensor decomposition, canonical model, discrete convolution, adaptive orthogonal tensor-product basis, multigrid acceleration

AMS subject classifications. 65F30, 65F50, 65N35, 65F10

DOI. 10.1137/080730408

1. Introduction.

In recent years *tensor approximation methods* have opened challenging perspectives for a feasible numerical solution of multidimensional problems arising in large scale electronic and molecular structure calculations [1, 2, 7, 8, 10, 12, 15, 4, 18, 28], in stochastic PDEs [24], and in financial mathematics. Traditional fields of applications of tensor methods include information technologies, chemometrics, signal processing, as well as stochastic models [25, 6, 20, 21].

The success of tensor methods based on low-rank separable approximation can be explained by their intrinsic nearly one-dimensional (1D) data structure organization. In fact, this allows us to avoid the so-called curse of dimensionality inherent in traditional numerical methods, which usually scale exponentially in the physical dimension d. However, there are fundamental difficulties in the systematic promotion of tensor techniques due to

- challenging problems in a rigorous theoretical analysis of tensor methods,
- high-dimensional nonlinear approximation problems,
- the need of the relevant and physically consistent, though normally hidden, well-formatted data-sparse representations of large and highly redundant data in $\mathbb{R}^d$.

Hence efficient rank-structured tensor methods are required for approximation of high-order tensors representing physically relevant functions and operators in $\mathbb{R}^d$,

*Received by the editors July 16, 2008; accepted for publication (in revised form) March 25, 2009; published electronically July 8, 2009. This research was partially supported by the DFG-SPP 1324 (Germany).
http://www.siam.org/journals/sisc/31-4/73040.html
†Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22-26, D-04103 Leipzig, Germany (bokh@mis.mpg.de, vekh@mis.mpg.de).

discretized on large spatial grids of size n^d. The commonly used tensor methods are based on the so-called orthogonal Tucker and canonical models. The main advantage of the orthogonal Tucker approximation is the (robust) construction of a problem adapted orthogonal basis that simultaneously resolves the peculiarities of the approximated object with an optimal tensor rank. In turn, the canonical tensor format allows us to get rid of the curse of dimensionality.

Efficiency of the Tucker-type tensor decomposition for the low-rank approximation of some classical functions and operators in $\mathbb{R}^d$ was demonstrated in [19]. In [18] the Tucker tensor approximation has been applied to the task of accurate computation of integral operators in electronic structure calculations on three-dimensional (3D) Cartesian grids. These computations have been performed on the $n \times n \times n$ uniform grids with the univariate grid size $n \leq 800$, which was sufficient to achieve high accuracy in the case of *pseudopotentials* of some simple molecules. However, computations involving the *all electron densities* of molecules, which contain strong cusps due to the core electron contribution, require much larger spatial grids, thus motivating substantial improvement of the numerical schemes.

Here we describe and analyze a novel tensor approximation method for discretized multidimensional functions and operators in $\mathbb{R}^d$, based on a multigrid acceleration technique.[1] This approach allows us to overcome the limitations of the single grid schemes in higher dimensions and for large grid size. The algorithm is based on the successive reiteration of the orthogonal Tucker tensor approximation on a sequence of nested refined grids. It resembles the so-called nested iterations in the multigrid (MG) method usually applied as an elliptic problem solver or as a preconditioner. Along with a good initial guess for the nonlinear approximation process it provides the transfer of the important data structure information from the coarse-to-fine grids, based on the *maximum energy principle*; see section 2.4. In particular, we construct the initial guess for the alternating least squares (ALS) iteration by using the interpolated side-matrices calculated on the coarser grid. Furthermore, based on the maximum energy principle we extract from the coarser grids the location of dominating columns of the ℓ-mode unfolding matrices. This leads to a fast nonlinear ALS iteration over an almost minimal sufficient subset of directional fibers in the target tensor.

Notice that our approach can be viewed as a tensor method with constraints, namely, the Tucker model with the constraint on the core that is assumed to be presented in the low-rank canonical format. In this regard we mention that another class of tensor decompositions, CANDELINC, considering canonical tensors with constraints on factor matrices, was introduced in [3]; see also the review paper [21].

We apply our technique to both tensors in the canonical rank-R representation and to full format tensors:

- When the target tensor is represented by a rank-R canonical model, the MG accelerated scheme combined with the two-level Tucker-canonical representation described in [19] is proved to have linear scaling in the grid size n and in the input rank R, $O(rRn)$, where r is the Tucker rank. Note that the unigrid approach leads to the polynomial-exponential complexity $O(dRn \min\{R, n\} + dr^{d-1}n \min\{r^{d-1}, n\})$, which might be computationally infeasible for large parameters d, R, and n. We demonstrate the performance of our approach by fast and highly accurate numerical computations in electronic structure calculations, with $R \sim 5000$, $r \sim 20$, and $n \sim 10^4$ (see

[1] This method was presented by the authors at the Conference on Tensor Methods, MPI MIS, Leipzig, 22–24 January, 2008.

section 3).

- For fully populated tensors of size n^d, the MG accelerated tensor method leads to a nested rank-r Tucker-to-canonical approximation that has only linear storage requirements $O(drn + drR)$, where R is now the canonical rank of the Tucker core. The computational cost of such an approximation scales linearly in the size of input data $O(n^d)$, instead of $O(n^{d+1})$-scaling for the standard Tucker approximation. We present the numerical illustrations for the full format tensors of size 512^3 on the example of computing the low-rank representation of a function of electron density ρ, $\rho^{1/3}$, which is an important issue in the problems of density functional theory (see section 2.5).

It should be emphasized that our technique is particularly efficient for the rank-R initial data, which is the case in electronic structure calculations. As was shown in [19], the orthogonal Tucker decomposition reduces dramatically the effective rank for a class of discretized analytic functions with point singularities, which is the case for the solutions of the Hartree–Fock equation. Our numerical tests presented in this paper show that the MG accelerated Tucker approximation method is a good candidate for fast and accurate computation of the Hartree potential with all electron densities for some simple molecules. Based on the low Tucker rank approximation of electron density, we obtain high resolution of its convolution product with the Newton potential in three dimensions up to 10^{-6} in the max-norm and for large $n \times n \times n$ grids in the range $n \le 1.6 \cdot 10^4$. The total computational time on such grids amounts to only several minutes in MATLAB implementation on a standard SUN cluster. For verification of our results we used the corresponding quantities computed by the MOLPRO program with reference accuracy.

The rest of the paper is organized as follows. In section 2 we discuss the basic methods of rank-structured tensor approximation. Theorem 2.5 proves the error estimate for the reduced higher-order SVD (HOSVD) applied to the canonical target tensor. We describe the MG accelerated Tucker approximation applied to different tensor classes. Theorem 2.7 proves linear complexity for the MG algorithm applied to the rank-R canonical input. Section 3 presents extensive numerical tests on the data from electronic structure calculations, illustrating the efficiency of the new algorithm applied to the large 3D spatial grids.

2. Rank-structured tensor approximation.

2.1. Basic definitions. A tensor of order d is a multidimensional array of data whose elements are referred to by using a tensor-product index set $\mathcal{I} = I_1 \times \cdots \times I_d$. We use the common notation

$$V = [v_{i_1,\dots,i_d} : i_\ell \in I_\ell] \in \mathbb{R}^{\mathcal{I}}, \quad I_\ell = \{1,\dots,n_\ell\}, \quad \ell = 1,\dots,d,$$

to denote a dth-order tensor and $\mathbf{n}$ for the d-tuple $(n_1,\dots,n_d)$. A tensor V is an element of the linear space $\mathbb{V}_{\mathbf{n}} = \otimes_{\ell=1}^{d} \mathbb{V}_\ell$ of real-valued (complex-valued) dth-order tensors with $\mathbb{V}_\ell = \mathbb{R}^{I_\ell}$, and is equipped with the Euclidean *inner product* $\langle \cdot, \cdot \rangle : \mathbb{V}_{\mathbf{n}} \times \mathbb{V}_{\mathbf{n}} \to \mathbb{R}$, defined as

$$\langle V, U \rangle := \sum_{(i_1,\dots,i_d) \in \mathcal{I}} v_{i_1,\dots,i_d} u_{i_1,\dots,i_d} \quad \text{for } V, U \in \mathbb{V}_{\mathbf{n}}. \tag{2.1}$$

The related Frobenius norm is $\|V\|_F := \sqrt{\langle V, V \rangle}$. Notice that a vector is an order-1 tensor, while a matrix is an order-2 tensor such that our tensor norm coincides with the 2-norm of vectors and the Frobenius norm of matrices, respectively.

Multilinear algebraic operations (including visualization) with tensors of order d ($d \geq 3$) can be reduced to the standard linear algebra by the *unfolding of a tensor* into a matrix. The unfolding of a tensor along mode ℓ is a matrix of dimension $n_\ell \times (n_{\ell+1} \dots n_d n_1 \dots n_{\ell-1})$, further denoted by $V_{(\ell)}$, whose columns are the respective columns of V along the ℓth mode.

Another important tensor operation is the so-called *contracted product* of two tensors. In the following, we frequently use its special case of *tensor-matrix multiplication* along mode ℓ. Given a tensor $V \in \mathbb{R}^{I_1 \times \dots \times I_d}$ and a matrix $M \in \mathbb{R}^{J_\ell \times I_\ell}$, we define the respective mode-ℓ tensor-matrix product by

$$U = V \times_\ell M \in \mathbb{R}^{I_1 \times \dots \times I_{\ell-1} \times J_\ell \times I_{\ell+1} \dots \times I_d},$$

where

$$u_{i_1,\dots,i_{\ell-1},j_\ell,i_{\ell+1},\dots,i_d} = \sum_{i_\ell=1}^{n_\ell} v_{i_1,\dots,i_{\ell-1},i_\ell,i_{\ell+1},\dots,i_d} m_{j_\ell,i_\ell}, \quad j_\ell \in J_\ell.$$

In terms of unfolding matrices we then obtain

$$U_{(\ell)} = M V_{(\ell)}.$$

The tensor-matrix product can be applied successively along several modes, and it can be shown to be commutative:

$$(V \times_\ell M) \times_m P = (V \times_m P) \times_\ell M = V \times_\ell M \times_m P, \quad \ell \neq m.$$

The repeated (iterated) mode-ℓ tensor-matrix product for matrices M and P of appropriate dimensions can be simplified as follows:

$$(V \times_\ell M) \times_\ell P = V \times_\ell (MP).$$

Assume for simplicity that $\dim \mathbb{V}_\ell = \#I_\ell = n$ for all $\ell = 1, \dots, d$; then the number of entries in V amounts to n^d, hence growing exponentially in d.

To get rid of exponential scaling in the dimension, approximate representations in some classes $\mathcal{S} \subset \mathbb{V}_{\mathbf{n}}$ of data-sparse "rank structured" tensors will be applied. As the simplest rank structured ansatz, we make use of rank-1 tensors. Specifically, the *outer product* of vectors $t_\ell = \{t_{\ell,i_\ell}\}_{i_\ell \in I_\ell} \in \mathbb{V}_\ell$ ($\ell = 1, \dots, d$) forms the canonical rank-1 tensor

$$T \equiv [t_{\mathbf{i}}]_{\mathbf{i} \in \mathcal{I}} = t_1 \otimes \dots \otimes t_d \in \mathbb{V}_{\mathbf{n}} \quad \text{with entries} \quad t_{\mathbf{i}} = t_{1,i_1} \cdots t_{d,i_d},$$

which requires only dn numbers to store it (now linear scaling in the dimension). When $d = 2$, the outer product of two vectors represents a rank-1 matrix.

In the present paper we apply data-sparse representation of high-order tensors based on the Tucker, canonical, and mixed models.

Given the vector-valued rank parameter $\mathbf{r} = (r_1, \dots, r_d)$, the *rank-*$(r_1, \dots, r_d)$ *Tucker approximation* [26, 6] is based on subspaces

$$\mathbb{T}_{\mathbf{r}} := \otimes_{\ell=1}^d \mathbb{T}_\ell \quad \text{of } \mathbb{V}_{\mathbf{n}} \quad \text{for certain } \mathbb{T}_\ell \subset \mathbb{V}_\ell$$

with $r_\ell := \dim \mathbb{T}_\ell \leq n$. We denote by $\boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$ (or simply $\boldsymbol{\mathcal{T}}_{\mathbf{r}}$) the subset of tensors in $\mathbb{V}_{\mathbf{n}}$ represented in the so-called Tucker format

$$V_{(\mathbf{r})} = \sum\nolimits_{\nu_1=1}^{r_1} \dots \sum\nolimits_{\nu_d=1}^{r_d} \beta_{\nu_1,\dots,\nu_d} t_1^{\nu_1} \otimes \dots \otimes t_d^{\nu_d}, \tag{2.2}$$

with some vectors $t_\ell^{\nu_\ell} \in \mathbb{V}_\ell = \mathbb{R}^{I_\ell}$ $(1 \le \nu_\ell \le r_\ell)$, which form the orthonormal basis of $\mathbb{T}_\ell = \mathrm{span}\{t_\ell^\nu\}_{\nu=1}^{r_\ell}$ $(\ell = 1, \dots, d)$. With fixed $\mathbf{r}$ and $\mathbf{n}$, we can write

$$\boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}} := \{\otimes_{\ell=1}^d \mathbb{T}_\ell : \dim \mathbb{T}_\ell = r_\ell\}.$$

In this paper the parameter

$$r = \max_\ell\{r_\ell\}$$

is called the *maximal Tucker rank*. In our applications, we normally have $r \ll n$, say, $r = O(\log n)$. The coefficient tensor $\boldsymbol{\beta} = [\beta_{\nu_1,\dots,\nu_d}]$, which is an element of a tensor space $\mathbb{B}_\mathbf{r} = \mathbb{R}^{r_1 \times \cdots \times r_d}$, is called the *core tensor*. The space $\mathbb{B}_\mathbf{r}$ can be viewed as the *dual* space with respect to the *primal* tensor space $\mathbb{V}_\mathbf{n}$. Introducing the (orthogonal) side-matrices $T^{(\ell)} = [t_\ell^1 \dots t_\ell^{r_\ell}]$, we then use a tensor-by-matrix contracted product to represent the Tucker decomposition of $V \in \boldsymbol{\mathcal{T}}_\mathbf{r}$:

$$V = \boldsymbol{\beta} \times_1 T^{(1)} \times_2 T^{(2)} \cdots \times_d T^{(d)}. \tag{2.3}$$

Remark 2.1. Notice that the representation (2.3) is not unique, since the tensor V is invariant under directional rotations. In fact, for any set of orthogonal $r_\ell \times r_\ell$ matrices Y_ℓ $(\ell = 1, \dots, d)$, we have the equivalent representation

$$V = \widehat{\boldsymbol{\beta}} \times_1 \widehat{T}^{(1)} \times_2 \widehat{T}^{(2)} \cdots \times_d \widehat{T}^{(d)},$$

with

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} \times_1 Y_1 \times_2 Y_2 \cdots \times_d Y_d, \quad \widehat{T}^{(\ell)} = T^{(\ell)} Y_\ell^T, \quad \ell = 1, \dots, d.$$

Remark 2.2. If the subspaces $\mathbb{T}_\ell = \mathrm{span}\{t_\ell^\nu\}_{\nu=1}^{r_\ell}$ are fixed, then the approximation $V_{(\mathbf{r})} \in \boldsymbol{\mathcal{T}}_\mathbf{r}$ of a given tensor $V \in \mathbb{V}_\mathbf{n}$ is reduced to the orthogonal projection of V onto the particular linear space $\mathbb{T}_\mathbf{r} = \otimes_{\ell=1}^d \mathbb{T}_\ell \subset \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$, that is,

$$V_{(\mathbf{r})} = \sum_{\nu_1,\dots,\nu_d} \langle t_1^{\nu_1} \otimes \cdots \otimes t_d^{\nu_d}, V\rangle\, t_1^{\nu_1} \otimes \cdots \otimes t_d^{\nu_d}.$$

This property is crucial in the computation of the best orthogonal Tucker approximation, where the "optimal" subspaces $\mathbb{T}_\ell$ are recalculated within a nonlinear approximation process.

Given a rank parameter $R \in \mathbb{N}$, we denote by $\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} = \boldsymbol{\mathcal{C}}_R \subset \mathbb{V}_\mathbf{n}$ a set of tensors which can be represented in the *canonical format*

$$V_{(R)} = \sum\nolimits_{\nu=1}^R \xi_\nu u_1^\nu \otimes \cdots \otimes u_d^\nu, \quad \xi_\nu \in \mathbb{R}, \tag{2.4}$$

with normalized vectors $u_\ell^\nu \in \mathbb{V}_\ell$ $(\ell = 1, \dots, d)$. The minimal parameter R in (2.4) is called the rank (or canonical rank) of a tensor. Introducing the side-matrices corresponding to representation (2.4),

$$U^{(\ell)} = [u_\ell^1 \dots u_\ell^R]$$

and the diagonal tensor $\boldsymbol{\xi} := \mathrm{diag}\{\xi_1, \dots, \xi_R\}$ such that $\xi_{\nu_1,\dots,\nu_d} = 0$ except when $\nu_1 = \cdots = \nu_d$ with $\xi_{\nu,\dots,\nu} = \xi_\nu$ $(\nu = 1, \dots, R)$, we obtain the equivalent representation

$$V_{(R)} = \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d U^{(d)}. \tag{2.5}$$

In the following, to simplify the discussion of complexity issues, we assume that $r_\ell = r$ $(\ell = 1, \dots, d)$. The storage requirements for the Tucker (resp., canonical) decomposition is given by $r^d + drn$ (resp., $R + dRn$), where usually $r \ll R$. In turn, the maximal canonical rank of the Tucker representation is r^{d-1}.

Since the Tucker core still presupposes r^d storage space, we introduce further the approximation methods using the mixed (two-level) representation which gainfully combines the beneficial features of both the Tucker and the canonical models [19, 17].

DEFINITION 2.3 (the two-level (primal-dual) Tucker-canonical format). *Given the rank parameters $\mathbf{r}, R$ (normally, $r \ll R$), we denote by $\boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}}$ the subclass of tensors in $\boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$ with the core $\boldsymbol{\beta}$ represented in the canonical format, $\boldsymbol{\beta} \in \boldsymbol{\mathcal{C}}_{R,\mathbf{r}} \subset \mathbb{B}_{\mathbf{r}}$. An explicit representation for $V \in \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}}$ is given by*

$$V = \left(\sum\nolimits_{\nu=1}^{R} \xi_\nu u_1^\nu \otimes \cdots \otimes u_d^\nu \right) \times_1 T^{(1)} \times_2 T^{(2)} \cdots \times_d T^{(d)}. \tag{2.6}$$

Clearly, we have the imbedding $\boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}} \subset \boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$ with the corresponding (nonorthogonal) side-matrices $U^{(\ell)} = [T^{(\ell)} u_\ell^1 \cdots T^{(\ell)} u_\ell^R]$ and scaling coefficients ξ_ν $(\nu = 1, \dots, R)$.

The target tensor $V \in \mathbb{V}_{\mathbf{n}}$ can be approximated by a sum of rank-1 tensors as in (2.2), (2.4), or by using the two-level format $\boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}}$ as in (2.6). In the next sections we present fast and efficient methods to compute the corresponding rank structured approximations in different problem settings.

2.2. Best orthogonal Tucker approximation. Notice that the target tensor A to be approximated may itself belong to a certain class $\mathcal{S}_0 \subset \mathbb{V}_{\mathbf{n}}$ of data structured tensors. Since both $\boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$ and $\boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$ are not linear spaces we are led to the challenging *nonlinear approximation* problem on estimation

$$A \in \mathcal{S}_0 \subset \mathbb{V}_{\mathbf{n}}: \quad \sigma(A, \mathcal{S}) := \inf_{T \in \mathcal{S}} \|A - T\| \tag{2.7}$$

with $\mathcal{S} \in \{\boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}, \boldsymbol{\mathcal{C}}_{R,\mathbf{n}}, \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}}\}$. Subsets of symmetric/antisymmetric or nonnegative tensors can also be utilized. The target tensor A might inherit certain data-sparse structure as follows: $\mathcal{S}_0 \subset \{\mathbb{V}_{\mathbf{n}}, \boldsymbol{\mathcal{C}}_{R_0,\mathbf{n}}, \boldsymbol{\mathcal{T}}_{\mathbf{r}_0,\mathbf{n}}\}$.

We are interested in approximation of function related tensors defined on a sequence of large spatial grids. As the basic nonlinear approximation scheme, we consider the best orthogonal rank-$(r_1, \dots, r_d)$ Tucker model. The main motivations are the following:

- The orthogonal Tucker approximation normally leads to robust minimization processes.
- For the wide class of function related tensors we observe fast exponential convergence in the Tucker rank (almost optimal tensor rank).
- Moreover, the approximations on a sequence of grids usually provide a stable representation of the side-matrices with respect to the grid parameter (good initial guess for the ALS iteration on the fine grid).
- For the case of $\mathcal{S}_0 \subset \boldsymbol{\mathcal{C}}_{R_0,\mathbf{n}}$ input data, by using the *maximal energy principle*, it is possible to transfer from the coarse-to-fine grid the important structured information on the location of dominating columns (fibers) of the ℓ-mode unfolding matrices (reduction of computational cost).

The key point for the efficient solving of the minimization problem (2.7) with $\mathcal{S} = \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$ is its equivalence to the *dual maximization problem* [6],

$$[Z^{(1)}, \dots, Z^{(d)}] = \operatorname{argmax} \left\| [\langle v_1^{\nu_1} \otimes \cdots \otimes v_d^{\nu_d}, A \rangle]_{\nu=\mathbf{1}}^{\mathbf{r}} \right\|_{\mathbb{B}_{\mathbf{r}}}^2, \tag{2.8}$$

over the set of side-matrices $V^{(\ell)}$ in the Stiefel manifold,

$$V^{(\ell)} = [v_\ell^{\nu_1} \dots v_\ell^{\nu_d}] \in \mathcal{M}_\ell := \{Y \in \mathbb{R}^{n\times r_\ell} : Y^T Y = I_{r_\ell \times r_\ell}\} \quad (\ell = 1, \dots, d). \tag{2.9}$$

In view of Remark 2.1, the rotational uniqueness of the maximizer in (2.8) is achieved if one solves this maximization problem in the so-called Grassmann manifold that is the factor space of $\mathcal{M}_\ell$ with respect to the rotational transforms. Under the natural compatibility conditions

$$r_\ell \le \bar{r}_\ell := r_1 \dots r_{\ell-1} r_{\ell+1} \dots r_d, \quad \ell = 1, \dots, d, \tag{2.10}$$

the dual maximization problem (2.8) can be proven to have a global maximum (see [16]). We use the notation $\mathcal{I}_{\bar{r}_\ell}$ for the "single hole" index set of size $r_1 \times \dots \times r_{\ell-1} \times r_{\ell+1} \times \dots \times r_d$, associated with the dual space of coefficient tensors.

The best (nonlinear) Tucker approximation (BTA) based on solving the dual maximization problem (2.8) is usually solved numerically by the ALS iteration with the initial guess computed by HOSVD [5]. In this case, the generic ALS algorithm G_BTA reads as follows [6].

ALGORITHM G_BTA ($\mathbb{V}_{\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$). Given the input tensor $A \in \mathbb{V}_{\mathbf{n}}$.

1. Compute an initial guess $V_0^{(\ell)}$ ($\ell = 1, \dots, d$) for the ℓ-mode side-matrices by "truncated" SVD applied to matrix unfolding $A_{(\ell)}$ (cost $O(n^{d+1})$).
2. For each $q = 1, \dots, d$, and with fixed side-matrices $V^{(\ell)} \in \mathbb{R}^{n\times r_\ell}$, $\ell \ne q$, the ALS iteration optimizes the q-mode matrix $V^{(q)}$ via computing the dominating r_q-dimensional subspace (truncated SVD) for the respective matrix unfolding

   $$B_{(q)} \in \mathbb{R}^{n\times \bar{r}_q}, \quad \bar{r}_q = r_1 \dots r_{q-1} r_{q+1} \dots r_d = O(r^{d-1}), \tag{2.11}$$

   corresponding to the q-mode contracted product

   $$B = A \times_1 {V^{(1)}}^T \times_2 \dots \times_{q-1} {V^{(q-1)}}^T \times_{q+1} {V^{(q+1)}}^T \dots \times_d {V^{(d)}}^T.$$

   Each iteration has the cost $O(dr^{d-1} n \min\{r^{d-1}, n\})$.
3. Compute the core $\boldsymbol{\beta}$ as the representation coefficients of the orthogonal projection of A onto $\mathbb{T}_{\mathbf{n}} = \otimes_{\ell=1}^d \mathbb{T}_\ell$ with $\mathbb{T}_\ell = \operatorname{span}\{v_\ell^\nu\}_{\nu=1}^{r_\ell}$ (see Remark 2.2),

   $$\boldsymbol{\beta} = A \times_1 {V^{(1)}}^T \times_2 \dots \times_d {V^{(d)}}^T \in \mathbb{B}_{\mathbf{r}},$$

   at the cost $O(r^d n)$.

In the molecular structure calculations, we normally observe fast and robust local convergence of the ALS iteration, though it is not always the case in traditional applications of the Tucker decomposition. This fact can, probably, be illuminated by the exponential error bound in the Tucker rank for the rank-**r** orthogonal approximation, which is often observed in applications to the "smooth" physically relevant data [19].

However, notice that the Tucker model applied to the general fully populated tensor of size n^d requires $O(n^{d+1})$ arithmetical operations due to the presence of complexity dominating HOSVD. Hence, normally, this algorithm applies only to small d and small n.

2.3. Two-level (primal-dual) BTA. Further reduction of numerical complexity for the Tucker model is based on the concept of the *two-level (primal-dual) Tucker approximation* [19]. The main idea of the two-level approximation consists of a rank-structured representation in certain tensor classes in the dual space, $\mathcal{S} \subset \mathbb{B}_{\mathbf{r}}$, applied to the Tucker core $\boldsymbol{\beta} \in \mathbb{B}_{\mathbf{r}}$. In particular, we consider a class $\mathcal{S} = \mathcal{C}_{R,\mathbf{r}}$ of rank-R canonical tensors in the dual space, i.e., $\boldsymbol{\beta} \in \mathcal{C}_{R,\mathbf{r}}$. The target tensor may either be represented entrywise or it may inherit certain data-sparse structure like in the Tucker or canonical models.

2.3.1. General target tensor. First, we describe the two-level Tucker model for the general input tensors. The two-level version of Algorithm G_BTA ($\mathbb{V}_{\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$) can be described as the following "model reduction" computational chain:

$$\mathbb{V}_{\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}} \subset \mathcal{C}_{R,\mathbf{n}},$$

where Level-I is understood as an application of Algorithm G_BTA ($\mathbb{V}_{\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$) and Level-II includes the rank-R canonical approximation to the small-size Tucker core $\boldsymbol{\beta} \in \mathbb{B}_{\mathbf{r}}$. Figure 2.1 illustrates the computational scheme of the two-level Tucker approximation.

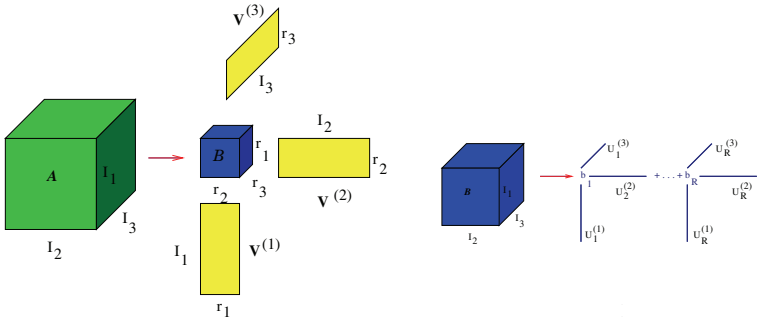

FIG. 2.1. *Level-*I*: Tucker model; Level-*II*: canonical approximation of* $\boldsymbol{\beta}$.

In the case of function related tensors, our goal is to compute the Level-I approximation with linear cost in the size of the input data (see section 2.4).

2.3.2. Target tensor in Tucker format. If the input tensor A_0 is already presented in the rank-$\mathbf{r}$ Tucker format, then one can apply the following lemma.

LEMMA 2.4 (two-level Tucker-to-canonical approximation). (a) *See* [19]. *Let the target tensor* $A \in \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$ *in* (2.7) *have the form* $A = \boldsymbol{\beta} \times_1 T^{(1)} \times_2 \cdots \times_d T^{(d)}$ *with the side-matrices* $T^{(\ell)} \in \mathbb{R}^{n \times r_\ell}$ *in the Grassman manifold and with* $\boldsymbol{\beta} \in \mathbb{R}^{r_1 \times \cdots \times r_d}$. *Then, for a given* $R < r^{d-1}$,

$$\min_{Z \in \mathcal{C}_{R,\mathbf{n}}} \|A - Z\|^2 = \min_{\boldsymbol{\mu} \in \mathcal{C}_{R,\mathbf{r}}} \|\boldsymbol{\beta} - \boldsymbol{\mu}\|^2. \tag{2.12}$$

Moreover, the optimal rank-R *approximation* $A_{(R)} \in \mathcal{C}_{R,\mathbf{n}}$ *of* A *(if existing) and the optimal rank-*R *approximation* $\boldsymbol{\beta}_{(R)} \in \mathcal{C}_{R,\mathbf{r}}$ *of* $\boldsymbol{\beta}$ *are related by*

$$A_{(R)} = \boldsymbol{\beta}_{(R)} \times_1 T^{(1)} \times_2 \cdots \times_d T^{(d)}. \tag{2.13}$$

(b) *For a given* $\mathbf{q} = (q_1, \ldots, q_d)$, *such that* $\mathbf{q} \le \mathbf{r}$ *(in the componentwise sense), and satisfying the compatibility condition* (2.10), *the BTA of* A *in the tensor class*

$\boldsymbol{\mathcal{T}}_{\mathbf{q},\mathbf{n}}$ is reduced to the BTA $\boldsymbol{\beta}_{(\mathbf{q})} = \boldsymbol{\mu}_{(\mathbf{q})} \times_1 \Sigma^{(1)} \times_2 \Sigma^{(2)} \cdots \times_d \Sigma^{(d)}$ of $\boldsymbol{\beta}$ in $\boldsymbol{\mathcal{T}}_{\mathbf{q},\mathbf{r}}$, where $\Sigma^{(\ell)} \in \mathbb{R}^{r_\ell \times q_\ell}$. The corresponding approximant can be represented by

$$A_{(\mathbf{q})} = \boldsymbol{\mu}_{(\mathbf{q})} \times_1 (T^{(1)}\Sigma^{(1)}) \times_2 \cdots \times_d (T^{(d)}\Sigma^{(d)}).$$

This representation is unique up to the rotational transform (see Remark 2.1*).*

Proof. Part (a) is proven in [19, Lemma 2.5]. To prove (b) it is enough to check the identity

$$\left(\boldsymbol{\mu} \times_1 \Sigma^{(1)} \times_2 \cdots \times_d \Sigma^{(d)}\right) \times_1 T^{(1)} \times_2 \cdots \times_d T^{(d)} = \boldsymbol{\mu} \times_1 (T^{(1)}\Sigma^{(1)}) \times_2 \cdots \times_d (T^{(d)}\Sigma^{(d)})$$

and then verify that $T^{(\ell)}\Sigma^{(\ell)} \in \mathbb{R}^{n \times q_\ell}$ $(\ell = 1, \ldots, d)$ are orthogonal matrices. The rest argument is similar to those in the proof of Lemma 2.5 in [19]. ∎

Lemma 2.4 means that the corresponding low-rank Tucker-canonical approximation of A can be reduced to the canonical approximation of a small-size core tensor. Likewise, the reduction of the Tucker rank of the target tensor $A \in \boldsymbol{\mathcal{T}}_{\mathbf{r},\mathbf{n}}$ can be reduced to the Tucker approximation of a small-size core tensor. In this way the general G_BTA algorithm can be easily adapted.

2.3.3. Rank-$\boldsymbol{R}$ canonical input. In applications related to the solution of high-dimensional PDEs the typical situation may arise when the target tensor is already presented in the rank-R canonical format, $A \in \boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$, but with relatively large R. In this case the computational cost of the two-level method can be reduced dramatically (see section 2.4.2 in [19]). The corresponding approximation scheme is presented as the following two-level chain:

$$\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R},\mathbf{r}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R'},\mathbf{r}}. \tag{2.14}$$

Here, on Level-I, we compute the best orthogonal Tucker approximation with $\boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$-type input, so that the resultant core is represented in the $\boldsymbol{\mathcal{C}}_{R,\mathbf{r}}$ format. On Level-II, the small-size Tucker core in $\boldsymbol{\mathcal{C}}_{R,\mathbf{r}}$ is approximated by an element in $\boldsymbol{\mathcal{C}}_{R',\mathbf{r}}$ with $R' < R$. In section 2.4, we will describe the computational algorithm on Level-I (which is, in fact, the most laborious part in computational scheme (2.14)) that has polynomial cost in the size of input data in $\boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$ (see Remark 2.6).

The next statement gives the characterization on the *solution structure* for the Level-I scheme in (2.14) and provides the key ingredients for constructing its efficient numerical implementation, provided that the target A is represented by (2.4). It also presents the *error estimates* for the reduced rank-**r** HOSVD-type approximation. Suppose for definiteness that $n \le R$, so that an SVD of the side-matrix $U^{(\ell)}$ is given by

$$U^{(\ell)} = Z^{(\ell)} D_\ell {V^{(\ell)}}^T = \sum_{k=1}^{n} \sigma_{\ell,k} z_\ell^k \, {v_\ell^k}^T, \quad z_\ell^k \in \mathbb{R}^n, \; v_\ell^k \in \mathbb{R}^R,$$

with the orthogonal matrices $Z^{(\ell)} = [z_\ell^1, \ldots, z_\ell^n]$, and $V^{(\ell)} = [v_\ell^1, \ldots, v_\ell^n]$, $\ell = 1, \ldots, d$. Introduce the truncated SVD of the side-matrices $U^{(\ell)}$, $Z_0^{(\ell)} D_{\ell,0} {V_0^{(\ell)}}^T$ $(\ell = 1, \ldots, d)$, where $D_{\ell,0} = \mathrm{diag}\{\sigma_{\ell,1}, \sigma_{\ell,2}, \ldots, \sigma_{\ell,r_\ell}\}$ and $Z_0^{(\ell)} \in \mathbb{R}^{n \times r_\ell}$, ${V_0}^{(\ell)} \in \mathbb{R}^{R \times r_\ell}$ represent the orthogonal factors being the respective submatrices in the SVD of $U^{(\ell)}$.

In the following we denote by $\mathcal{G}_\ell$ the Grassman manifold that is a factor space to the Stiefel manifold $\mathcal{M}_\ell$ $(\ell = 1, \ldots, d)$ in (2.9), with respect to all possible rotations.

THEOREM 2.5 (canonical-to-Tucker approximation). (a) *Let* $A \in \mathcal{C}_{R,\mathbf{n}}$ *be given by* (2.4). *Then the minimization problem*

$$A \in \mathcal{C}_{R,\mathbf{n}} \subset \mathbb{V}_{\mathbf{n}} : \quad A_{(\mathbf{r})} = \operatorname{argmin}_{T \in \mathcal{T}_{\mathbf{r},\mathbf{n}}} \|A - T\|_{\mathbb{V}_{\mathbf{n}}} \tag{2.15}$$

is equivalent to the dual maximization problem,

$$[W^{(1)}, \dots, W^{(d)}] = \operatorname{argmax}_{Y^{(\ell)} \in \mathcal{G}_\ell} \left\| \sum_{\nu=1}^{R} \xi_\nu \left({Y^{(1)}}^T u_1^\nu \right) \otimes \cdots \otimes \left({Y^{(d)}}^T u_d^\nu \right) \right\|_{\mathbb{B}_{\mathbf{r}}}^2, \tag{2.16}$$

over the Grassman manifolds $\mathcal{G}_\ell$, $Y^{(\ell)} = [y_\ell^1 \dots y_\ell^{r_\ell}] \in \mathcal{G}_\ell$ $(\ell = 1, \dots, d)$, *and where* ${Y^{(\ell)}}^T u_\ell^\nu \in \mathbb{R}^{r_\ell}$.

(b) *The compatibility condition* (2.10) *is simplified to*

$$r_\ell \le \operatorname{rank}(U^{(\ell)}) \quad \textit{with} \quad U^{(\ell)} = [u_\ell^1 \dots u_\ell^R] \in \mathbb{R}^{n \times R},$$

and this condition ensures the solvability of (2.16).

The maximizer is given by orthogonal matrices $W^{(\ell)} = [w_\ell^1 \cdots w_\ell^{r_\ell}] \in \mathbb{R}^{n \times r_\ell}$, *which can be computed by Algorithm G_BTA, where the HOSVD at step* 1 *is reduced to computation of truncated SVD of the side-matrices* $U^{(\ell)}$, $Z_0^{(\ell)} D_{\ell,0} {V_0^{(\ell)}}^T$ $(\ell = 1, \dots, d)$.

(c) *The minimizer in* (2.15) *is then calculated by the orthogonal projection*

$$A_{(\mathbf{r})} = \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{r}} \mu_{\mathbf{k}} w_1^{k_1} \otimes \cdots \otimes w_d^{k_d}, \quad \mu_{\mathbf{k}} = \langle w_1^{k_1} \otimes \cdots \otimes w_d^{k_d}, A \rangle,$$

so that the core tensor $\boldsymbol{\mu} = [\mu_{\mathbf{k}}]$ *can be represented in the rank-R canonical format*

$$\boldsymbol{\mu} = \sum_{\nu=1}^{R} \xi_\nu ({W^{(1)}}^T u_1^\nu) \otimes \cdots \otimes ({W^{(d)}}^T u_d^\nu) \in \mathcal{C}_{R,\mathbf{r}}. \tag{2.17}$$

(d) *Let* $\sigma_{\ell,1} \ge \sigma_{\ell,2} \cdots \ge \sigma_{\ell,\min(n,R)}$ *be the singular values of the* ℓ*-mode side-matrix* $U^{(\ell)} \in \mathbb{R}^{n \times R}$ $(\ell = 1, \dots, d)$. *Then the reduced HOSVD approximation* $A_{(\mathbf{r})}^0$ *at step* 1 *of Algorithm G_BTA (see item* (b)*), obtained by the projection of A onto the matrices of singular vectors* $Z_0^{(\ell)}$, *exhibits the error estimate*

$$\|A - A_{(\mathbf{r})}^0\| \le \|\boldsymbol{\xi}\| \sum_{\ell=1}^{d} \left(\sum_{k=r_\ell+1}^{\min(n,R)} \sigma_{\ell,k}^2 \right)^{1/2}, \tag{2.18}$$

where $\|\boldsymbol{\xi}\|^2 = \sum_{\nu=1}^{R} \xi_\nu^2$.

Proof. (a) The generic dual maximization problem (2.8), with A given by (2.4), now takes the form (2.16) due to the relation

$$\langle y_1^{k_1} \otimes \cdots \otimes y_d^{k_d}, A \rangle = \sum_{\nu=1}^{R} \xi_\nu \langle y_1^{k_1}, u_1^\nu \rangle \dots \langle y_d^{k_d}, u_d^\nu \rangle.$$

(b) The compatibility condition ensures the size consistency of all matrix unfoldings. Let us assume that $n \le R$ for definiteness. To justify the choice of $Z_0^{(\ell)}$, we notice that using the contracted product representation (2.5) of the canonical tensors $A \in \mathcal{C}_{R,\mathbf{n}}$,

$$A = \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d U^{(d)},$$

we have, by the construction, the following expansion for the reduced HOSVD approximation:

$$A^0_{(\mathbf{r})} = \boldsymbol{\xi} \times_1 \left[Z_0^{(1)} D_{1,0} {V_0^{(1)}}^T\right] \times_2 \left[Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T\right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T\right].$$

Now we start from the error representation

$$\begin{aligned}
A - A^0_{(\mathbf{r})} &= \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d U^{(d)} \\
&\quad - \boldsymbol{\xi} \times_1 \left[Z_0^{(1)} D_{1,0} {V_0^{(1)}}^T\right] \times_2 \left[Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T\right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T\right] \\
&\quad = \boldsymbol{\xi} \times_1 \left[U^{(1)} - Z_0^{(1)} D_{1,0} {V_0^{(1)}}^T\right] \times_2 \left[Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T\right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T\right] \\
&\quad + \boldsymbol{\xi} \times_1 U^{(1)} \times_2 \left[U^{(2)} - Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T\right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T\right] + \cdots \\
&\quad + \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d \left[U^{(d)} - Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T\right].
\end{aligned}$$

To proceed, we introduce

$$\Delta^{(\ell)} = U^{(\ell)} - Z_0^{(\ell)} D_{\ell,0} {V_0^{(\ell)}}^T, \quad W^{(\ell)} = Z_0^{(\ell)} D_{\ell,0} {V_0^{(\ell)}}^T;$$

then the ℓth summand in the right-hand side above takes the form

$$B_\ell = \boldsymbol{\xi} \times_1 U^{(1)} \cdots \times_{\ell-1} U^{(\ell-1)} \times_\ell \Delta^{(\ell)} \times_{\ell+1} W^{(\ell+1)} \cdots \times_d W^{(d)}.$$

This leads to the error bound (by the triangle inequality)

$$\begin{aligned}
\|A - A^0_{(\mathbf{r})}\| \le \sum_{\ell=1}^{d} \|B_\ell\| &= \|\boldsymbol{\xi} \times_1 \Delta^{(1)} \times_2 W^{(2)} \cdots \times_d W^{(d)}\| \\
&\quad + \|\boldsymbol{\xi} \times_1 U^{(1)} \times_2 \Delta^{(2)} \cdots \times_d W^{(d)}\| + \cdots \\
&\quad + \|\boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d \Delta^{(d)}\|.
\end{aligned}$$

Here the ℓth term B_ℓ can be represented by

$$\sum_{\nu=1}^{R} \xi_\nu u_1^\nu \cdots \times_{\ell-1} u_{\ell-1}^\nu \times_\ell \sum_{k=r_\ell+1}^{n} \sigma_{\ell,k} z_\ell^k v_{\ell,\nu}^k \times_{\ell+1} \sum_{k=1}^{r_{\ell+1}} \sigma_{\ell+1,k} z_{\ell+1}^k v_{\ell+1,\nu}^k \cdots \times_d \sum_{k=1}^{r_d} \sigma_{d,k} z_d^k v_{d,\nu}^k,$$

giving the estimate (take into account that $\|u_\ell^\nu\| = 1$, $\ell = 1, \ldots, d$, $\nu = 1, \ldots, R$)

$$\|B_\ell\| \le \sum_{\nu=1}^{R} |\xi_\nu| \left(\sum_{k=r_\ell+1}^{n} \sigma_{\ell,k}^2 {v_{\ell,\nu}^k}^2\right)^{1/2} \cdot \left(\sum_{k=1}^{r_{\ell+1}} \sigma_{\ell+1,k}^2 {v_{\ell+1,\nu}^k}^2\right)^{1/2} \cdots \left(\sum_{k=1}^{r_d} \sigma_{d,k}^2 {v_{d,\nu}^k}^2\right)^{1/2}.$$

Recall that $U^{(\ell)}$ ($\ell = 1, \ldots, d$) has normalized columns, i.e., $1 = \|u_\ell^\nu\| = \|\sum_{k=1}^{n} \sigma_{\ell,k} z_\ell^k v_{\ell,\nu}^k\|$, implying $\sum_{k=1}^{n} \sigma_{\ell,k}^2 {v_{\ell,\nu}^k}^2 = 1$ for $\ell = 1, \ldots, d$ and $\nu = 1, \ldots, R$.

Hence, we finalize the error bound as follows:

$$\begin{aligned}\|A - A^0_{(\mathbf{r})}\| &\le \sum_{\ell=1}^{d}\sum_{\nu=1}^{R}|\xi_\nu|\left(\sum_{k=r_\ell+1}^{n}\sigma^2_{\ell,k}{v^k_{\ell,\nu}}^2\right)^{1/2}\\ &\le \sum_{\ell=1}^{d}\left(\sum_{\nu=1}^{R}\xi^2_\nu\right)^{1/2}\left(\sum_{\nu=1}^{R}\sum_{k=r_\ell+1}^{n}\sigma^2_{\ell,k}{v^k_{\ell,\nu}}^2\right)^{1/2}\\ &= \sum_{\ell=1}^{d}\|\boldsymbol{\xi}\|\left(\sum_{k=r_\ell+1}^{n}\sigma^2_{\ell,k}\sum_{\nu=1}^{R}{v^k_{\ell,\nu}}^2\right)^{1/2}\\ &= \|\boldsymbol{\xi}\|\sum_{\ell=1}^{d}\left(\sum_{k=r_\ell+1}^{n}\sigma^2_{\ell,k}\right)^{1/2}.\end{aligned}$$

The case $R < n$ can be analyzed along the same lines. Now item (d) follows. ∎

We notice that the error estimate (2.18) in Theorem 2.5 actually provides the control of the reduced (simplified) HOSVD approximation error via the computable ℓ-mode error bounds since, by the construction, we have

$$\|U^{(\ell)} - Z^{(\ell)}_0 D_{\ell,0} V^{(\ell)}_0\|^2_F = \sum_{k=r_\ell+1}^{n}\sigma^2_{\ell,k}, \quad \ell = 1,\dots,d.$$

This result is similar to the well-known error estimate for the HOSVD approximation (see Property 10 in [5]).

Based on Theorem 2.5 the corresponding C_BTA algorithm for the rank-R input data reads as follows.

ALGORITHM C_BTA ($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$). Given $A \in \boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$ in the form (2.4) and the rank parameter $\mathbf{r}$.

1. For $\ell = 1,\dots,d$ compute the truncated SVD of $U^{(\ell)}$ to obtain orthogonal matrices $Z^{(\ell)}_0 \in \mathbb{R}^{n\times r_\ell}$, representing the rank-$r_\ell$ HOSVD approximation of dominating subspaces $\mathbb{T}_\ell$ (cost $O(dRn\min\{R,n\})$).
2. Perform ALS iteration as at step 2 in the general G_BTA algorithm to obtain the maximizer $W^{(\ell)}$ ($\ell = 1,\dots,d$) (cost $O(dr^{d-1}n\min\{r^{d-1},n\})$ per iteration).
3. Calculate projections of $\operatorname{Im} U^{(\ell)}$ onto the orthogonal basis of $W^{(\ell)}$ as the matrix product ${W^{(\ell)}}^T U^{(\ell)}$ ($\ell = 1,\dots,d$), at the cost $O(drRn)$.
4. Compute the rank-R core tensor $\boldsymbol{\beta} \in \boldsymbol{\mathcal{C}}_{R,\mathbf{r}}$ as in (2.17), in $O(drRn)$ operations and with $O(drR)$-storage.

Notice that step 2 in Algorithm C_BTA ($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$) above is not mandatory. It can be omitted if the initial guess $Z^{(\ell)}_0$ turns out to be "good enough" with respect to the chosen threshold criterion (see estimate (2.18)).

The following remark addresses the complexity issues.

Remark 2.6. Algorithm C_BTA ($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$) exhibits polynomial cost in R, r, n,

$$O(dRn\min\{n,R\} + dr^{d-1}n\min\{r^{d-1},n\}),$$

but with exponential scaling in d. In the absence of step 2, it does not contain iteration loops, and for any $d \ge 2$ it is a finite algorithm (now scales linearly in d) taking into

account that the matrix SVD is the finite algorithm with cubic cost in the matrix size.

Numerical tests show that Algorithm C_BTA($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$) is efficient for moderate R and n (in particular, it works well in electronic structure calculations in three dimensions for moderate grid size $n \lesssim 10^3$ and for $R \le 10^3$). However, in large scale simulations (say, in quantum chemistry) one may require computations in the range $n \lesssim 3 \cdot 10^4$, $R \le 10^4$. Hence, to get rid of polynomial scaling in R, n, r, we develop the new generation of BTA methods based on the idea of MG acceleration of nonlinear ALS iteration.

2.4. MG accelerated Tucker approximation with linear scaling. The concept of multigrid acceleration (MGA) can be applied to the multidimensional data obtained as a discretization of some smooth enough functions on a sequence of refined spatial grids. Typical application areas are the solution of integral-differential equations in $\mathbb{R}^d$, approximation of multidimensional operators and functionals, and data-structured representation of physically relevant quantities (say, molecular or electron densities, the Hartree and exchange potentials). Note that for such applications, we usually have the following estimate on the Tucker rank, $r = O(\log n) \ll n$. This concept can be applied to the fully populated as well as to the canonical rank-R target tensors. In the case of rank-R input data it can be understood as an adaptive tensor approximation method running over an incomplete set of data in the dual space. It resembles the multidimensional "adaptive cross approximation" (see, e.g., [23] and [9] related to the 3D case).

We introduce the equidistant tensor grid $\omega_{\mathbf{d},n} := \omega_1 \times \omega_2 \cdots \times \omega_d$, where $\omega_\ell := \{-X_A + (m-1)h : m = 1, \dots, n+1\}$ $(\ell = 1, \dots, d)$ with mesh-size $h = 2X_A/n$. Define a set of collocation points $\{x_{\mathbf{m}}\}$ in $\Omega \in \mathbb{R}^d$, $\mathbf{m} \in \mathcal{I} := \{1, \dots, n\}^d$, located at the midpoints of the grid-cells numbered by $\mathbf{m} \in \mathcal{I}$. For fixed n, the target tensor $A_n = [a_{n,\mathbf{m}}] \in \mathbb{R}^{\mathcal{I}}$ is defined as the trace of the given continuous multivariate function $f : \Omega \to \mathbb{R}$ on the set of collocation points $\{x_{\mathbf{m}}\}$ as follows:

$$a_{n,\mathbf{m}} = f(x_{\mathbf{m}}), \quad \mathbf{m} \in \mathcal{I}.$$

Notice that the projected Galerkin discretization method can be applied as well. For further constructions, we also need an "accurate" 1D interpolation operator $\mathcal{I}_{m-1 \to m}$ from the coarse-to-fine grids and in each spatial direction. For example, this might be the interpolation by either piecewise linear or cubic splines.

The basic idea of the *MG accelerated* best orthogonal Tucker approximation can be described by the following principles (topics 3–4 below apply to $\boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$ initial data):

1. *General MG concept.* We solve a *sequence of nonlinear approximation problems* for $A = A_n$ as in (2.7) with $n = n_m := n_0 2^m$, $m = 0, 1, \dots, M$, corresponding to a sequence of (d-adic) refined spatial grids $\omega_{\mathbf{d},n_m}$. The sequence of approximation problems is treated successively in one run from the coarse-to-fine grids (brings to mind the cascadic version of the MG method).
2. *Coarse initial approximation to the side-matrices $U^{(q)}$.* We use a *coarse initial approximation* to the q-mode dominating subspaces on finer grids, represented by the orthogonal side-matrices $U^{(q)}$ (see Definition 2.3). Specifically, initial approximation of $U^{(q)}$ on finer grid $\omega_{\mathbf{d},n_m}$ is obtained by the linear interpolation from coarser grid $\omega_{\mathbf{d},n_{m-1}}$, up to interpolation accuracy $O(n_m^{-\alpha})$, $\alpha > 0$.
3. *Most important fibers.* We employ the idea of most important fibers (MIFs) of the q-mode unfolding matrices $B_{(q)} \in \mathbb{R}^{n \times \overline{r}_q}$ (see (2.11) in step 2 of basic algorithm G_BTA, section 2.2.), whose positions are extracted from the

coarser grids. To identify the location of MIFs we apply the so-called *maximal energy principle* as follows. On the coarse grid, we calculate a projection of the q-mode unfolding matrix $B_{(q)}$ onto the true q-mode orthogonal subspace $\mathrm{Im}U^{(q)} = \mathrm{span}\{u_q^1, \ldots, u_q^{r_q}\}$, which is computed as the matrix product,

$$\boldsymbol{\beta}_{(q)} = {U^{(q)}}^T B_{(q)} \in \mathbb{R}^{r_q \times \overline{r}_q}. \tag{2.19}$$

Now the maximal energy principle specifies the location of MIFs by finding pr columns in $\boldsymbol{\beta}_{(q)}$ with maximal Euclidean norms (supposing that $pr \ll \overline{r}_q$); see Figures 2.2 and 2.3. The positions of MIFs are numbered by the index set $\mathcal{I}_{q,p}$ with $\#\mathcal{I}_{q,p} = pr$ being the subset of the larger index set,

$$\mathcal{I}_{q,p} \subset \mathcal{I}_{\overline{r}_q} := I_{r_1} \times \cdots \times I_{r_{q-1}} \times I_{r_{q+1}} \times \cdots \times I_{r_d}, \quad \#\mathcal{I}_{\overline{r}_q} = \overline{r}_q = O(r^{d-1}).$$

This strategy allows a "blind search" sampling of a fixed portion of q-mode fibers in the Tucker core which accumulate the maximum part of ℓ^2-energy. The union of selected fibers from every space dimension (specified by the index set $\mathcal{I}_{q,p}$, $q = 1, \ldots, d$) accumulates the most important information about the structure of the rank R-tensor in the dual space $\mathbb{R}^{r_1 \times \cdots \times r_d}$. This *knowledge* enormously reduces the amount of computational work on fine grids (SVD with matrix-size $n \times pr$ instead of $n \times \overline{r}_q$).

4. *Performing restricted ALS iteration.* The proposed choice of MIFs allows us to accelerate the ALS iteration to solving the problem of best rank-r approximation to the large unfolding matrix $B_{(q)} \in \mathbb{R}^{n \times \overline{r}_q}$ with dominating second dimension $\overline{r}_q = r^{d-1}$ (always the case for large d). Our approach allows us to reduce ALS iteration to the problem of computation of the r-dimensional dominating subspace of small $n \times pr$ submatrices $B_{(q,p)}$ of $B_{(q)}$ ($q = 1, \ldots, d$), where $p = O(1)$ is some fixed parameter.

The invention of the above principles leads to dramatical complexity reduction of the standard tensor algorithms G_BTA($\mathbb{V}_{\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathbb{B}_{\mathbf{r}}}$) and C_BTA($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$). In the latter case, this approach leads to the efficient tensor approximation method with linear scaling in parameters n, R, and r up to the computational cost on the "very coarse" level. In the case of fully populated tensors we arrive, at least, at the linear cost $O(n^d)$, corresponding to the storage space for the initial tensor (cf. also [23]); see also section 2.5.

In the following we focus on the case of rank-R input. The proposed algorithm of MG accelerated BTA for $A \in \boldsymbol{\mathcal{C}}_{R,\mathbf{n}}$ can be outlined as follows.

ALGORITHM MG_C_BTA ($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}_M} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$) (MG accelerated canonical-to-Tucker approximation).

1. Given $A_m \in \boldsymbol{\mathcal{C}}_{R,\mathbf{n}_m}$ in the form (2.4), corresponding to a sequence of grid parameters $n_m := n_0 2^m$, $m = 0, 1, \ldots, M$, fix a reliability threshold parameter $\varepsilon > 0$, a structural constant $p = O(1)$, the critical grid level $m_0 < M$, and the Tucker parameter $\mathbf{r}$.
2. For $m = 0$, solve C_BTA($\mathcal{C}_{R,\mathbf{n}_0} \to \mathcal{T}_{\mathcal{C}_R,\mathbf{r}}$) and compute the index set $\mathcal{I}_{q,p}(n_0) \subset \mathcal{I}_{\overline{r}_q}$ via identification of MIFs in the matrix unfolding $B_{(q)}$, $q = 1, \ldots, d$, using *the maximum energy principle* applied to the "preliminary core" $\boldsymbol{\beta}_{(q)}$ in (2.19).
3. For $m = 1, \ldots, m_0$, perform the *cascadic MG nonlinear approximation* by the ALS iteration:
   (3a) Compute initial orthogonal basis by interpolation (say, using cubic splines)
   $$\{U^{(1)}, \ldots, U^{(d)}\}_m = \mathcal{I}_{m-1 \to m}(\{U^{(1)}, \ldots, U^{(d)}\}_{m-1}).$$

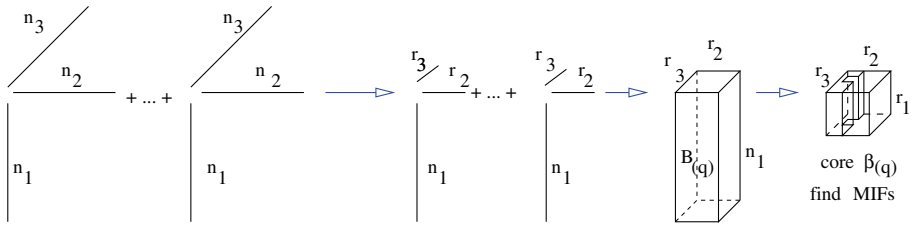

FIG. 2.2. *Illustration for $d = 3$. Finding MIFs in the "preliminary core" $\boldsymbol{\beta}_{(q)}$ for $q = 1$ for the rank-R initial data on the coarse grid $\mathbf{n} = \mathbf{n}_0$, $\mathbf{n} = (n_1, n_2, n_3)$. $B_{(q)}$ is presented in a tensor form for explanatory reasons.*

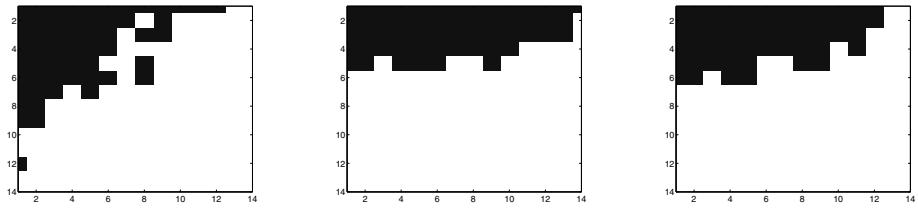

FIG. 2.3. *MIFs: selected projections of the fibers of the preliminary "cores" for computing $U^{(1)}$ (left), $U^{(2)}$ (middle), and $U^{(3)}$ (right). The example is taken from the MGA rank compression during the computation of the Hartree potential for the water molecule, $r = 14$, $p = 4$.*

For each $q = 1, \ldots, d$, and with fixed $U^{(\ell)}$ $(\ell = 1, \ldots, d,\ \ell \neq q)$, perform the following:

(3b) Define the index set $\mathcal{I}_{q,p}(n_m) = \mathcal{I}_{q,p}(n_{m-1}) \subset \mathcal{I}_{\overline{r}_q}$ and check the reliability (approximation) criteria by SVD analysis of the small-size matrix (see illustration in Figure 2.5, right)

$$B_{(q,p)} = B_{(q)|\mathcal{I}_{q,p}(n_m)} \in \mathbb{R}^{n_m \times pr}.$$

If $\sigma_{min}(B_{(q,p)}) \leq \varepsilon$, then the index set $\mathcal{I}_{q,p}$ is admissible. If for $m = m_0$ the approximation criteria above is not satisfactory, choose $p = p + 1$ and repeat steps $m = 0, \ldots, m_0$.

(3c) Determine the orthogonal matrix $U^{(q)} \in \mathbb{R}^{n \times r}$ via computing the r-dimensional dominating subspace for the *"restricted"* matrix unfolding $B_{(q,p)}$.

4. For levels $m = m_0 + 1, \ldots, M$, perform the MGA Tucker approximation by ALS iteration as in steps (3a), (3c), but with fixed positions of MIFs specified by the index set $\mathcal{I}_{q,p}(n_{m_0})$, i.e., by discarding all fibers in $B_{(q)}$ corresponding to the "less important" index set $\mathcal{I}_{\overline{r}_q} \setminus \mathcal{I}_{q,p}$.
5. Compute the rank-R core tensor $\boldsymbol{\beta} \in \boldsymbol{\mathcal{C}}_{R,\mathbf{r}}$, as in step 3 of basic algorithm C_BTA $(\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}})$.

Figure 2.4 shows the flow chart of Algorithm MG_C_BTA for the simplest case $m_0 = 1$. Numerical illustrations to the above algorithm will be presented in section 3.

The next statement proves the linear complexity of Algorithm MG_C_BTA.

THEOREM 2.7. *Algorithm MG_C_BTA($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}_M} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$) amounts to*

$$O(dRrn_M + dp^2r^2n_M)$$

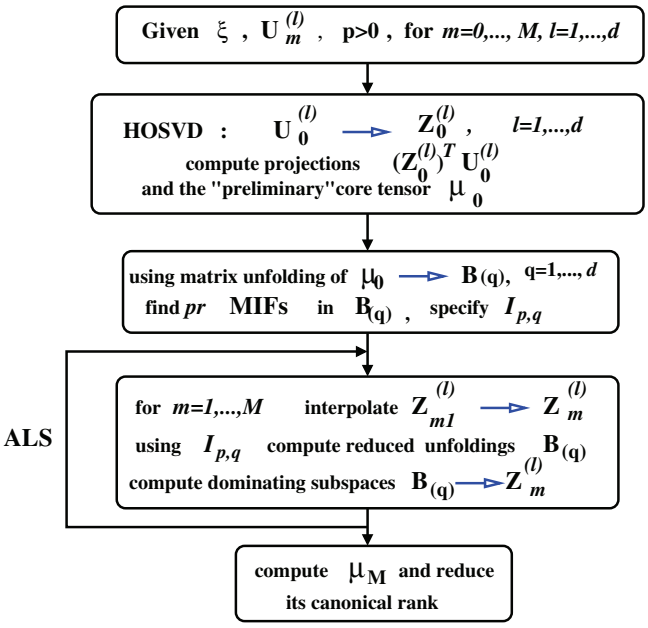

FIG. 2.4. *Flow chart of the algorithm for the MG accelerated Tucker approximation of the rank-R target, corresponding to the case* $m_0 = 1$.

operations per ALS loop, plus extra cost of the coarse mesh solver BTA ($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}_0} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$). *It requires* $O(drn_M + drR)$ *storage to represent the result.*

Proof. Step (3a) requires $O(drn_M)$ operations and memory. Notice that for large M, we have $pr \le n_M$; hence the complexity of step (3c) is bounded by $O(dRrn_M + prn_M + p^2r^2n_M)$ per iteration loop, and the same for step (3b). Rank-R representation of $\boldsymbol{\beta} \in \boldsymbol{\mathcal{C}}_{R,\mathbf{r}}$ requires $O(drRn_M)$ operations and $O(drR)$-storage. Summing up these costs over levels $m = 0, \dots, M$ proves the result. ∎

Notice that the complexity and error of the MGA Tucker approximation can be effectively controlled by the adaptive choice of the governing parameters p, m_0, and n_0.

Figure 2.5(left) demonstrates linear scaling of the MGA Tucker approximation in the input rank R and in the grid size n (electron density for the CH_4 molecule). Figure 2.5(right) plots the singular values of the mode-1 matrix unfolding $B_{(1,p)}$ with the choice $p = 4$, which demonstrates the reliability of the maximal energy principal in the error control. Similar fast decay of respective singular values is typical in most of our numerical examples in electronic structure calculations considered so far. Remarkably, the "representative subset" of fibers $\mathcal{I}_{(q,p)}$ normally has the size pr of several r's with $p \ll r$. In our applications, the usual choice of the critical grid level is $m_0 = 1$.

2.5. Remarks on general types of input data. We distinguish three particular versions of the MG accelerated algorithm, MG_BTA($\boldsymbol{\mathcal{S}}_0 \to \boldsymbol{\mathcal{T}}_{\mathbf{r}}$), adapted to different classes $\mathcal{S}_0$ of input tensors:

(F) Full (pointwise) tensor format, i.e., $\boldsymbol{\mathcal{S}}_0 = \mathbb{V}_{\mathbf{n}}$.

Initial guess on the coarse grid: (a) truncated ℓ-mode SVD (cf. [6]), (b) approximation with smaller Tucker rank, or (c) three-way cross-approximation algorithm as proposed in [23] (applies to $d = 3$).

(C) Type $\boldsymbol{\mathcal{C}}_R$ with some $R > r$ (may correspond to an analytic approximation

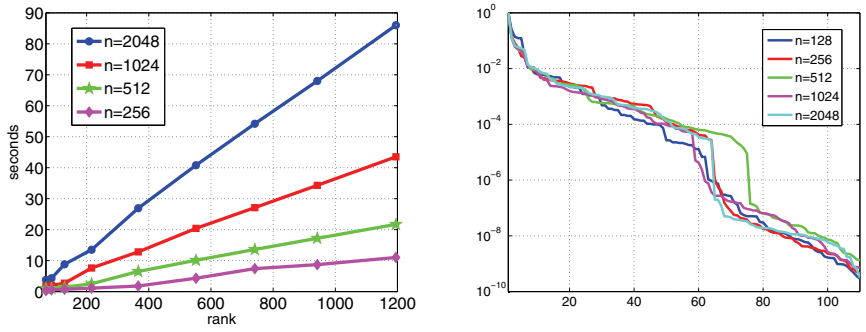

FIG. 2.5. *Linear scaling in R and in n (left). Plot of SVD for the mode-1 matrix unfolding $B_{(1,p)}$, $p = 4$ (right).*

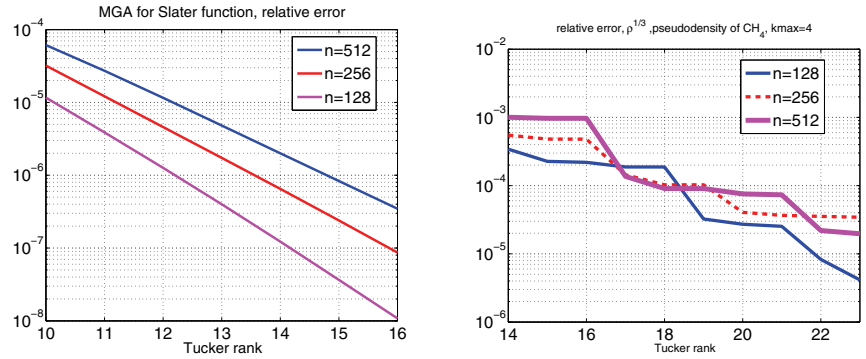

FIG. 2.6. *Approximation error of the MG Tucker tensor approximation for the 3D Slater function (left) and for the $\rho^{1/3}$ function (right).*

via *sinc*-quadratures or representation by exponential fitting as in electronic structure calculations).

Initial guess on the coarse grid: ℓ-mode QR-decomposition with truncated SVD.

(T) Type $\boldsymbol{\mathcal{T}}_{\mathbf{R}}$ (may correspond to an analytic approximation via *tensor-product interpolation*).

Initial guess on the coarse grid : truncated HOSVD of the Tucker core, or approximation with smaller Tucker rank accomplished with best rank-1 approximation to the initial increment.

In view of Lemma 2.4, the case (T) can be reduced to the case (F). The case (F) will be addressed in more detail in a forthcoming paper.

Here we give only the numerical examples of the MG accelerated Tucker approximation for fully populated tensors given by 3D Slater function, $e^{-\|x\|}$, Figure 2.6(left), and by the transcendental function of electron density ρ, $\rho^{1/3}$ (see (3.3) below), Figure 2.6(right). Note that the low-rank representation of the potential $\rho^{1/3}$ is an important issue in the density functional theory computations. For both functions we compute the Tucker approximation on the grid n^3 up to $n = 512$. As was mentioned before, the convergence upon the Tucker rank depends on physical "relevance" (smoothness) of the function. Our MG accelerated scheme requires the SVD with the complexity $O(nr^{d-1}\min\{n, r^{d-1}\})$, where $r \ll n$, instead of $O(n^{d+1})$ in the unigrid approach. In fact, the grid size of the MG accelerated Tucker decomposition for

full tensors is bounded only by the amount of available computer memory for tensor representation.

3. Numerical examination of Algorithm MG_C_BTA. As the basic example, let us consider some quantities in the Hartree–Fock equation for pairwise orthogonal electronic orbitals $\psi_i : \mathbb{R}^3 \to \mathbb{R}$, which reads as

$$\mathcal{F}_\Phi \psi_i(x) = \lambda_i \psi_i(x), \quad \int_{\mathbb{R}^3} \psi_i \psi_j = \delta_{ij},\ i,j = 1,\dots,N, \tag{3.1}$$

with $\mathcal{F}_\Phi$ being the nonlinear Fock operator

$$\mathcal{F}_\Phi := -\frac{1}{2}\Delta - V_c + V_H - \mathcal{K}.$$

Here we use the definitions

$$\tau(x,y) := \sum_{i=1}^{N} \psi_i^*(x)\psi_i(y), \quad \rho(x) := \tau(x,x),$$

$$(\mathcal{K}\psi)(x) := -\frac{1}{2}\sum_{i=1}^{N}\left(\psi\,\psi_i \star \frac{1}{\|\cdot\|}\right)\psi_i^*(x) = -\frac{1}{2}\int_{\mathbb{R}^3} \frac{\tau(x,y)}{\|x-y\|}\,\psi(y)dy$$

with the density matrix $\tau(x,y)$, the electron density $\rho(x)$, the atomic potential $V_c(x) = \sum_{\nu=1}^{M} \frac{Z_\nu}{|x-a_\nu|}$, the Hartree potential $V_H(x)$ given by

$$V_H(x) := \int_{\mathbb{R}^3} \frac{\rho(y)}{|x-y|}\,dy, \quad x \in \mathbb{R}^3, \tag{3.2}$$

and the nonlocal exchange operator $\mathcal{K}$.

The MG_CBTA algorithm has been evoked by the problem of computation of the Hartree potential V_H in the Tucker and canonical tensor formats (cf. [18]). In the cited paper we used the tensor-product convolution described in [17, 19] for calculation of the Hartree potential for the *pseudodensities* of some simple molecules. These computations provide sufficient accuracies of order 10^{-6} Hartree in the max-norm for the Hartree potential over $n \times n \times n$ grid, already with $n = 400$ to 800.

However, when computing the *all electron densities* of molecules, which contain strong cusps due to the core electron contribution (see all electron density for the water molecule in Figure 3.1), large grids of the order of several thousands are required to ensure the high resolution of local singularities. In the recent work [4], where the wavelet basis set is used for approximation of the electronic structure quantities, the Hartree potential of CH_4 and C_2H_6 molecules is resolved in the cusp area with the accuracies of the order of 10^{-3}, by computations in the volume of $[-X_A, X_A]^3$ with $X_A = 20$ atomic units (au) and with the univariate grid size $n = 5.1 \cdot 10^3$; that is, the step-size should be as small as $h = 2X_A/n \le 8 \cdot 10^{-3}$ au. In the present work we perform electronic structure calculations in tensor-product format using uniform grids with the step-size up to $h = 1.3 \cdot 10^{-3}$ au (corresponding to the univariate grid size $n = 16384$ with $X_A = 10.6$ au).

Our computational scheme for the convolution integral (3.2) involves the following steps:

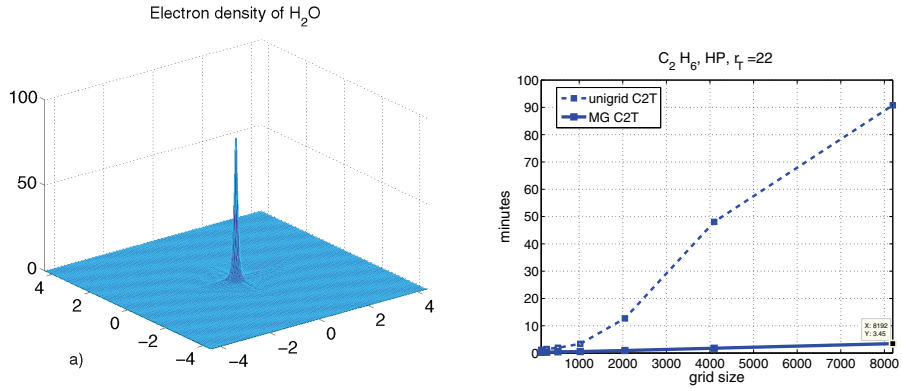

FIG. 3.1. (a) *Electron density of H_2O in the section $\Omega = [-4,4] \times [-4,4] \times \{0\}$.* (b) *MG vs. single grid CPU times for the C-T transform in computations of the Hartree potential of the C_2H_6 molecule, $r = 22$, $p = 4$.*

1. Compute the electron density of the molecule using an initial Gaussian-type representation (GTO) for the orbitals with discretization on the corresponding grid size. The resulting canonical rank of the data is $R_\rho \sim 10^3 \div 10^4$, depending on the molecule.
2. Reduce the rank of the electron density by consequent canonical-to-Tucker (C-T) and Tucker-to-canonical transform. Then the canonical rank of ρ is reduced by the order of magnitude, which is sufficient for fast discrete convolution. Note that in the single grid approach, the complexity of a direct C-T transformation depends polynomially on the main model parameters. To avoid this limitation, we apply the *MGA technique* (see section 2.4) which yields the linear complexity scaling with increasing grid size n, canonical rank R_ρ, and dimension parameter d.
3. Compute the convolution (3.2) (representing the Hartree potential) in tensor-product format using rank-R_N canonical representation of the Newton potential with $R_N \sim 19 \div 32$ and canonical representation of the electron density obtained in step 2.

Then one can compute accurately physically relevant functionals which built the Fock operator in the framework of the Hartree–Fock model, say, the Coulomb matrix.

We obtain the total electron density in a form of separable representation of orbitals in Gaussian-type basis set (GTO) given by the exponential sum

$$\rho(x) := \sum_{\nu=1}^{M} \left(\sum_{k=1}^{R_0} c_{\nu,k} (x - x_k)^{\beta_k} e^{-\lambda_k (x - x_k)^2} \right)^2, \quad x \in \mathbb{R}^3, \quad M = N/2, \tag{3.3}$$

where x_k correspond to the locations of the atoms in a molecule, M is the number of electron pairs, and R_0 is the number of GTO basis functions.

The number of GTO basis functions (Gaussians) for orbitals are given by $R_0 = 55, 41$, and 96 for CH_4, H_2O, and C_2H_6 molecules, respectively. We assume that a particular molecule is imbedded in an appropriate computational box $[-X_A, X_A]^3$. We use $X_A = 10.6$ au for H_2O and $X_A = 14.6$ au for CH_4 and C_2H_6 molecules. In the following we represent the convolving tensor corresponding to $\rho(x)$ in the canonical format. Since the products of two Gaussians in (3.3) can be written in the form of

single Gaussians by recomputing the coefficients as

$$e^{-\lambda(x-a)^2}\cdot e^{-\beta(x-b)^2}=e^{\sigma}\cdot e^{-\gamma(x-c)^2},\quad \sigma=\frac{\lambda\beta(a-b)^2}{\lambda+\beta},\ \gamma=\lambda+\beta,\ c=\frac{a\lambda+b\beta}{\lambda+\beta},$$

this leads to the following bound on the initial rank of the input tensor:

$$R=R(\rho)=\frac{R_0(R_0+1)}{2}. \tag{3.4}$$

So, we have the following ranks of the discrete canonical representation of all electron densities for the considered molecules: $R_{CH_4}=1540$, $R_{C_2H_6}=4656$, $R_{H_2O}=861$.

We apply the piecewise constant approximation to discretize the Gaussians of the type

$$g_k(y)=(y-A_k)^{a_k}e^{-\alpha_k(y-A_k)^2},\quad y\in\mathbb{R}^3,\quad k=1,\dots,R, \tag{3.5}$$

by rank-1 $n\times n\times n$ tensors by computing the sampling values of the corresponding Gaussians at the centers of intervals corresponding to the equidistant tensor grid $\omega_{\mathbf{3},n}:=\omega_1\times\omega_2\times\omega_3$, $\omega_\ell:=\{-X_A+(m-1)h : m=1,\dots,n+1\}$ $(\ell=1,\dots,3)$ with mesh-size $h=2X_A/n$. We define the sampling points as $\{y_{\mathbf{i}}=(y^1_{i_1},y^2_{i_2},y^3_{i_3})\}$, $i_\ell\in I_\ell=\{1,\dots,n\}$ for $\ell=1,2,3$. Here we have $a_k=(a^1_k,a^2_k,a^3_k)$, $A_k=(A^1_k,A^2_k,A^3_k)$. The rank-1 tensor representing the single Gaussian g_k is given by the canonical rank-1 tensor

$$G_k\equiv[t_{\mathbf{i}}]_{\mathbf{i}\in\mathcal{I}}=t_1\otimes t_2\otimes t_3\in\mathbb{V}_{\mathbf{n}}\quad\text{with entries}\quad t_{\mathbf{i}}=t_{1,i_1}\cdot t_{2,i_2}\cdot t_{3,i_3},$$

where

$$t_\ell=\{t_{\ell,i}\}_{i\in I_\ell}\in\mathbb{V}_\ell,\quad t_{\ell,i}=(y^\ell_i-A^\ell_k)^{a^\ell_k}e^{-\alpha_k(y^\ell_i-A^\ell_k)^2},\quad \ell=1,2,3.$$

We apply the standard collocation scheme to discretize the convolution product on tensor grid $\omega_{\mathbf{3},n}$ of collocation points $\{x_{\mathbf{m}}\}$ in $\omega_{\mathbf{3},n}$, $\mathbf{m}\in\mathcal{M}:=\{1,\dots,n+1\}^3$ [17]. For given piecewise constant basis functions $\{\phi_{\mathbf{i}}\}$, $\mathbf{i}\in\mathcal{I}:=\{1,\dots,n\}^3$, associated with $\omega_{\mathbf{3},n}$, let $f_{\mathbf{i}}=f(y_{\mathbf{i}})$ be the representation coefficients of f in $\{\phi_{\mathbf{i}}\}$,

$$f(y)\approx\sum_{\mathbf{i}\in\mathcal{I}}f_{\mathbf{i}}\phi_{\mathbf{i}}(y),$$

where $y_{\mathbf{i}}$ is the midpoint of the grid-cell numbered by $\mathbf{i}\in\mathcal{I}$. Now the *collocation scheme* reads as

$$f*g\approx\{W_{\mathbf{m}}\}_{\mathbf{m}\in\mathcal{M}},\quad W_{\mathbf{m}}:=\sum_{\mathbf{i}\in\mathcal{I}}f_{\mathbf{i}}\int_{\mathbb{R}^3}\phi_{\mathbf{i}}(y)g(x_{\mathbf{m}}-y)dy,\quad x_{\mathbf{m}}\in\omega_{\mathbf{3},n}.$$

As a first step, we precompute the coefficients

$$g_{\mathbf{i}}=\int_{\mathbb{R}^3}\frac{\phi_{\mathbf{i}}(y)}{|y|}dy,\quad \mathbf{i}\in\mathcal{I}.$$

The coefficient tensor $G=[g_{\mathbf{i}}]\in\mathbb{R}^{\mathcal{I}}$ for the Coulomb potential $\frac{1}{|x-y|}$ is approximated in the rank-R_N canonical tensor format (see section 2) using optimized *sinc*-quadratures [12, 13, 17], where the rank parameter $R_N=O(|\log\varepsilon|\log n)$ depends

logarithmically on both the required accuracy and the problem-size n. In all computations presented below we choose the tensor rank in the range $R_N \in [20, 30]$ to provide the desired accuracy. The third-order tensor $F = [f_{\mathbf{i}}] \in \mathbb{R}^{\mathcal{I}}$ is approximated either in the rank $\mathbf{r} = (r, r, r)$ Tucker format or via the canonical model with the tensor rank R.

Following [17, 19], we compute $W_{\mathbf{m}}$ by copying the corresponding portion of the *discrete convolution* in $\mathbb{R}^3$,

$$F * G := \{z_{\mathbf{j}}\}, \quad z_{\mathbf{j}} := \sum_{\mathbf{i}\in\mathcal{I}} f_{\mathbf{i}} g_{\mathbf{j}-\mathbf{i}+\mathbf{1}}, \quad \mathbf{j} \in \mathcal{J} := \{1, \dots, 2n-1\}^3, \tag{3.6}$$

centered at $\mathbf{j} = \mathbf{n}$, where the sum is over all $\mathbf{i}, \mathbf{j} \in \mathcal{I}$, which lead to legal subscripts for $g_{\mathbf{j}-\mathbf{i}+\mathbf{1}}$, i.e., $\mathbf{j} - \mathbf{i} + \mathbf{1} \in \mathcal{I}$.

Approximating F in the rank-R canonical format (see (2.4)) enables us to compute $F * G$ in the form

$$F * G = \sum_{k=1}^{R_N} \sum_{m=1}^{R} c_k b_m \left(u_1^k * v_1^m\right) \otimes \left(u_2^k * v_2^m\right) \otimes \left(u_3^k * v_3^m\right),$$

which leads to the cost

$$\mathcal{N}_{C*C} = O(RR_N n \log n).$$

Thus, computation of the convolution product with such ranks of the input tensors has practical limitations since the exact rank of the resulting tensor is a product of those for the corresponding convolving arrays. For the purpose of reducing the tensor rank of the convolving density, we apply the fast algorithm MG_C_BTA($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}_M} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$) for the C-T MG transform defined on the sequence of refined grids, followed by the Tucker-to-canonical transform of the core tensor. In this way, the canonical rank R can be reduced by the order of magnitude, from several thousands to a few hundreds or even tens, depending on the input data. As was mentioned, in general, the complexity of the single grid C-T transform depends polynomially on the parameters of the model; see Algorithm C_BTA ($\boldsymbol{\mathcal{C}}_{R,\mathbf{n}} \to \boldsymbol{\mathcal{T}}_{\mathcal{C}_R,\mathbf{r}}$). The *MGA technique* implemented by Algorithm MG_CBTA provides calculations of the 3D integral transform (3.2) using record grid sizes up to $n = 16384$ in small computational times of both the MG preprocessing and discrete 3D convolution; see Figure 3.2(b). Figure 3.1 shows a comparison of the C-T preprocessing times for BTA and MG_BTA algorithms for computations with the same accuracy.

The table below shows the advantage of the fast tensor-product convolution method with the rank recompression via Algorithm MG_C_BTA, compared with those based on 3D FFT. We present the CPU times for high accuracy computation of the Hartree potential for the H_2O molecule calculated in MATLAB on a Sun Fire X4600 computer with 2.6 GHz processor. The CPU time for the FFT-based scheme with $n \geq 1024$ is obtained by extrapolation.

n^3	128^3	256^3	512^3	1024^3	2048^3	4096^3	8192^3	16384^3
3D FFT (sec)	4.3	55.4	582.8	~ 6000	–	–	–	~ 2 years
$Conv_{CC}$ (sec)	1.0	3.1	5.3	21.9	43.7	127.1	368.6	700.2

Figures 3.2(a), 3.3(a), and 3.4(a) show the accuracy of the computation of the Hartree potential of H_2O, CH_4, and C_2H_6 molecules on logarithmic scale. We present the absolute error for V_H in the subinterval along the x-axis compared with the corresponding

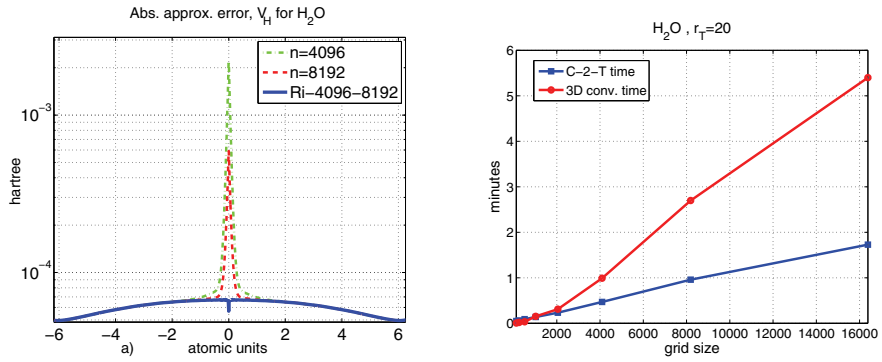

FIG. 3.2. (a) *Absolute approximation error of the tensor-product computation of the Hartree potential of the water molecule in the subinterval* $\Omega = [-6, 6] \times \{0\} \times \{0\}$ *and* (b) *the corresponding CPU times.*

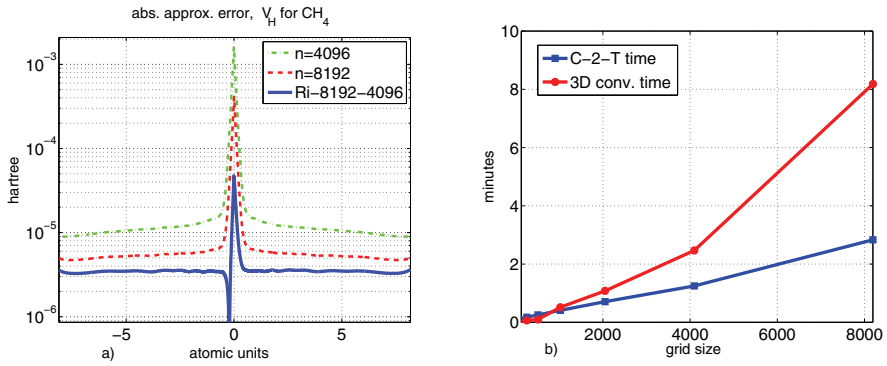

FIG. 3.3. (a) *Absolute approximation error of the tensor-product computation of the Hartree potential of the* CH_4 *molecule in the subinterval* $\Omega = [-8, 8] \times \{0\} \times \{0\}$ *and* (b) *the corresponding CPU times.*

values of V_H computed by the MOLPRO package. We compare the computational error for the grid sizes $n = 4096, 8192$ and for the corresponding Richardson extrapolation of order $O(h^3)$; see Theorems 2.2 and 2.3 from [17]. We observe the accuracy $8 \cdot 10^{-5}$ Hartree at the cusp region corresponding to carbon atoms in CH_4, $5 \cdot 10^{-5}$ Hartree at the cusp corresponding to oxygen atoms in H_2O, and $5 \cdot 10^{-6}$ Hartree at the cusps corresponding to carbon atoms in C_2H_6. Note that maximum values of the Hartree potential at the cusp points are $V_H(0, 0, 0) = 8.35$, 11.73, and 9.49 Hartree for the methane, water, and ethane molecules, respectively. This yields the relative accuracy of the order 10^{-6} for the considered molecules.

Figures 3.2(b), 3.3(b), and 3.4(b) show CPU times for the CH_4, H_2O, and C_2H_6 molecules, respectively. "C2T" lines correspond to the preprocessing time, which consists mostly of the MG rank compression, while the "3D conv." lines show the respective convolution times. The total computational time for V_H at a fixed grid size $n = n_f$ consists of a sum of preprocessing times from all previous levels, starting from the initial grid with, say, $n_0 = 64$, up to n_f, plus the convolution time for the level with $n = n_f$.

Computational time depends mainly on the chosen accuracy of approximation, which is defined by the MG parameters, namely, by the structural constant p and

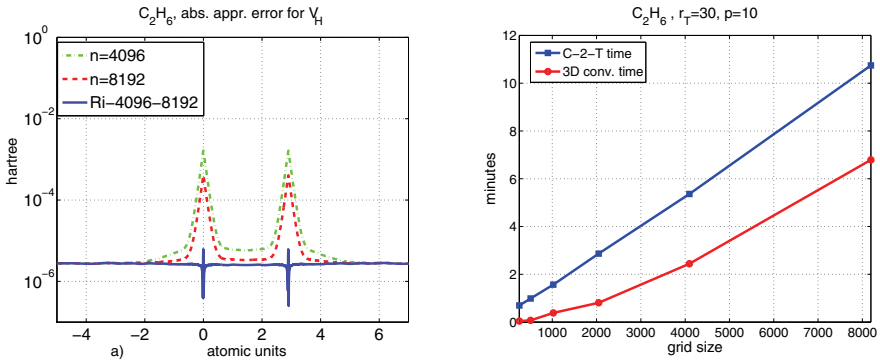

FIG. 3.4. (a) *Absolute approximation error of the tensor-product computation of the Hartree potential of the C_2H_6 molecule in the subinterval $\Omega = [-5, 7] \times \{0\} \times \{0\}$ and* (b) *the corresponding CPU times.*

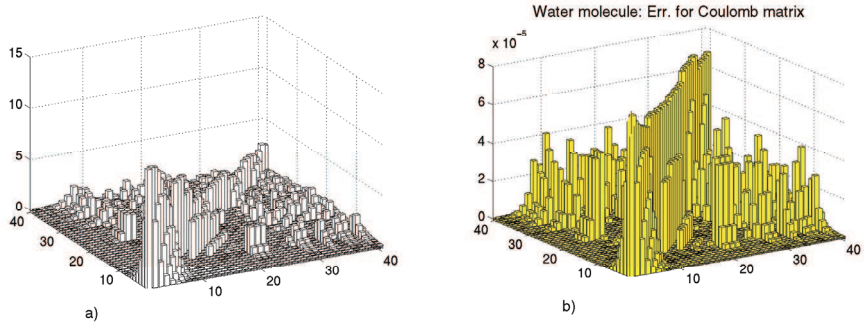

FIG. 3.5. (a) *Coulomb matrix (absolute values) for H_2O and* (b) *the absolute approximation error of the MG tensor-product computation of the Coulomb matrix for H_2O.*

the respective Tucker rank; see Theorem 2.7. It should be noted that the CPU times can be crucially reduced if the accuracy of the result is not so demanding, say, up to 10^{-3}. In applications discussed above the computational task was challenging due to accuracy requirements, which led to the increase of the Tucker rank and of the structural constant p in the MG scheme, and consequently the demands on computational resources.

Next, we computed the Coulomb matrix for the Hartree potential of the considered molecules in the GTO basis set, using the tensor-product scheme [19] as discussed in [18]. The Coulomb matrix for V_H with respect to the set of normalized Gaussians $\{\widetilde{g}_k\}$ is given by

$$J_{km} := \int_{\mathbb{R}^3} \widetilde{g}_k(x)\widetilde{g}_m(x)V_H(x)dx, \quad k, m = 1, \ldots, R_0, \quad x \in \mathbb{R}^3. \tag{3.7}$$

Figure 3.5 shows the absolute approximation error of the Coulomb matrix J_{km} for the water molecule. Figures 3.6 and 3.7 show absolute errors of diagonal elements of the Coulomb matrices for CH_4 and C_2H_6 compared with the MOLPRO computations. It should be noticed that the absolute approximation errors for the diagonal matrix elements corresponding to cusp areas are below $4 \cdot 10^{-6}$, $3 \cdot 10^{-5}$, and $8 \cdot 10^{-5}$ for the CH_4, C_2H_6, and water molecules, respectively.

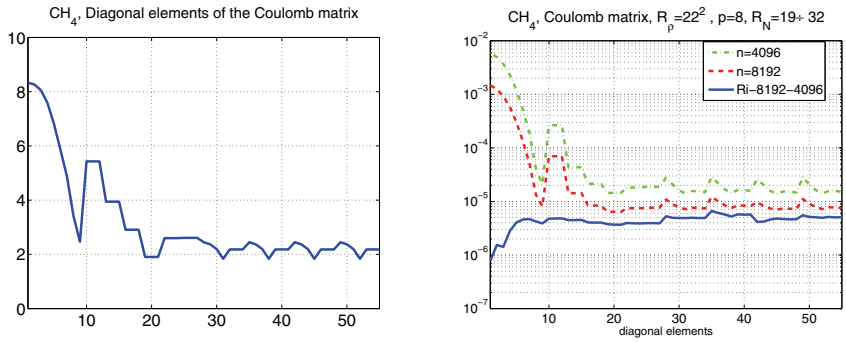

FIG. 3.6. *Diagonal elements of the Coulomb matrix of the methane molecule (left) and absolute approximation error (right).*

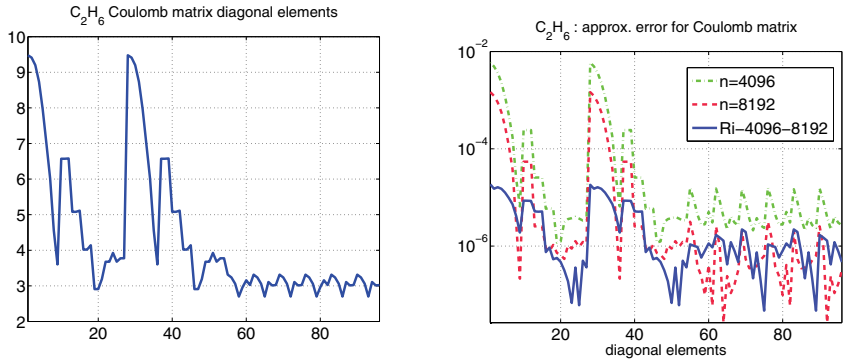

FIG. 3.7. *Diagonal elements of the Coulomb matrix of the ethane molecule (left) and absolute approximation error (right).*

Acknowledgments. The authors gratefully acknowledge Prof. Dr. W. Hackbusch and Prof. Dr. R. Schneider for fruitful discussions and useful comments. We would like to thank Dr. H.-J. Flad and Dipl.-Phys. R. Chinnamsetty for collaboration and providing the MOLPRO data. We appreciate various comments of the anonymous referees and the managing editor, which have led to substantial improvement of the presentation.

REFERENCES

[1] G. BEYLKIN AND M. J. MOHLENKAMP, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251.

[2] G. BEYLKIN AND M. J. MOHLENKAMP, *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.

[3] J. D. CARROL, S. PRUZANSKY, AND J. B. KRUSKAL, *CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters*, Psychometrika, 45 (1980), pp. 3–24.

[4] S. R. CHINNAMSETTY, M. ESPIG, W. HACKBUSCH, B. N. KHOROMSKIJ, AND H.-J. FLAD, *Kronecker tensor product approximation in quantum chemistry*, J. Chem. Phys., 127 (2007), 084110.

[5] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.

[7] H.-J. Flad, W. Hackbusch, B. N. Khoromskij, and R. Schneider, *Concept of data-sparse tensor-product approximation in many-particle modeling*, in Matrix Methods: Theory, Algorithms, and Applications, V. Olshevsky and E. Tyrtyshnikov, eds., World Scientific, Singapore, 2008, pp. 313–347.

[8] H.-J. Flad, W. Hackbusch, and R. Schneider, *Best N-term approximation in electronic structure calculations*: I. *One-electron reduced density matrix*, M2AN Math. Model. Numer. Anal., 40 (2006), pp. 49–61.

[9] S. Friedland and V. Mehrmann, *Best Subspace Tensor Approximations*, http:/arxiv.org/abs/0805.4220v1 (2008).

[10] M. Griebel and J. Hamaekers, *Sparse grids for the Schrödinger equation*, M2AN Math. Model. Numer. Anal., 41 (2007), pp. 215–247.

[11] W. Hackbusch, *Fast and exact projected convolution for non-equidistant grids*, Computing, 80 (2007), pp. 137–168.

[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), pp. 177–202.

[13] W. Hackbusch and B. N. Khoromskij, *Tensor-product approximation to operators and functions in high dimensions*, J. Complexity, 23 (2007), pp. 697–714.

[14] W. Hackbusch, B. N. Khoromskij, and E. Tyrtyshnikov, *Approximate iteration for structured matrices*, Numer. Math., 109 (2008), pp. 365–383.

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

[16] B. N. Khoromskij, *Structured rank-$(r_1, \ldots, r_d)$ decomposition of function-related tensors in $\mathbb{R}^d$*, Comput. Methods Appl. Math., 6 (2006), pp. 194–220.

[17] B. N. Khoromskij, *Fast and Accurate Tensor Approximation of Multivariate Convolution with Linear Scaling in Dimension*, Preprint MPI MIS 36, MPI MIS, Leipzig, Germany, 2008; J. Comput. Appl. Math., to appear.

[18] B. N. Khoromskij, V. Khoromskaia, S. R. Chinnamsetty, and H.-J. Flad, *Tensor decomposition in electronic structure calculations on 3D Cartesian grids*, J. Comput. Phys., 228 (2009), pp. 5749–5762.

[19] B. N. Khoromskij and V. Khoromskaia, *Low rank Tucker tensor approximation to the classical potentials*, Cent. Eur. J. Math., 5 (2007), pp. 1–28.

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

[21] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., to appear.

[22] C. Le Bris, ed., *Handbook of Numerical Analysis, Vol.* X, *Computational Chemistry*, North–Holland, Amsterdam, 2003.

[23] I. V. Oseledets, D. V. Savostianov, and E. E. Tyrtyshnikov, *Tucker dimensionality reduction of three-dimensional arrays in linear time*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 939–956.

[24] C. Schwab and R. Todor, *Sparse finite elements for elliptic problems with stochastic loading*, Numer. Math., 95 (2003), pp. 707–713.

[25] A. Smilde, R. Bro, and P. Geladi, *Multiway Analysis*, Wiley, New York, 2004.

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

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

[28] H. Yserentant, *The hyperbolic cross space approximation of electronic wavefunctions*, Numer. Math., 105 (2007), pp. 659–690.