



# Ubiquitous Nature of the Reduced Higher Order SVD in Tensor-Based Scientific Computing

Venera Khoromskaia\* and Boris N. Khoromskij

Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany

## OPEN ACCESS

### Edited by:

Edoardo Angelo Di Napoli,  
Helmholtz Association of German  
Research Centres (HZ), Germany

### Reviewed by:

Maxim Rakhuba,  
National Research University Higher  
School of Economics, Russia  
Katharina Kormann,  
Uppsala University, Sweden  
Akum Onwunta,  
Lehigh University, United States

### \*Correspondence:

Venera Khoromskaia  
vekh@mis.mpg.de

### Specialty section:

This article was submitted to  
Mathematics of Computation and  
Data Science,  
a section of the journal  
Frontiers in Applied Mathematics and  
Statistics

Received: 01 December 2021

Accepted: 03 March 2022

Published: 19 April 2022

### Citation:

Khoromskaia V and Khoromskij BN  
(2022) Ubiquitous Nature of the  
Reduced Higher Order SVD in  
Tensor-Based Scientific Computing.  
Front. Appl. Math. Stat. 8:826988.  
doi: 10.3389/fams.2022.826988

Tensor numerical methods, based on the rank-structured tensor representation of  $d$ -variate functions and operators discretized on large  $n^{\otimes d}$  grids, are designed to provide  $O(dn)$  complexity of numerical calculations contrary to  $O(n^d)$  scaling by conventional grid-based methods. However, multiple tensor operations may lead to enormous increase in the tensor ranks (curse of ranks) of the target data, making calculation intractable. Therefore, one of the most important steps in tensor calculations is the robust and efficient rank reduction procedure which should be performed many times in the course of various tensor transforms in multi-dimensional operator and function calculus. The rank reduction scheme based on the Reduced Higher Order SVD (RHOSVD) introduced by the authors, played a significant role in the development of tensor numerical methods. Here, we briefly survey the essentials of RHOSVD method and then focus on some new theoretical and computational aspects of the RHOSVD and demonstrate that this rank reduction technique constitutes the basic ingredient in tensor computations for real-life problems. In particular, the stability analysis of RHOSVD is presented. We introduce the multi-linear algebra of tensors represented in the range-separated (RS) tensor format. This allows to apply the RHOSVD rank-reduction techniques to non-regular functional data with many singularities, for example, to the rank-structured computation of the collective multi-particle interaction potentials in bio-molecular modeling, as well as to complicated composite radial functions. The new theoretical and numerical results on application of the RHOSVD in scattered data modeling are presented. We underline that RHOSVD proved to be the efficient rank reduction technique in numerous applications ranging from numerical treatment of multi-particle systems in material sciences up to a numerical solution of PDE constrained control problems in  $\mathbb{R}^d$ .

**Keywords:** low-rank tensor product approximation, multi-variate functions, tensor calculus, rank reduction, tucker format, canonical tensors, interaction potentials, scattered data modeling

## 1. INTRODUCTION

The mathematical models in large-scale scientific computing are often described by steady state or dynamical PDEs. The underlying physical, chemical or biological systems usually live in 3D physical space  $\mathbb{R}^3$  and may depend on many structural parameters. The solution of arising discrete systems of equations and optimization of the model parameters lead to the challenging numerical

problems. Indeed, the accurate grid-based approximation of operators and functions involved requires large spatial grids in  $\mathbb{R}^d$ , resulting in considerable storage space and implementation of various algebraic operations on huge vectors and matrices. For further discussion we shall assume that all functional entities are discretized on  $n^{\otimes d}$  spatial grids where the univariate grid size  $n$  may vary in the range of several thousands. The linear algebra on  $N$ -vectors and  $N \times N$  matrices with  $N = n^d$  quickly becomes non-tractable as  $n$  and  $d$  increase.

Tensor numerical methods [1, 2] provide means to overcome the problem of the exponential increase of numerical complexity in the dimension of the problem  $d$ , due to their intrinsic feature of reducing the computational costs of multi-linear algebra on rank-structured data to merely linear scaling in both the grid-size  $n$  and dimension  $d$ . They appeared as bridging of the algebraic tensor decompositions initiated in chemometrics [3–10] and of the nonlinear approximation theory on separable low-rank representation of multi-variate functions and operators [11–13]. The canonical [14, 15], Tucker [16], tensor train (TT) [17, 18], and hierarchical Tucker (HT) [19] formats are the most commonly used rank-structured parametrizations in applications of modern tensor numerical methods. Further data-compression to the logarithmic scale can be achieved by using the quantized-TT (QTT) [20, 21] tensor approximation. At present there is an active research toward further progress of tensor numerical methods in scientific computing [1, 2, 22–26]. In particular, there are considerable achievements of tensor-based approaches in computational chemistry [27–31], in bio-molecular modeling [32–35], in optimal control problems (including the case of fractional control) [36–39], and in many other fields [6, 40–44].

Here, we notice that tensor numerical methods proved to be efficient when all input data and all intermediate quantities within the chosen computational scheme are presented in a certain low-rank tensor format with controllable rank parameters, i.e., on low-rank tensor manifolds. In turn, tensor decomposition of the full format data arrays is considered as an N-P hard problem. For example, the truncated HOSVD [7] of an  $n^{\otimes d}$ -tensor in the Tucker format amounts to  $O(n^{d+1})$  arithmetic operations while the respective cost of the TT and HT higher-order SVD [18, 45] is estimated by  $O(n^{\frac{3}{2}d})$ , indicating that rank decomposition of full format tensors still suffers from the “curse of dimensionality” and practically could not be applied in large scale computations.

On the other hand, often, the initial data for complicated numerical algorithms may be chosen in the canonical/Tucker tensor formats, say as a result of discretization of a short sum of Gaussians or multi-variate polynomials, or as a result of the analytical approximation by using Laplace transform representation and sinc-quadratures [1]. However, the ranks of tensors are multiplied in the course of various tensor operations, leading to dramatic increase in the rank parameter (“curse of ranks”) of a resulting tensor, thus making tensor-structured calculation intractable. Therefore, fast and stable rank reduction schemes are the main prerequisite for the success of rank-structured tensor techniques.

Invention of the Reduced Higher Order SVD (RHOSVD) in [46] and the corresponding rank reduction procedure based on the canonical-to-Tucker transform and subsequent canonical approximation of the small Tucker core (Tucker-to-canonical transform) was a decisive step in development of the tensor numerical methods in scientific computing. In contrast to the conventional HOSVD, the RHOSVD does not need a construction of the full size tensor for finding the orthogonal subspaces of the Tucker tensor representation. Instead, RHOSVD applies to numerical data in the canonical tensor format (with possibly large initial rank  $R$ ) and exhibits the  $O(dnR \min\{n, R\})$  complexity, uniformly in the dimensionality of the problem,  $d$ , and it was an essential step ahead in evolution of the tensor-structured numerical techniques.

In particular, this rank reduction scheme was applied to calculation of 3D and 6D convolution integrals in tensor-based solution of the Hartree-Fock equation [27, 46]. Combined with the Tucker-to-canonical transform, this algorithm provides a stable procedure for the rank reduction of possibly huge ranks in tensor-structured calculations of the Hartree potential. The RHOSVD based rank reduction scheme for the canonical tensors is specifically useful for 3D problems, which are most often in real-life applications. However, the RHOSVD-type procedure can be also efficiently applied in the construction of the TT tensor format from the canonical tensor input, which often appears in tensor calculations<sup>1</sup>.

The RHOSVD is the basic tool for the construction of the range-separated (RS) tensor format introduced in [32] for the low-rank tensor representation of the bio-molecular long-range electrostatic potentials. Recent example on the RS representation of the multi-centered Dirac delta function [34] paves the way for efficient solution decomposition scheme introduced for the Poisson-Boltzmann equation [33, 35].

In some applications the data could be presented as a sum of highly localized and rank-structured components so that their further numerical treatment again requires the rank reduction procedure (see Section 4.5 concerning the long-range potential calculation for many-particle system). Here, we present the constructive description of multi-linear operations on tensors in RS format which allow to compute the short- and long-range parts of resulting combined tensors. In particular, this applies to commonly used addition of tensors, Hadamard and contracted products as well as to composite functions of RS tensors. We then introduce tensor-based modeling of the scattered data by a sum of Slater kernels and show the existence of the low-rank representation for such data in the RS tensor format. The numerical examples demonstrate the practical efficiency of such kind of tensor interpolation. This approach may be efficiently used in many applications in data science and in stochastic data modeling.

Rank reduction procedure by using the RHOSVD is a mandatory part in solving the three-dimensional elliptic and pseudo-differential equations in the rank-structured tensor format. In the course of preconditioned iterations, the tensor

<sup>1</sup>Otherwise one can not avoid the “curse of dimensionality”, see the cost of the HT/TT SVD above.

ranks of the governing operator, the preconditioner and of the current iterand are multiplied at each iterative step, and, therefore, a fast and robust rank reduction techniques is the prerequisite for such methodology applied in the framework of iterative elliptic problem solvers. In particular, this approach was applied to the PDE constrained (including the case of fractional operators) optimal control problems [36, 39]. As result, the computational complexity can be reduced to almost linear scale,  $O(nR)$ , contrary to conventional  $O(n^3)$  complexity, as demonstrated by numerics in [36, 39].

Tensor-based algorithms and methods are now being widely used and developed further in the communities of scientific computing and data science. Tensor techniques evolve in traditional tensor decompositions in data processing [5, 42, 47], and they are actively promoted for tensor-based solution of the multi-dimensional problems in numerical analysis and quantum chemistry [1, 24, 29, 38, 39, 48, 49]. Notice that in the case of higher dimensions the rank reduction in the canonical format can be performed directly (i.e., without intermediate use of the Tucker approximation) by using the cascading ALS iteration in the CP format (see [50] concerning the tensor-structured solution of the stochastic/parametric PDEs).

The rest of this article is organized as follows. In Section 2, we sketch some results on the construction of the RHOSVD and present some old and new results on the stability of error bounds. In Section 2.2, we recollect the mixed canonical-Tucker tensor format and the Tucker-to-canonical transform. Section 3 recalls the results from Khoromskaia [27] on calculation of the multi-dimensional convolution integrals with the Newton kernel arising in computational quantum chemistry. Section 4 addresses the application of RHOSVD to RS parametrized tensors. In Section 4.2, we discuss the application of RHOSVD in multi-linear operations of data in the RS tensor format. The scattered data modeling is considered in section 4.5 from both theoretical and computational aspects. Application of RHOSVD for tensor-based representation of Greens kernels is discussed in Section 5. Section 6 gives a short sketch of RHOSVD in application to tensor-structured elliptic problem solvers.

## 2. REDUCED HOSVD AND CP-TO-TUCKER TRANSFORM

### 2.1. Reduced HOSVD: Error Bounds

In computational schemes including bilinear tensor-tensor or matrix-tensor operations the increase of tensor ranks leads to the critical loss of efficiency. Moreover, in many applications, for example in electronic structure calculations, the canonical tensors with large rank parameters arise as the result of polynomial type or convolution transforms of some function related tensors (say, electron density, the Hartree potential, etc.) In what follows, we present the new look on the direct method of rank reduction for the canonical tensors with large initial rank, the reduced HOSVD, first introduced and analyzed in [46].

In what follows, we consider the vector space of  $d$ -fold real-valued data arrays  $\mathbb{R}^{n_1 \times \dots \times n_d}$  endorsed by the Euclidean scalar product  $\langle \cdot, \cdot \rangle$  with the related norm  $\|u\| = \langle u, u \rangle^{1/2}$ . We denote

by  $\mathcal{T}_{\mathbf{r}, \mathbf{n}}$  the class of tensors  $\mathbf{A} \in \mathbb{R}^{n \otimes d}$  parametrized in the rank- $\mathbf{r}$ ,  $\mathbf{r} = (r_1, \dots, r_d)$  orthogonal Tucker format,

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

with the orthogonal side-matrices  $V^{(\ell)} = [\mathbf{v}_1^{(\ell)} \dots \mathbf{v}_{r_\ell}^{(\ell)}] \in \mathbb{R}^{n \times r_\ell}$  and with the core coefficient tensor  $\boldsymbol{\beta} \in \mathbb{R}^{r_1 \times \dots \times r_d}$ . Here and thereafter  $\times_\ell$  denotes the contracted tensor-matrix product in the dimension  $\ell$ , and  $\mathbb{R}^{n \otimes d}$  denotes the Euclidean vector space of  $n_1 \times \dots \times n_d$ -tensors with equal mode size  $n_\ell = n$ ,  $\ell = 1, \dots, d$ .

Likewise,  $\mathcal{C}_{R, \mathbf{n}}$  denotes the class of rank- $R$  canonical tensors. For given  $\mathbf{A} \in \mathcal{C}_{R, \mathbf{n}}$  in the rank- $R$  canonical format,

$$\mathbf{A} = \sum_{v=1}^R \xi_v \mathbf{u}_v^{(1)} \otimes \dots \otimes \mathbf{u}_v^{(d)}, \quad \xi_v \in \mathbb{R}, \quad (1)$$

with normalized canonical vectors, i.e.,  $\|\mathbf{u}_v^{(\ell)}\| = 1$  for  $\ell = 1, \dots, d$ ,  $v = 1, \dots, R$ .

The standard algorithm for the Tucker tensor decomposition [7] is based on HOSVD applied to full tensors of size  $n^d$  which exhibits  $O(n^{d+1})$  computational complexity. The question is how to simplify the HOSVD Tucker approximation in the case of canonical input tensor in the form Equation (1) without use of the full format representation of  $\mathbf{A}$ , and in the situation when the CP rank parameter  $R$  and the mode sizes  $n$  of the input can be sufficiently large.

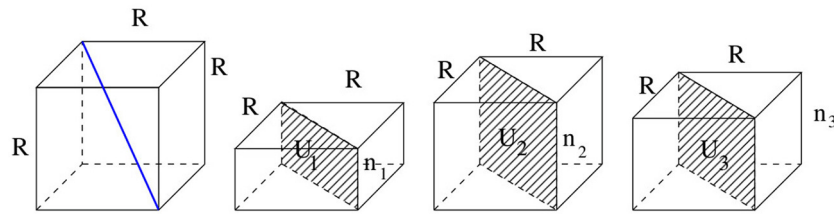
First, let us use the equivalent (nonorthogonal) rank- $\mathbf{r} = (R, \dots, R)$  Tucker representation of the tensor Equation (1),

$$\mathbf{A} = \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_d U^{(d)}, \quad \boldsymbol{\xi} = \text{diag}\{\xi_1, \dots, \xi_R\}, \quad (2)$$

via contraction of the diagonal tensor  $\boldsymbol{\xi} = \text{diag}\{\xi_1, \dots, \xi_R\} \in \mathbb{R}^{R \otimes d}$  with  $\ell$ -mode side matrices  $U^{(\ell)} = [\mathbf{u}_1^{(\ell)}, \dots, \mathbf{u}_R^{(\ell)}] \in \mathbb{R}^{n \times R}$  (see **Figure 1**). By definition the tensor  $\text{diag}\{\xi_1, \dots, \xi_R\} \in \mathbb{R}^{R \otimes d}$  is called diagonal if it has all zero entries except the diagonal elements given by  $\xi(i_\ell, \dots, i_\ell) = \xi_{i_\ell}$ ,  $\ell = 1, \dots, R$ . Then the problem of canonical to Tucker approximation can be solved by the method of reduced HOSVD (RHOSVD) introduced in [46]. The basic idea of the reduced HOSVD is that for large (function related) tensors given in the canonical format their HOSVD does not require the construction of a tensor in the full format and SVD based computation of its matrix unfolding. Instead, it is sufficient to compute the SVD of the directional matrices  $U^{(\ell)}$  in Equation (2) composed by only the vectors of the canonical tensor in every dimension separately, as shown in **Figure 1**. This will provide the initial guess for the Tucker orthogonal basis in the given dimension. For the practical applicability, the results of the approximation theory on the low-rank approximation to the multi-variate functions, exhibiting exponential error decay in the Tucker rank, are of the principal significance [51].

In the following, we suppose that  $n \leq R$  and denote the SVD of the side-matrix  $U^{(\ell)}$  by

$$U^{(\ell)} = Z^{(\ell)} D_\ell V^{(\ell)T} = \sum_{k=1}^n \sigma_{\ell, k} \mathbf{z}_k^{(\ell)} \mathbf{v}_k^{(\ell)T}, \quad \mathbf{z}_k^{(\ell)} \in \mathbb{R}^n, \quad \mathbf{v}_k^{(\ell)} \in \mathbb{R}^R, \quad (3)$$



**FIGURE 1** | Illustration to the contracted product representation Equation (2) of the rank- $R$  canonical tensor. The first factor corresponds to the diagonal coefficient tensor  $\xi$ .

with the orthogonal matrices  $Z^{(\ell)} = [\mathbf{z}_1^{(\ell)}, \dots, \mathbf{z}_n^{(\ell)}] \in \mathbb{R}^{n \times n}$ , and  $V^{(\ell)} = [\mathbf{v}_1^{(\ell)}, \dots, \mathbf{v}_n^{(\ell)}] \in \mathbb{R}^{R \times n}$ ,  $\ell = 1, \dots, d$ . We use the following notations for the vector entries,  $v_k^{(\ell)}(v) = v_{k,v}^{(\ell)}$  ( $v = 1, \dots, R$ ).

To fix the idea, we introduce the vector of rank parameters,  $\mathbf{r} = (r_1, \dots, r_d)$ , and let

$$U^{(\ell)} \approx W^{(\ell)} := Z_0^{(\ell)} D_{\ell,0} V_0^{(\ell)T}, \quad (4)$$

be the rank- $r_\ell$  truncated SVD of the side-matrix  $U^{(\ell)}$  ( $\ell = 1, \dots, d$ ). Here, the matrix  $D_{\ell,0} = \text{diag}\{\sigma_{\ell,1}, \sigma_{\ell,2}, \dots, \sigma_{\ell,r_\ell}\}$  is the submatrix of  $D_\ell$  in Equation (3) and

$$Z_0^{(\ell)} = [\mathbf{z}_1^{(\ell)}, \dots, \mathbf{z}_{r_\ell}^{(\ell)}] \in \mathbb{R}^{n \times r_\ell}, \quad V_0^{(\ell)} \in \mathbb{R}^{R \times r_\ell},$$

represent the respective dominating ( $n \times r_\ell$ )-submatrices of the left and right factors in the complete SVD decomposition in Equation (3).

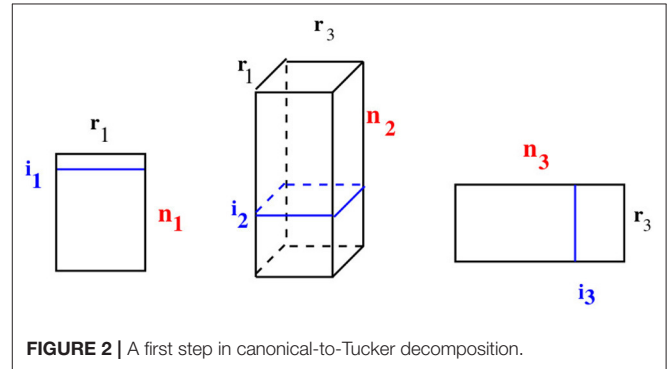
**Definition 2.1.** (Reduced HOSVD, [46]). Given the canonical tensor  $\mathbf{A} \in \mathcal{C}_{R,\mathbf{n}}$ , the truncation rank parameter  $\mathbf{r}$ , ( $r_\ell \leq R$ ), and rank- $r_\ell$  truncated SVD of  $U^{(\ell)}$ , see Equation (4), then the RHOSVD approximation of  $\mathbf{A}$  is defined by the rank- $\mathbf{r}$  orthogonal Tucker tensor

$$\begin{aligned} \mathbf{A}_{(\mathbf{r})}^0 &:= \xi \times_1 W^{(1)} \times_2 \cdots \times_d W^{(d)} = \xi \times_1 \left[ Z_0^{(1)} D_{1,0} V_0^{(1)T} \right] \\ &\times_2 \cdots \times_d \left[ Z_0^{(d)} D_{d,0} V_0^{(d)T} \right] \\ &= \left( \xi \times_1 [D_{1,0} V_0^{(1)T}] \times_2 \cdots \times_d [D_{d,0} V_0^{(d)T}] \right) \times_1 Z_0^{(1)} \\ &\times_2 \cdots \times_d Z_0^{(d)} \in \mathcal{T}_{\mathbf{r}}, \end{aligned} \quad (5)$$

obtained by the projection of canonical side matrices  $U^{(\ell)}$  onto the left orthogonal singular matrices  $Z_0^{(\ell)}$ , defined in Equation (4).

Notice that the general error bound for the RHOSVD approximation will be presented by Theorem 2.3, see also the discussion afterwards. Corollary 2.4 provides the conditions which guarantee the stability of RHOSVD.

The sub-optimal Tucker approximand Equation (5) is simple to compute and it provides accurate approximation to the initial canonical tensor even with rather small Tucker rank. Moreover,



**FIGURE 2** | A first step in canonical-to-Tucker decomposition.

this provides the good initial guess to calculate the best rank- $\mathbf{r}$  Tucker approximation by using the ALS iteration. In our numerical practice, usually, only one or two ALS iterations are required for convergence. For example, in case  $d = 3$ , algorithmically, the one step of the canonical-to-Tucker ALS algorithm reduces to the following operations. Substituting the orthogonal matrices  $Z_0^{(1)}$  and  $Z_0^{(3)}$  from Equation (5) into Equation (2), we perform the initial step of the first ALS iteration

$$\mathbf{A} \mapsto \mathbf{A}_{(1)} = Z_0^{(1)} \times_1 \mathbf{A}_2 \times_3 Z_0^{(3)}, \quad (6)$$

where  $\mathbf{A}_2$  is given by the contraction

$$\mathbf{A}_2 = \xi \times_1 D_{1,0} V_0^{(1)T} \times_2 U^{(2)} \times_3 D_{3,0} V_0^{(3)T} \in \mathbb{R}^{r_1 \times n_2 \times r_3},$$

as illustrated in Figure 2. Then we optimize the orthogonal subspace in the second variable by calculating the best rank- $r_2$  approximation to the  $r_1 r_3 \times n_2$  matrix unfolding of the tensor  $\mathbf{A}_2$ . The similar contracted product representation can be used when  $d > 3$ , as well as for the construction of the TT representation for the canonical input.

Here, we notice that the core tensor in the RHOSVD decomposition can be represented in the CP data-sparse format.

**Proposition 2.2.** The core tensor

$$\beta_0 = \xi \times_1 [D_{1,0} V_0^{(1)T}] \times_2 \cdots \times_d [D_{d,0} V_0^{(d)T}] \in \mathbb{R}^{r_1 \times \cdots \times r_d},$$

in the orthogonal Tucker representation Equation (5),  $\mathbf{A}_{(\mathbf{r})}^0 = \beta_0 \times_1 Z_0^{(1)} \times_2 \cdots \times_d Z_0^{(d)} \in \mathcal{T}_{\mathbf{r},\mathbf{n}}$ , can be recognized as

the rank- $R$  canonical tensor of size  $r_1 \times \dots \times r_d$  with the storage request  $R(\sum_{\ell=1}^d r_\ell)$ , which can be calculated entry-wise in  $O(Rr_1 \dots r_d)$  operations.

Indeed, introducing the matrices  $U_0^{(\ell)T} = D_{\ell,0} V_0^{(\ell)T} \in \mathbb{R}^{r_\ell \times R}$ , for  $\ell = 1, \dots, d$ , we conclude that the canonical core tensor  $\beta_0$  is determined by the  $\ell$ -mode side matrices  $U_0^{(\ell)T}$ . In the other words, the tensor  $\mathbf{A}_{(r)}^0$  is represented in the mixed Tucker-canonical format getting rid of the ‘‘curse of dimensionality’’ (see also Section 2.2 below).

The accuracy of the RHOSVD approximation can be controlled by the given  $\varepsilon$ -threshold in truncated SVD of side matrices  $U^{(\ell)}$ . The following theorem proves the absolute error bound for the RHOSVD approximation.

**Theorem 2.3.** (RHOSVD error bound, [46]). For given  $\mathbf{A} \in \mathcal{C}_{R,n}$  in Equation (1), let  $\sigma_{\ell,1} \geq \sigma_{\ell,2} \dots \geq \sigma_{\ell,\min(n,R)}$  be the singular values of  $\ell$ -mode side matrices  $U^{(\ell)} \in \mathbb{R}^{n \times R}$  ( $\ell = 1, \dots, d$ ) with normalized skeleton vectors. Then the error of RHOSVD approximation,  $\mathbf{A}_{(r)}^0$ , is bounded by

$$\|\mathbf{A} - \mathbf{A}_{(r)}^0\| \leq \|\xi\| \sum_{\ell=1}^d \left( \sum_{k=r_\ell+1}^{\min(n,R)} \sigma_{\ell,k}^2 \right)^{1/2}, \quad \|\xi\| = \sqrt{\sum_{v=1}^R \xi_v^2}. \quad (7)$$

The complete proof can be found in Section 8 (see Appendix).

The accuracy of the RHOSVD can be controlled in terms of the  $\varepsilon$ -criteria. To that end, given  $\varepsilon > 0$ , chose the Tucker ranks such that  $\sum_{\ell=1}^d \left( \sum_{k=r_\ell+1}^{\min(n,R)} \sigma_{\ell,k}^2 \right)^{1/2} \leq \varepsilon$  is satisfied, then Theorem 2.3 provided the error bound adapted to the  $\varepsilon$ -threshold.

The error estimate in Theorem 2.3 differs from the case of complete HOSVD by the extra factor  $\|\xi\|$ , which is the payoff for the lack of orthogonality in the canonical input tensor. Hence, Theorem 2.3 does not provide, in general, the stable control of relative error since for the general canonical tensors there is no uniform upper bound on the constant  $C$  in the estimate

$$\|\xi\| \leq C\|\mathbf{A}\|. \quad (8)$$

The problem is that Equation (8) applies to the general non-orthogonal canonical decomposition.

The stable RHOSVD approximation can be proven in the case of the so-called partially orthogonal or monotone decompositions. With partially orthogonal decomposition we mean that for each pair of indexes  $v, \mu$  in Equation (1) there holds  $\prod_{\ell=1}^d \langle \mathbf{u}_v^{(\ell)}, \mathbf{u}_\mu^{(\ell)} \rangle = 0$ . For monotone decompositions we assume that all coefficients and skeleton vectors in Equation (1) have non-negative values.

**Corollary 2.4.** (Stability of RHOSVD) Assume the conditions of Theorem 2.3 are satisfied. (A) Suppose that at least one of the side matrices  $U^{(\ell)}$ ,  $\ell = 1, \dots, d$ , in Equation (2), is orthogonal or

the decomposition Equation (1) is partially orthogonal. Then the RHOSVD error can be bounded by

$$\|\mathbf{A} - \mathbf{A}_{(r)}^0\| \leq C\|\mathbf{A}\| \sum_{\ell=1}^d \left( \sum_{k=r_\ell+1}^{\min(n,R)} \sigma_{\ell,k}^2 \right)^{1/2}. \quad (9)$$

(B) Let decomposition Equation (1) be monotone. Then (9) holds.

*Proof:* (A) The partial orthogonality assumption combined with normalization constraints for the canonical skeleton vectors imply

$$\begin{aligned} \|\mathbf{A}\|^2 &= \langle \sum_{v=1}^R \xi_v \mathbf{u}_v^{(1)} \otimes \dots \otimes \mathbf{u}_v^{(d)}, \sum_{v=1}^R \xi_v \mathbf{u}_v^{(1)} \otimes \dots \otimes \mathbf{u}_v^{(d)} \rangle \\ &= \sum_{v=1}^R \xi_v^2 \prod_{\ell=1}^d \langle \mathbf{u}_v^{(\ell)}, \mathbf{u}_v^{(\ell)} \rangle \\ &\quad + \sum_{v,\mu=1, v \neq \mu}^R \xi_v \xi_\mu \prod_{\ell=1}^d \langle \mathbf{u}_v^{(\ell)}, \mathbf{u}_\mu^{(\ell)} \rangle \\ &= \sum_{v=1}^R \xi_v^2 = \|\xi\|^2. \end{aligned}$$

The above relation also holds in the case of orthogonality of the side matrix  $U^{(\ell)}$  for some fixed  $\ell$ . Then the result follows by (7).

(B) In case of monotone decomposition we conclude that the pairwise scalar product of all summands in Equation (1) is non-negative, while the norm of each  $v$ -term is equal to  $\xi_v$ . Then the upper bound

$$\langle \mathbf{u}_1, \mathbf{u}_1 \rangle + \dots + \langle \mathbf{u}_R, \mathbf{u}_R \rangle \leq \left\langle \sum_{v=1}^R \mathbf{u}_v, \sum_{v=1}^R \mathbf{u}_v \right\rangle,$$

holds for vectors  $\mathbf{u}_v = \xi_v \mathbf{u}_v^{(1)} \otimes \dots \otimes \mathbf{u}_v^{(d)}$ ,  $v = 1, \dots, R$ , with non-negative entries applied to the case of  $R$  summands, thus implying  $\|\xi\|^2 \leq \|\mathbf{A}\|^2$ . Now, the result follows.

Clearly, the orthogonality assumption may lead to slightly higher separation rank, however, this constructive decomposition stabilizes the RHOSVD approximation method applied to the canonical format tensor (i.e., it allows the stable control of relative error). The case of monotone canonical sums typically arises in the sinc-based canonical approximation to radially symmetric Green’s kernels by a sum of Gaussians. On the other hand, in long term computational practice the numerical instability of RHOSVD approximation was not observed in case of physically relevant data.

## 2.2. Mixed Tucker Tensor Format and Tucker-to-CP Transform

In the procedure for the canonical tensor rank reduction the goal is to have a result in a canonical tensor format with a smaller rank. By converting the core tensor to CP format, one can use the mixed two-level Tucker data format [12, 27], or canonical CP format. Figure 3 illustrates the computational scheme of the two-level Tucker approximation.

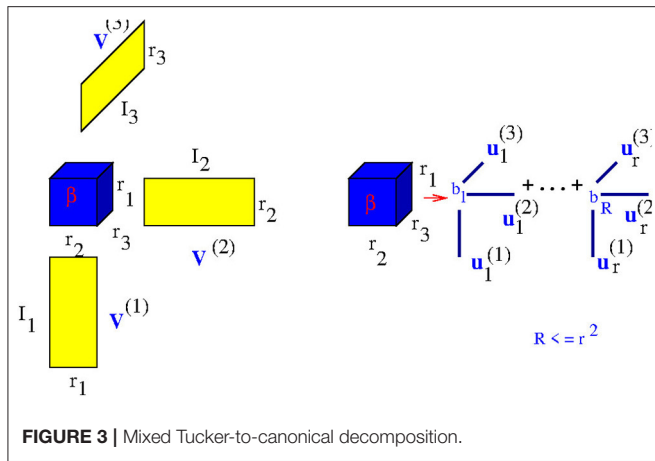


FIGURE 3 | Mixed Tucker-to-canonical decomposition.

We define by  $\bar{n}_\ell$  the single-hole product of dimension-modes,

$$\bar{n}_\ell = n_1 \cdots n_{\ell-1} n_{\ell+1} \cdots n_d. \tag{10}$$

The same definition applies to the quantity  $\bar{r}_\ell$ .

Next lemma describes the approximation of the Tucker tensor by using canonical representation [12, 27].

**Lemma 2.5.** (Mixed Tucker-to-canonical approximation, [27]).

(A) Let the target tensor  $\mathbf{A}$  have the form  $\mathbf{A} = \boldsymbol{\beta} \times_1 V^{(1)} \times_2 \dots \times_d V^{(d)} \in \mathcal{T}_{\mathbf{r}, \mathbf{n}}$ , with the orthogonal side-matrices  $V^{(\ell)} = [v_1^{(\ell)} \dots v_{r_\ell}^{(\ell)}] \in \mathbb{R}^{n \times r_\ell}$  and  $\boldsymbol{\beta} \in \mathbb{R}^{r_1 \times \dots \times r_d}$ . Then, for a given  $R \leq \min_{1 \leq \ell \leq d} \bar{r}_\ell$ ,

$$\min_{\mathbf{Z} \in \mathcal{C}_{R, \mathbf{n}}} \|\mathbf{A} - \mathbf{Z}\| = \min_{\boldsymbol{\mu} \in \mathcal{C}_{R, \mathbf{r}}} \|\boldsymbol{\beta} - \boldsymbol{\mu}\|. \tag{11}$$

(B) Assume that there exists the best rank- $R$  approximation  $\mathbf{A}_{(R)} \in \mathcal{C}_{R, \mathbf{n}}$  of  $\mathbf{A}$ , then there is the best rank- $R$  approximation  $\boldsymbol{\beta}_{(R)} \in \mathcal{C}_{R, \mathbf{r}}$  of  $\boldsymbol{\beta}$ , such that

$$\mathbf{A}_{(R)} = \boldsymbol{\beta}_{(R)} \times_1 V^{(1)} \times_2 \dots \times_d V^{(d)}. \tag{12}$$

The complete proof can be found in Section 8 (see Appendix). Notice that condition  $R \leq \min_{1 \leq \ell \leq d} \bar{r}_\ell$  simply means that the canonical rank does not exceed the maximal CP rank of the Tucker core tensor.

Combination of Theorem 2.3 and Lemma 2.5 paves the way to the rank optimization of canonical tensors with the large mode-size arising, for example, in the grid-based numerical methods for multi-dimensional PDEs with non-regular (singular) solutions. In such applications the univariate grid-size (i.e., the mode-size) may be about  $n = 10^4$  and even larger.

Notice that the Tucker (for moderate  $d$ ) and canonical formats allow to perform basic multi-linear algebra using one-dimensional operations, thus reducing the exponential scaling in  $d$ . Rank-truncated transforms between different formats can be applied in multi-linear algebra on mixed tensor representations as well, see Lemma 2.5. The particular application to tensor

convolution in many dimensions was discussed, for example, in [1, 2].

We summarize that the direct methods of tensor approximation can be classified by:

- (1) Analytic Tucker approximation to some classes of function-related  $d$ th order tensors ( $d \geq 2$ ), say, by multi-variate polynomial interpolation [1].
- (2) Sinc quadrature based approximation methods in the canonical format applied to a class of analytic function related tensors [11].
- (3) Truncated HOSVD and RHOSVD, for quasi-optimal Tucker approximation of the full-format, respectively, canonical tensors [46].

Direct analytic approximation methods by sinc quadrature/interpolation are of principal importance. Basic examples are given by the tensor representation of Green's kernels, the elliptic operator inverse and analytic matrix-valued functions. In all cases, the algebraic methods for rank reduction by the ALS-type iterative Tucker/canonical approximation can be applied.

Further improvement and enhancement of algebraic tensor approximation methods can be based on the combination of advanced nonlinear iteration, multigrid tensor methods, greedy algorithms, hybrid tensor representations, and the use of new problem adapted tensor formats.

### 2.3. Tucker-to-Canonical Transform

In the rank reduction scheme for the canonical rank- $R$  tensors, we use successively the canonical-to-Tucker (C2T) transform and then the Tucker-to-canonical (T2C) tensor approximation.

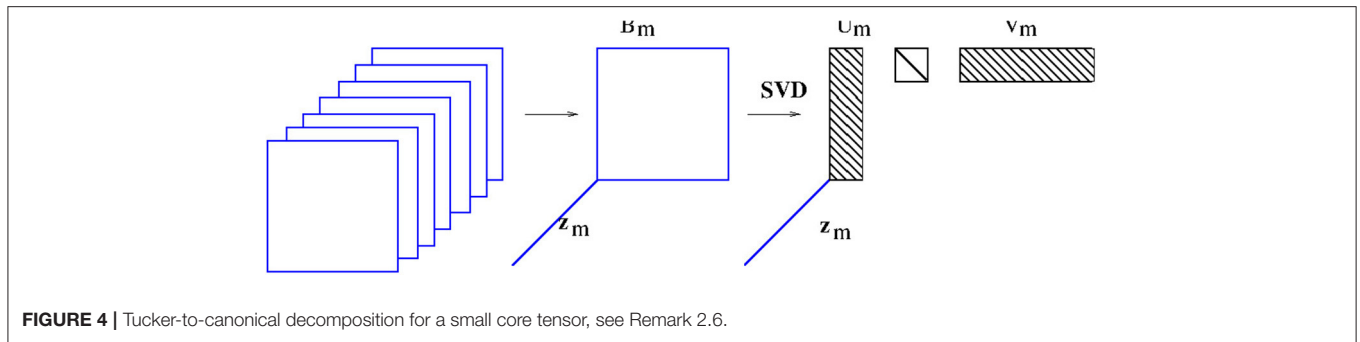
First, we notice that the canonical rank of a tensor  $\mathbf{A} \in \mathbb{V}_{\mathbf{n}}$  has the upper bound (see [27, 46]),

$$R \leq \min_{1 \leq \ell \leq d} \bar{n}_\ell, \tag{13}$$

where  $\bar{n}_\ell$  is given by Equation (10). Rank bound (13) applied to the Tucker core tensor of the size  $r \times r \times r$ , indicates that the ultimate canonical rank of a large-size tensor in  $\mathbb{V}_{\mathbf{n}}$  has the upper bound  $r^2$ . Notice that for function related tensors the Tucker rank scales logarithmically in both approximation accuracy and the discretization grid size (see the proof for some classes of function in [51]).

The following remark shows that the maximal canonical rank of the Tucker core of 3rd order tensor can be easily reduced to the value less than  $r^2$  by the SVD-based procedure applied to the matrix slices of the Tucker core tensor  $\boldsymbol{\beta}$ . Though, being not practically attractive for arbitrary high order tensors, the simple algorithm described in Remark 2.6 below is proved to be useful for the treatment of small size 3rd order Tucker core tensors within the rank reduction algorithms described in the previous sections.

**Remark 2.6.** Let  $d = 3$  for the sake of clarity [27, 46]. There is a simple procedure based on SVD to reduce the canonical rank of the core tensor  $\boldsymbol{\beta}$ , within the accuracy  $\varepsilon > 0$ . Denote by  $B_m \in \mathbb{R}^{r \times r}$ ,



**FIGURE 4** | Tucker-to-canonical decomposition for a small core tensor, see Remark 2.6.

$m = 1, \dots, r$  the two-dimensional slices of  $\beta$  in each fixed mode and represent

$$\beta = \sum_{m=1}^r B_m \otimes \mathbf{z}_m, \quad \mathbf{z}_m \in \mathbb{R}^r, \quad (14)$$

where  $\mathbf{z}_m(m) = 1, \mathbf{z}_m(j) = 0$  for  $j = 1, \dots, r, j \neq m$  (there are exactly  $d$  possible decompositions). Let  $p_m$  be the minimal integer, such that the singular values of  $B_m$  satisfy  $\sigma_k^{(m)} \leq \frac{\epsilon}{r^{3/2}}$  for  $k = p_m + 1, \dots, r$  (if  $\sigma_r^{(m)} > \frac{\epsilon}{r^{3/2}}$ , then set  $p_m = r$ ). Then, denoting by

$$B_{p_m} = \sum_{k=1}^{p_m} \sigma_{k_m}^{(m)} \mathbf{u}_{k_m} \otimes \mathbf{v}_{k_m},$$

the corresponding rank- $p_m$  approximation to  $B_m$  (by truncation of  $\sigma_{p_m+1}^{(m)}, \dots, \sigma_r^{(m)}$ ), we arrive at the rank- $R$  canonical approximation to  $\beta$ ,

$$\beta_{(R)} := \sum_{m=1}^r B_{p_m} \otimes \mathbf{z}_m, \quad \mathbf{z}_m \in \mathbb{R}^r, \quad (15)$$

providing the error estimate

$$\begin{aligned} \|\beta - \beta_{(R)}\| &\leq \sum_{m=1}^r \|B_m - B_{p_m}\| = \sum_{m=1}^r \sqrt{\sum_{k=p_m+1}^r (\sigma_{k_m}^{(m)})^2} \\ &\leq \sum_{m=1}^r \sqrt{r \frac{\epsilon^2}{r^3}} = \epsilon \end{aligned}$$

Representation (15) is a sum of rank- $p_m$  terms so that the total rank is bounded by  $R \leq p_1 + \dots + p_r \leq r^2$ . The approach can be extended to arbitrary  $d \geq 3$  with the bound  $R \leq r^{d-1}$ .

**Figure 4** illustrates the canonical decomposition of the core tensor by using the SVD of slices  $B_m$  of the core tensor  $\beta$ , yielding matrices  $U_m = \{\mathbf{u}_{k_m}\}_{k=1}^{p_m}$ ,  $V_m = \{\mathbf{v}_{k_m}\}_{k=1}^{p_m}$  and a diagonal matrix of small size  $p_m \times p_m$  containing the truncated singular values. It also shows the vector  $\mathbf{z}_m = [0, \dots, 0, 1, 0, \dots, 0]$ , containing all entries equal to 0 except 1 at the  $m$ th position.

It is worse to note that the rank reduction for the rank- $R$  core tensor of small size  $r_1 \times \dots \times r_d$ , can be also performed

by using the cascading ALS algorithms in CP format applied to the canonical input tensor, as it was applied in [50]. Moreover, a number of numerical examples presented in the present paper and in the included literature (applied to function generated tensors) demonstrate the substantial reduction of the initial canonical rank  $R$ .

### 3. CALCULATION OF 3D INTEGRALS WITH THE NEWTON KERNEL

The first application of the RHOSVD was calculation of the 3D grid-based Hartree potential operator in the Hartree-Fock equation,

$$V_H(x) := \int_{\mathbb{R}^3} \frac{\rho(y)}{\|x - y\|} dy, \quad (16)$$

where the electron density,

$$\rho(x) = 2 \sum_{a=1}^{N_{orb}} (\varphi_a)^2, \quad (17)$$

is represented in terms of molecular orbitals, presented in the Gaussian-type basis (GTO),  $\varphi_a(x) = \sum_{k=1}^{N_b} c_{a,k} g_k(x)$ . The Hartree potential describes the repulsion energy of the electrons in a molecule. The intermediate goal here is the calculation of the so-called Coulomb matrix,

$$J_{km} := \int_{\mathbb{R}^3} g_k(x) g_m(x) V_H(x) dx, \quad k, m = 1, \dots, N_b \quad x \in \mathbb{R}^3,$$

which represents the Hartree potential in the given GTO basis.

In fact, calculation of this 3D convolution operator with the Newton kernel, requires high accuracy and it should be repeated multiply in the course of the iterative solution of the Hartree-Fock nonlinear eigenvalue problem. The presence of nuclear cusps in the electron density makes additional challenge to computation of the Hartree potential operator. Traditionally, these calculations are based on involved analytical evaluation of the corresponding integral in a separable Gaussian basis set by using erf function. Tensor-structured calculation of the multi-dimensional convolution integral operators with the Newton kernel have been introduced in [27, 29, 46].

The molecule is embedded in a computational box  $\Omega = [-b, b]^3 \in \mathbb{R}^3$ . The equidistant  $n \times n \times n$  tensor grid  $\omega_{3,n} = \{x_i\}$ ,  $\mathbf{i} \in \mathcal{I} := \{1, \dots, n\}^3$ , with the mesh-size  $h = 2b/(n + 1)$  is used. In calculations of integral terms, the Gaussian basis functions  $g_k(x)$ ,  $x \in \mathbb{R}^3$ , are approximated by sampling their values at the centers of discretization intervals using one-dimensional piecewise constant basis functions  $g_k(x) \approx \bar{g}_k(x) = \prod_{\ell=1}^3 \bar{g}_k^{(\ell)}(x_\ell)$ ,  $\ell = 1, 2, 3$ , yielding their rank-1 tensor representation,

$$\mathbf{G}_k = \mathbf{g}_k^{(1)} \otimes \mathbf{g}_k^{(2)} \otimes \mathbf{g}_k^{(3)} \in \mathbb{R}^{n \times n \times n}, \quad k = 1, \dots, N_b. \quad (18)$$

Given the discrete tensor representation of basis functions (18), the electron density is approximated using 1D Hadamard products of rank-1 tensors as

$$\begin{aligned} \rho \approx \Theta &= 2 \sum_{a=1}^{N_{orb}} \sum_{k=1}^{N_b} \sum_{m=1}^{N_b} c_{a,m} c_{a,k} (\mathbf{g}_k^{(1)} \odot \mathbf{g}_m^{(1)}) \otimes \dots \otimes (\mathbf{g}_k^{(3)} \odot \mathbf{g}_m^{(3)}) \\ &\in \mathbb{R}^{n \times n \times n}. \end{aligned} \quad (19)$$

For convolution operator, the representation of the Newton kernel  $\frac{1}{\|x-y\|}$  by a canonical rank- $R_N$  tensor [1] is used (see Section 4.1 for details),

$$\mathbf{P}_R = \sum_{q=1}^{R_N} \mathbf{p}_q^{(1)} \otimes \mathbf{p}_q^{(2)} \otimes \mathbf{p}_q^{(3)} \in \mathbb{R}^{n \times n \times n}. \quad (20)$$

The initial rank of the electron density in the canonical tensor format  $\Theta$  in Equation (17) is large even for small molecules. Rank reduction by using RHOSVD C2T plus T2C reduces the rank  $\Theta \mapsto \Theta'$  by several orders of magnitude, from  $N_b^2/2$  to  $R_\rho \ll N_b^2/2$ , from  $\sim 10^4$  to  $\sim 10^2$ . Then the 3D tensor representation of the Hartree potential is calculated by using the 3D tensor product convolution, which is a sum of tensor products of 1D convolutions,

$$\begin{aligned} V_H \approx \mathbf{V}_H &= \Theta' * \mathbf{P}_R = \sum_{m=1}^{R_\rho} \sum_{q=1}^{R_N} c_m (\mathbf{u}_m^{(1)} * \mathbf{p}_q^{(1)}) \otimes (\mathbf{u}_m^{(2)} * \mathbf{p}_q^{(2)}) \\ &\otimes (\mathbf{u}_m^{(3)} * \mathbf{p}_q^{(3)}). \end{aligned}$$

The Coulomb matrix entries  $J_{km}$  are obtained by 1D scalar products of  $\mathbf{V}_H$  with the Galerkin basis consisting of rank-1 tensors,

$$J_{km} \approx \langle \mathbf{G}_k \odot \mathbf{G}_m, \mathbf{V}_H \rangle, \quad k, m = 1, \dots, N_b.$$

The cost of 3D tensor product convolution is  $O(n \log n)$  instead of  $O(n^3 \log n)$  for the standard benchmark 3D convolution using the 3D FFT. Table 1 shows CPU times (sec) for the Matlab computation of  $V_H$  for  $H_2O$  molecule [46] on a SUN station using 8 Opteron Dual-Core/2600 processors (times for 3D FFT for  $n \geq 1024$  are obtained by extrapolation). C2T shows the time for the canonical-to-Tucker rank reduction.

**TABLE 1** | Times (sec) for the C2T transform and the 3D tensor product convolution vs. 3D FFT convolution.

$n^3$	1024 <sup>3</sup>	2048 <sup>3</sup>	4096 <sup>3</sup>	8192 <sup>3</sup>	16384 <sup>3</sup>
FFT <sub>3</sub>	~ 6000	–	–	–	~ 2 years
C*C	8.8	20.0	61.0	157.5	299.2
C2T	6.9	10.9	20.0	37.9	86.0

The grid-based tensor calculation of the multi-dimensional integrals in quantum chemistry provides the required high accuracy by using large grids and the ranks are controlled by the required  $\varepsilon$  in the rank truncation algorithms. The results of the tensor-based calculations have been compared with the results of the benchmark standard computations by the MOLPRO package. It was shown that the accuracy is of the order of  $10^{-7}$  hartree in the resulting ground state energy (see [2, 27]).

## 4. RHOSVD IN THE RANGE-SEPARATED TENSOR FORMATS

The range-separated (RS) tensor formats have been introduced in [32] as the constructive tool for low-rank tensor representation (approximation) of function related data discretized on Cartesian grids in  $\mathbb{R}^d$ , which may have multiple singularities or cusps. Such highly non-regular data typically arise in computational quantum chemistry, in many-particle dynamics simulations and many-particle electrostatics calculations, in protein modeling and in data science. The key idea of the RS representation is the splitting of the short- and long-range parts in the functional data and further low-rank approximation of the rather regular long-range part in the classical tensor formats.

In this concern RHOSVD method becomes an essential ingredient of the rank reduction algorithms for the “long-range” input tensor, which usually inherits the large initial rank.

### 4.1. Low-Rank Approximation of Radial Functions

First, we recall the grid-based method for the low-rank canonical representation of a spherically symmetric kernel functions  $p(\|x\|)$ ,  $x \in \mathbb{R}^d$  for  $d = 2, 3, \dots$ , by its projection onto the finite set of basis functions defined on tensor grid. The approximation theory by a sum of Gaussians for the class of analytic potentials  $p(\|x\|)$  was presented in [1, 11, 51, 52]. The particular numerical schemes for rank-structured representation of the Newton and Slater kernels

$$p(\|x\|) = \frac{1}{4\pi \|x\|}, \quad \text{and} \quad p(\|x\|) = e^{-\lambda \|x\|}, \quad x \in \mathbb{R}^3, \quad (21)$$

discretized on a fine 3D Cartesian grid in the form of low-rank canonical tensor was described in [11, 51].

In what follows, for the ease of exposition, we confine ourselves to the case  $d = 3$ . In the computational domain  $\Omega = [-b, b]^3$ , let us introduce the uniform  $n \times n \times n$  rectangular Cartesian grid  $\Omega_n$  with mesh size  $h = 2b/n$  ( $n$  even). Let  $\{\psi_{\mathbf{i}} =$



$\prod_{\ell=1}^3 \psi_{i_\ell}^{(\ell)}(x_\ell)$  be a set of tensor-product piecewise constant basis functions, labeled by the 3-tuple index  $\mathbf{i} = (i_1, i_2, i_3)$ ,  $i_\ell \in I_\ell = \{1, \dots, n\}$ ,  $\ell = 1, 2, 3$ . The generating kernel  $p(\|x\|)$  is discretized by its projection onto the basis set  $\{\psi_{\mathbf{i}}\}$  in the form of a third order tensor of size  $n \times n \times n$ , defined entry-wise as

$$\mathbf{P} := [p_{\mathbf{i}}] \in \mathbb{R}^{n \times n \times n}, \quad p_{\mathbf{i}} = \int_{\mathbb{R}^3} \psi_{\mathbf{i}}(x) p(\|x\|) \, dx. \quad (22)$$

The low-rank canonical decomposition of the 3rd order tensor  $\mathbf{P}$  is based on using exponentially fast convergent sinc-quadratures for approximating the Laplace-Gauss transform to the analytic function  $p(z)$ ,  $z \in \mathbb{C}$ , specified by a certain weight  $\widehat{p}(t) > 0$ ,

$$p(z) = \int_{\mathbb{R}_+} \widehat{p}(t) e^{-t^2 z^2} \, dt \approx \sum_{k=-M}^M p_k e^{-t_k^2 z^2} \quad \text{for } |z| > 0, z \in \mathbb{R}, \quad (23)$$

with the proper choice of the quadrature points  $t_k$  and weights  $p_k$ . The *sinc*-quadrature based approximation to generating function by using the short-term Gaussian sums in Equation (23) are applicable to the class of analytic functions in certain strip  $|z| \leq D$  in the complex plane, such that on the real axis these functions decay polynomially or exponentially. We refer to basic results in [11, 52, 53], where the exponential convergence of the *sinc*-approximation in the number of terms (i.e., the canonical rank) was analyzed for certain classes of analytic integrands.

Now, for any fixed  $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ , such that  $\|x\| > a > 0$ , we apply the sinc-quadrature approximation Equation (23) to obtain the separable expansion

$$\begin{aligned} p(\|x\|) &= \int_{\mathbb{R}_+} \widehat{p}(t) e^{-t^2 \|x\|^2} \, dt \approx \sum_{k=-M}^M p_k e^{-t_k^2 \|x\|^2} \\ &= \sum_{k=-M}^M p_k \prod_{\ell=1}^3 e^{-t_k^2 x_\ell^2}, \end{aligned} \quad (24)$$

providing an exponential convergence rate in  $M$ ,

$$\left| p(\|x\|) - \sum_{k=-M}^M p_k e^{-t_k^2 \|x\|^2} \right| \leq \frac{C}{a} e^{-\beta \sqrt{M}}, \quad \text{with some } C, \beta > 0. \quad (25)$$

In the case of Newton kernel, we have  $p(z) = 1/z$ ,  $\widehat{p}(t) = \frac{2}{\sqrt{\pi}}$ , so that the Laplace-Gauss transform representation reads

$$\frac{1}{z} = \frac{2}{\sqrt{\pi}} \int_{\mathbb{R}_+} e^{-z^2 t^2} \, dt, \quad \text{where } z = \|x\|, \quad x \in \mathbb{R}^3, \quad (26)$$

which can be approximated by the sinc quadrature Equation (24) with the particular choice of quadrature points  $t_k$ , providing the exponential convergence rate as in Equation (25) [11, 51].

In the case of Yukawa potential the Laplace Gauss transform reads

$$\frac{e^{-\kappa z}}{z} = \frac{2}{\sqrt{\pi}} \int_{\mathbb{R}_+} e^{-\kappa^2/t^2} e^{-z^2 t^2} \, dt, \quad \text{where } z = \|x\|, \quad x \in \mathbb{R}^3. \quad (27)$$

The analysis of the sinc quadrature approximation error for this case can be found, in particular, in [1, 51], section 2.4.7.

Combining (22) and (24), and taking into account the separability of the Gaussian basis functions, we arrive at the low-rank approximation to each entry of the tensor  $\mathbf{P} = [p_{\mathbf{i}}]$ ,

$$\begin{aligned} p_{\mathbf{i}} &\approx \sum_{k=-M}^M p_k \int_{\mathbb{R}^3} \psi_{\mathbf{i}}(x) e^{-t_k^2 \|x\|^2} \, dx \\ &= \sum_{k=-M}^M p_k \prod_{\ell=1}^3 \int_{\mathbb{R}} \psi_{i_\ell}^{(\ell)}(x_\ell) e^{-t_k^2 x_\ell^2} \, dx_\ell. \end{aligned}$$

Define the vector (recall that  $p_k > 0$ )

$$\begin{aligned} \mathbf{p}_k^{(\ell)} &= p_k^{1/3} \left[ b_{i_\ell}^{(\ell)}(t_k) \right]_{i_\ell=1}^{n_\ell} \in \mathbb{R}^{n_\ell} \quad \text{with} \\ b_{i_\ell}^{(\ell)}(t_k) &= \int_{\mathbb{R}} \psi_{i_\ell}^{(\ell)}(x_\ell) e^{-t_k^2 x_\ell^2} \, dx_\ell, \end{aligned} \quad (28)$$

then the 3rd order tensor  $\mathbf{P}$  can be approximated by the  $R$ -term ( $R = 2M + 1$ ) canonical representation

$$\begin{aligned} \mathbf{P} \approx \mathbf{P}_R &= \sum_{k=-M}^M p_k \bigotimes_{\ell=1}^3 \mathbf{b}^{(\ell)}(t_k) \\ &= \sum_{k=-M}^M \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)} \in \mathbb{R}^{n \times n \times n}, \mathbf{p}_k^{(\ell)} \in \mathbb{R}^{n_\ell}. \end{aligned} \quad (29)$$

Given a threshold  $\varepsilon > 0$ , in view of Equation (25), we can choose  $M = O(\log^2 \varepsilon)$  such that in the max-norm

$$\|\mathbf{P} - \mathbf{P}_R\| \leq \varepsilon \|\mathbf{P}\|.$$

In the case of continuous radial function  $p(\|x\|)$ , say the Slater potential, we use the collocation type discretization at the grid points including the origin,  $x = 0$ , so that the univariate mode size becomes  $n \rightarrow n_1 = n + 1$ . In what follows, we use the same notation  $\mathbf{P}_R$  in the case of collocation type tensors (for example, the Slater potential) so that the particular meaning becomes clear from the context.

## 4.2. The RS Tensor Format Revisited

The range separated (RS) tensor format was introduced in [32] for efficient representation of the collective free-space electrostatic potential of large biomolecules. This rank-structured tensor representation of the collective electrostatic potential of many-particle systems of general type allows to reduce essentially computation of their interaction energy, and it provides convenient form for performing other algebraic transforms. The RS format proved to be useful for range-separated tensor representation of the Dirac delta [34] in  $\mathbb{R}^d$  and based on that, for regularization of the Poisson-Boltzmann equation (PBE) by decomposition of the solution into short- and long-range parts, where the short-range part of the solution is evaluated by simple tensor operations without solving the PDE. The smooth long-range part is calculated by solving the PBE with the modified

right-hand side by using the RS decomposition of the Dirac delta, so that now it does not contain singularities. We refer to papers [33, 35] describing the approach in details.

First, we recall the definition of the range separated (RS) tensor format, see [32], for representation of  $d$ -tensors  $\mathbf{A} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ . The RS format is served for the hybrid tensor approximation of discretized functions with multiple cusps or singularities. This allows the splitting of the target tensor onto the highly localized components approximating the singularity and the component with global support that allows the low-rank tensor approximation. Such functions typically arise in computational quantum chemistry, in many-particle modeling and in the interpolation of multi-dimensional data measured at certain set of spatial points in  $\mathbb{R}^{n \times n \times n}$ .

In the following definition of RS-canonical tensor format, we use the notion of localized canonical tensor  $\mathbf{U}_0$ , which is characterized by the small support whose diameter has a size of a few grid points. This tensor will be used as the reference one for presentation of the short-range part in the RS tensor. To that end we use the operation  $\text{Replica}_{x_v}(\mathbf{U}_0)$  which replicates  $\mathbf{U}_0$  into some given grid point  $x_v$ . In this construction, we assume that the chosen grid points  $x_v$  are well separated, i.e., the distance between each pair of points is not less than some given threshold  $n_\delta > 0$ .

**Definition 4.1.** (RS-canonical tensors, [32]). Given the rank- $R_s$  reference localized CP tensor  $\mathbf{U}_0$ . The RS-canonical tensor format defines the class of  $d$ -tensors  $\mathbf{A} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ , represented as a sum of a rank- $R_l$  CP tensor  $\mathbf{U}_{long} = \sum_{k=1}^{R_l} \xi_k \mathbf{u}_k^{(1)} \otimes \dots \otimes \mathbf{u}_k^{(d)}$ , and a cumulated CP tensor  $\mathbf{U}_{short} = \sum_{v=1}^{N_0} c_v \mathbf{U}_v$ , such that

$$\mathbf{A} = \mathbf{U}_{long} + \mathbf{U}_{short} = \sum_{k=1}^{R_l} \xi_k \mathbf{u}_k^{(1)} \otimes \dots \otimes \mathbf{u}_k^{(d)} + \sum_{v=1}^{N_0} c_v \mathbf{U}_v, \tag{30}$$

where  $\mathbf{U}_{short}$  is generated by the localized reference CP tensor  $\mathbf{U}_0$ , i.e.,  $\mathbf{U}_v = \text{Replica}_{x_v}(\mathbf{U}_0)$ , with  $\text{rank}(\mathbf{U}_v) = \text{rank}(\mathbf{U}_0) \leq R_s$ , where, given the threshold  $n_\delta > 0$ , the effective support of  $\mathbf{U}_v$  is bounded by  $\text{diam}(\text{supp}\mathbf{U}_v) \leq 2n_\delta$  in the index size.

Each RS-canonical tensor is, therefore, uniquely defined by the following parametrization: rank- $R_l$  canonical tensor  $\mathbf{U}_{long}$ , the rank- $R_s$  reference canonical tensor  $\mathbf{U}_0$  with the small mode size bounded by  $2n_\delta$ , list  $\mathcal{J}$  of the coordinates and weights of  $N_0$  particles in  $\mathbb{R}^d$ . The storage size is linear in both the dimension and the univariate grid size,

$$\text{stor}(\mathbf{A}) \leq dR_l n + (d + 1)N_0 + dR_s n_\delta.$$

The main benefit of the RS-canonical tensor decomposition is the almost uniform bound on the CP/Tucker rank of the long-range part  $\mathbf{U}_{long} = \sum_{k=1}^{R_l} \xi_k \mathbf{u}_k^{(1)} \otimes \dots \otimes \mathbf{u}_k^{(d)}$ , in the multi-particle potential discretized on fine  $n \times n \times n$  spatial grid. It was proven in [32] that the canonical rank  $R$  scales logarithmically in both the number of particles  $N_0$  and the approximation precision, see also Lemma 4.5.

Given the rank- $R$  CP decomposition Equation (29) based on the sinc-quadrature approximation Equation (24) of the discretized radial function  $p(\|x\|)$ , we define the two subsets of

indices,  $\mathcal{K}_l := \{k : t_k \leq 1\}$  and  $\mathcal{K}_s := \{k : t_k > 1\}$ , and then introduce the RS-representation of this tensor as follows,

$$\mathbf{P}_R = \mathbf{P}_{R_l} + \mathbf{P}_{R_s}, \quad R = R_l + R_s, \quad R_l = \#\mathcal{K}_l, \quad R_s = \#\mathcal{K}_s, \tag{31}$$

where

$$\mathbf{P}_{R_l} := \sum_{k \in \mathcal{K}_l} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)}, \quad \mathbf{P}_{R_s} := \sum_{k \in \mathcal{K}_s} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)}.$$

This representation allows to reduce the calculation of the multi-particle interaction energy of the many-particle system. Recall that the electrostatic interaction energy of  $N$  charged particles is represented in the form

$$E_N = E_N(x_1, \dots, x_N) = \sum_{i=1}^N \sum_{j < i}^N \frac{z_i z_j}{\|x_i - x_j\|}, \tag{32}$$

and it can be computed by direct summation in  $O(N^2)$  operations. The following statement is the modification of Lemma 4.2 in [32] (see [54] for more details).

**Lemma 4.2.** [54] Let the effective support of the short-range components in the reference potential  $\mathbf{P}_R$  for the Newton kernel does not exceed the minimal distance between particles,  $\sigma > 0$ . Then the interaction energy  $E_N$  of the  $N$ -particle system can be calculated by using only the long range part in the tensor  $\mathbf{P}$  representing on the grid the total potential sum,

$$E_N = \frac{1}{2} \sum_{j=1}^N z_j (\mathbf{P}_l(x_j) - z_j \mathbf{P}_{R_l}(0)) = \frac{1}{2} \langle \mathbf{z}, \mathbf{P}_l \rangle - \frac{\mathbf{P}_{R_l}(0)}{2} \sum_{j=1}^N z_j^2, \tag{33}$$

in  $O(R_l N)$  operations, where  $R_l$  is the canonical rank of the long-range component in  $\mathbf{P}$ ,  $\mathbf{P}_l$ .

Here,  $\mathbf{z} \in \mathbb{R}^N$  is a vector composed of all charges of the multi-particle systems, and  $\mathbf{p}_l \in \mathbb{R}^N$  is the vector of samples of the collective electrostatic long-range potential  $\mathbf{P}_l$  in the nodes corresponding to particle locations. Thus, the term  $\frac{1}{2} \langle \mathbf{z}, \mathbf{P}_l \rangle$  denotes the “non-calibrated” interaction energy associated with the long-range tensor component  $\mathbf{P}_l$ , while  $\mathbf{P}_{R_l}$  denotes the long-range part in the tensor representing the *single reference Newton kernel*, and  $\mathbf{P}_{R_l}(0)$  is its value at the origin.

Lemma 4.2 indicates that the interaction energy does not depend on the short-range part in the collective potential, and this is the key point for the construction of energy preserving regularized numerical schemes for solving the basic equations in bio-molecular modeling by using low-rank tensor decompositions.

### 4.3. Multi-Linear Operations in RS Tensor Formats

In what follows, we address the important question on how the *basic multi-linear operations* can be implemented in the RS tensor format by using the RHOSVD rank compression. The point is that various tensor operations arise in the

course of commonly used numerical schemes and iterative algorithms which usually include many sums and products of functions as well as the actions of differential/integral operators, always making the tensor structure of input data much more complicated requiring the robust rank reduction schemes.

The other important aspect is related to the use of large (fine resolution) discretization grids which is limited by the restriction on the size of the full input tensors,  $O(n^d)$  (curse of dimensionality), representing the discrete functions and operators to be approximated in low rank tensor format. Remarkably, that tensor decomposition for special class of functions, which allow the sinc-quadrature approximation, can be performed on practically indefinitely large grids because the storage and numerical costs of such numerical schemes scale linearly in the univariate grid size,  $O(dn)$ . Hence, having constructed such low rank approximations for certain set of “reproducing” radial functions, makes it possible to construct the low rank RS representation at linear complexity,  $O(dn)$ , for the wide class of functions and operators by using the rank truncated multi-linear operations. The examples of such “reproducing” radial functions are commonly used in our computational practice.

First, consider the Hadamard product of two tensors  $\mathbf{P}_R$  and  $\mathbf{Q}_{R_1}$  corresponding to the pointwise product of two generating multi-variate functions centered at the same point. The RS representation of the product tensor is based on the observation that the long-range part of the Hadamard product of two tensors in RS-format is basically determined by the product of their long-range parts.

**Lemma 4.3.** *Suppose that the RS representation Equation (31) of tensors  $\mathbf{P}_R$  and  $\mathbf{Q}_{R_1}$  is constructed based on the sinc-quadrature CP approximation Equation (29). Then the long-range part of the Hadamard product of these RS-tensors,*

$$\mathbf{Z} = (\mathbf{P}_s + \mathbf{P}_l) \odot (\mathbf{Q}_s + \mathbf{Q}_l),$$

can be represented by the product of their long-range parts,  $\mathbf{Z}_l = \mathbf{P}_l \odot \mathbf{Q}_l$ , with the subsequent rank reduction. Moreover, we have  $\text{rank}(\mathbf{Z}_l) \leq R_l Q_l$ .

*Proof:* We consider the case of collocation tensors and suppose that each skeleton vector in CP tensors  $\mathbf{P}_R$  and  $\mathbf{Q}_{R_1}$  is given by the restriction of certain Gaussians to the set of grid points. Chose the arbitrary short-range components in  $\mathbf{P}_R$  and some component in  $\mathbf{Q}_{R_1}$ , generated by Gaussians  $e^{-t_k x_k^2}$  and  $e^{-t_m x_m^2}$ , respectively. Then the effective support of the product of these two terms becomes smaller than that for each of the factors in view of the identity  $e^{-t_k x_k^2} e^{-t_m x_m^2} = e^{-(t_k+t_m)x_k^2}$  considered for arbitrary  $t_k, t_m > 0$ . This means that each term that includes the short-range multiple remains to be in the short range. Then the long range part in  $\mathbf{Z}$  takes a form  $\mathbf{Z}_l = \mathbf{P}_l \odot \mathbf{Q}_l$  with the subsequent rank reduction.

The sums of several tensors in RS format can be easily split into short- and long-range parts by grouping the respective components in the summands. The other important operation is the operator-function product in RS tensor format (see the

example in [34] related to the action of Laplacian with the singular Newton kernel resulting in the RS decomposition of the Dirac delta). This topic will be considered in detail elsewhere.

### 4.4. Representing the Slater Potential in RS Tensor Format

In what follows, we consider the RS-canonical tensor format for the rank-structured representation of the Slater function

$$G(x) = e^{-\lambda \|x\|}, \quad \lambda \in \mathbb{R}_+,$$

which has the principal significance in electronic structure calculations (say, based on the Hartree-Fock equation) since it represents the cusp behavior of electron density in the local vicinity of nuclei. This function (or its approximation) is considered as the best candidate to be used as the localized basis function for atomic orbitals basis sets. Another direction is related to the construction of the accurate low-rank global interpolant for big scattered data to be considered in the next section. In this way, we calculate the data adaptive basis set living on the fine Cartesian grid in the region of target data. The main challenge, however, is due to the presence of point singularities which are hard to approximate in the problem independent polynomial or trigonometric basis sets.

The construction of low-rank RS approximation to the Slater function is based on the generalized Laplace transform representation for the Slater function written in the form  $G(\rho) = e^{-2\sqrt{\alpha\rho}}$ ,  $\rho(x) = x_1^2 + \dots + x_d^2$ , reads

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

which corresponds to the choice  $\widehat{G}(\tau) = \frac{\sqrt{\alpha}}{\sqrt{\pi}} \tau^{-3/2} e^{-\alpha/\tau}$  in the canonical form of the Laplace transform representation for  $G(\rho)$ ,

$$G(\rho) = \int_{\mathbb{R}_+} \widehat{G}(\tau) e^{-\rho\tau} d\tau. \tag{34}$$

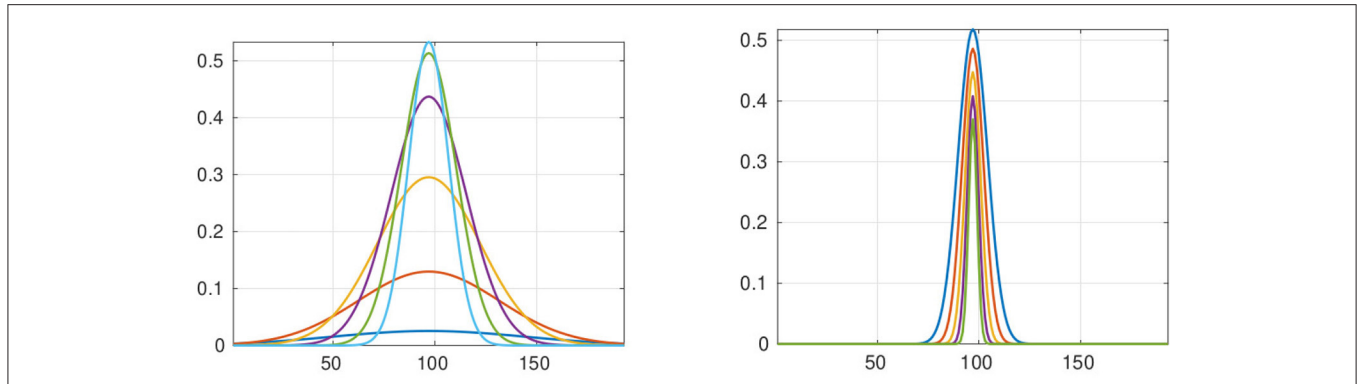
Denote by  $\mathbf{G}_R$  the rank- $R$  canonical approximation to the function  $G(\rho)$  discretized on the  $n \times n \times n$  Cartesian grid.

**Lemma 4.4.** ([51]) *For given threshold  $\varepsilon > 0$  let  $\rho \in [1, A]$ . Then the  $(2M + 1)$ -term sinc-quadrature approximation of the integral in (34) with*

$$M = O(|\log \varepsilon| (|\log \varepsilon| + \log A)),$$

ensures the max-error of the order of  $O(\varepsilon)$  for the corresponding rank- $(2M + 1)$  CP approximation  $\mathbf{G}_R$  to the tensor  $\mathbf{G}$ .

**Figure 5** illustrates the RS splitting for the tensor  $\mathbf{G}_R = \mathbf{G}_{R_l} + \mathbf{G}_{R_s}$  representing the Slater potential  $G(x) = e^{-\lambda \|x\|}$ ,  $\lambda = 1$ , discretized on the  $n \times n \times n$  grid with  $n = 1024$ . The rank parameters are chosen by  $R = 24, R_l = 6$  and  $R_s = 18$ . Notice that for this radial function the long-range part (**Figure 5**, left) includes much less canonical vectors comparing with the case of Newton kernel. This anticipates the smaller total canonical rank



**FIGURE 5** | Long-range (left) and short-range (right, a base 10 logarithmic scale) canonical vectors for the Slater function with the grid size  $n = 1024$ ,  $R = 24, R_l = 6, \lambda = 1$ .

for the long-range part in the large sum of Slater-like potentials arising, for example, in the representation of molecular orbitals and the electron density in electronic structure calculations. For instance, the wave function for the Hydrogen atom is given by the Slater function  $e^{-\mu\|x\|}$ . In the following section, we consider the application of RS tensor format to interpolation of scattered data in  $\mathbb{R}^d$ .

### 4.5. Application of RHOSVD to Scattered Data Modeling

In scattered data modeling the problem is in a low parametric approximation of multi-variate functions  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  by sampling at a finite set  $\mathcal{X} = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$  of piecewise distinct points. Here, the function  $f$  might be the surface of a solid body, the solution of a PDE, many-body potential field, multi-parametric characteristics of physical systems, or some other multi-dimensional data, etc.

Traditional ways of recovering  $f$  from a sampling vector  $f_{|\mathcal{X}} = (f(x_1), \dots, f(x_N))$  is the constructing a functional interpolant  $P_N: \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $P_N|_{\mathcal{X}} = f_{|\mathcal{X}} =: \mathbf{f} \in \mathbb{R}^N$ , i.e.,

$$P_N(x_j) = f(x_j), \quad \forall 1 \leq j \leq N. \tag{35}$$

Using radial basis (RB) functions one can find interpolants  $P_N$  in the form

$$P_N(x) = \sum_{j=1}^N c_j p(\|x - x_j\|) + Q(x),$$

$Q$  is some smooth function, say, polynomial, (36)

where  $p = p(r): [0, \infty) \rightarrow \mathbb{R}$  is a fixed RB function, and  $r = \|\cdot\|$  is the Euclidean norm on  $\mathbb{R}^d$ . In further discussion, we set  $Q(x) = 0$ . For example, the following RB functions are commonly used

$$p = r^\nu, \quad (1+r^2)^\nu, \quad (\nu \in \mathbb{R}), \quad \exp(-r^2), \quad \exp(-\lambda r), \quad r^2 \log(r).$$

The other examples of RB functions are defined by Green's kernels or by the class of Matérn functions [23].

We discuss the following computational tasks (A) and (B).

- (A) For a fixed coefficient vector  $\mathbf{c} = (c_1, \dots, c_N)^T \in \mathbb{R}^N$ , efficiently representing the interpolant  $P_N(x)$  on the fine tensor grid in  $\mathbb{R}^d$  providing
  - (a)  $O(1)$ -fast point evaluation of  $P_N$  in the computational volume  $\Omega$ ,
  - (b) computation of various integral-differential operations on that interpolant (say, gradients, scalar products, convolution integrals, etc.)
- (B) Finding the coefficient vector  $\mathbf{c}$  that solves the interpolation problem Equation (35) in the case of large number  $N$ .

Problem (A) exactly fits the RS tensor framework so that the RS tensor approximation solves the problem with low computational costs provided that the sum of long-range parts of the interpolating functions can be easily approximated in the low rank CP tensor format. We consider the case of interpolation by Slater functions  $\exp(-\lambda r)$  in the more detail.

Problem (B): Suppose that we use some favorable preconditioned iteration for solving coefficient vector  $\mathbf{c} = (c_1, \dots, c_N)^T$ ,

$$A_{p,\mathcal{X}} \mathbf{c} = \mathbf{f}, \quad \text{with } A_{p,\mathcal{X}} = A_{p,\mathcal{X}}^T = [p(\|x_i - x_j\|)]_{1 \leq i,j \leq N} \in \mathbb{R}^{N \times N}, \tag{37}$$

with the distance dependent symmetric system matrix  $A_{p,\mathcal{X}}$ . We assume  $\mathcal{X} = \Omega_h$  be the  $n^{\otimes d}$ -set of grid-points located on tensor grid, i.e.,  $N = n^d$ . Introduce the  $d$ -tuple multi-index  $i \mapsto \mathbf{i} = (i_1, \dots, i_d)$ , and  $j \mapsto \mathbf{j} = (j_1, \dots, j_d)$  and reshape  $A_{p,\mathcal{X}}$  into the tensor form

$$A_{p,\mathcal{X}} \mapsto \mathbf{A} = [a(i_1, j_1, \dots, i_d, j_d)] \in \bigotimes_{\ell=1}^d \mathbb{R}^{n \times n},$$

which can be decomposed by using the RS based splitting

$$\mathbf{A} = \mathbf{A}_{R_s} + \mathbf{A}_{R_l},$$

generated by the RS representation of the weighted potential sum in Equation (36). Here,  $\mathbf{A}_{R_s}$  is a banded diagonal matrix with dominating diagonal part, while  $\mathbf{A}_{R_l} = \sum_{k=1}^{R_l} A_k^{(1)} \otimes$

$\dots \otimes A_k^{(d)}$  is the low Kronecker rank matrix. This implies a bound on the storage,  $O(N + dR_l n)$ , and ensures a fast matrix-vector multiplication. Introducing the additional rank-structured representation in  $\mathbf{c}$ , the solution of Equation (37) can be further simplified.

The above approach can be applied to the data sparse representation for the class of large covariance matrices in the spatial statistics, see for example [23, 55].

In application of tensor methods to data modeling (see Section 4.5) we consider the interpolation of 3D scattered data by a large sum of Slater functions

$$G_N(x) = \sum_{j=1}^N c_j e^{-\lambda \|x-x_j\|}, \quad \lambda > 0. \quad (38)$$

Given the coefficients  $c_j$ , we address the question how to efficiently represent the interpolant  $G_N(x)$  on fine Cartesian grid in  $\mathbb{R}^3$  by using the low-rank (i.e., low-parametric) CP tensor format, such that each value on the grid can be calculated in  $O(1)$  operations. The main problem is that the generating Slater function  $e^{-\lambda \|x\|}$  has the cusp at the origin so that the considered interpolant has very low regularity. As result, the tensor rank of the function  $G_N(x)$  in Equation (38) discretized on a large  $n \times n \times n$  grid increases almost proportionally to the number  $N$  of sampling points  $x_j$ , which in general may be very large. This increase in the canonical rank has been observed in a number of numerical tests. Hence, the straightforward tensor approximation of  $G_N(x)$  does not work in this case.

Tables 2, 3 illustrate the stability of the canonical rank in the number  $N$  of sampling points in the case of random and function related distribution of the waiting coefficients  $c_j$  in the long-range part of the Slater interpolant Equation (38).

The generating Slater radial function can be proven to have the low-rank RS canonical tensor decomposition by using the sinc-approximation method (see section 4.1).

To complete this section, we present the numerical example demonstrating the application of RS tensor representation to scattered data modeling in  $\mathbb{R}^3$ . We denote by  $\mathbf{G}_R \in \mathbb{R}^{n \otimes 3}$  the rank- $R$  CP tensor approximation of the reference Slater potential  $e^{-\lambda \|x\|}$  discretized on  $n \times n \times n$  grid  $\Omega_n$ , and introduce its RS splitting  $\mathbf{G}_R = \mathbf{G}_{R_l} + \mathbf{G}_{R_s}$ , with  $R_l + R_s = R$ . Here,  $R_l \approx R/2$  is the rank parameter of the long-range part in  $\mathbf{G}_R$ . Assume that all measurement points  $x_j$  in Equation (38) are located on the discretization grid  $\Omega_n$ , then the tensor representation of the long-range part of the total interpolant  $P_N$  can be obtained as the sum of the properly replicated reference potential  $\mathbf{G}_l$ , via the shift-and-windowing transform  $\mathcal{W}_j, j = 1, \dots, N$ ,

$$\mathbf{G}_{N,l} = \sum_{j=1}^N c_j \mathbf{G}_{l,j}, \quad \mathbf{G}_{l,j} = \mathcal{W}_j \mathbf{G}_l, \quad (39)$$

that includes about  $N R_l$  terms. For large number of measurement points,  $N$ , the rank reduction is ubiquitous.

It can be proven (by slight modification of arguments in [32]) that both the CP and Tucker ranks of the  $N$ -term sum in Equation (39) depend only logarithmically (but not linearly) on  $N$ .

TABLE 2 | Reduced ranks for the case of random amplitudes.

$L_1 \times L_2 \times L_3$	$N$	Tucker ranks	$R_{ini}$	$R_{comp}$
$4 \times 4 \times 4$	64	$13 \times 13 \times 13$	192	56
$6 \times 6 \times 6$	216	$15 \times 15 \times 15$	649	95
$8 \times 8 \times 8$	512	$19 \times 19 \times 19$	1536	131
$16 \times 16 \times 8$	2048	$32 \times 32 \times 19$	6141	253
$16 \times 16 \times 16$	4096	$32 \times 32 \times 32$	12288	380

$$\varepsilon_{Tuck} = 10^{-3}, \varepsilon_{T2C} = 10^{-5}, R_L = 3, R = 5.$$

TABLE 3 | Reduced scanonical ranks for the case of functional amplitudes.

$L_1 \times L_2 \times L_3$	$N$	Tucker ranks	$R_{ini}$	$R_{comp}$
$4 \times 4 \times 4$	64	$7 \times 7 \times 7$	256	25
$6 \times 6 \times 6$	216	$7 \times 7 \times 7$	648	23
$8 \times 8 \times 8$	512	$7 \times 7 \times 7$	1536	30
$16 \times 16 \times 8$	2048	$10 \times 9 \times 6$	6144	47
$16 \times 16 \times 16$	4096	$10 \times 9 \times 8$	12288	55

$$R_L = 3, R_L = 5, \varepsilon_N = 10^{-5}, \varepsilon_{Tuck} = 10^{-3}, \varepsilon_{T2C} = 10^{-5}.$$

**Proposition 4.5.** (Uniform rank bounds for the long-range part in the Slater interpolant). Let the long-range part  $\mathbf{G}_{N,l}$  in the total Slater interpolant in Equation (39) be composed of those terms in Equation (24) which satisfy the relation  $t_k \leq 1$ , where  $M = O(\log^2 \varepsilon)$ . Then the total  $\varepsilon$ -rank  $\mathbf{r}_0$  of the Tucker approximation to the canonical tensor sum  $\mathbf{G}_{N,l}$  is bounded by

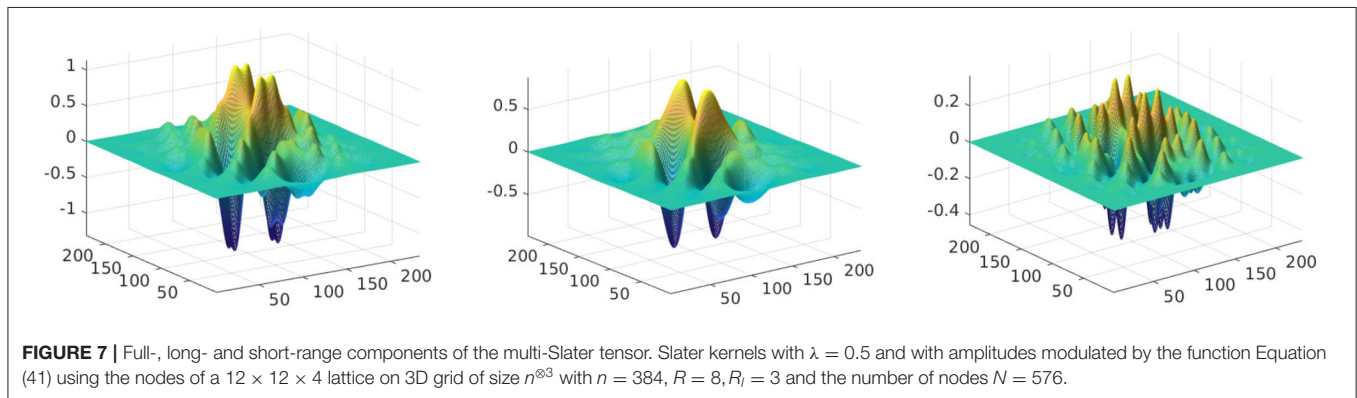
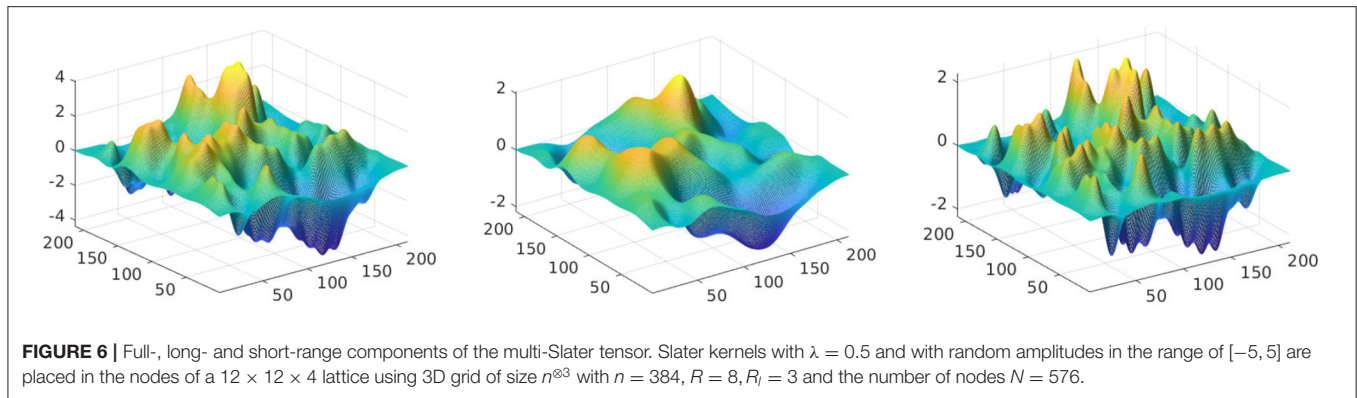
$$|\mathbf{r}_0| := \text{rank}_{Tuck}(\mathbf{G}_{N,l}) = C b \log^{3/2}(\log(N/\varepsilon)), \quad (40)$$

where the constant  $C$  does not depend on the number of particles  $N$ , as well as on the size of the computational box,  $[-b, b]^3$ .

*Proof:* (Sketch) The main argument of the proof is based on the fact that the grid function  $\mathbf{G}_{N,l}$  has the band-limited Fourier image, such that the frequency interval depends weakly (logarithmically) on  $N$ . Then we represent all Gaussians in the truncated Fourier basis and make the summation in the fixed set of orthogonal trigonometric basis functions, which defines the orthogonal Tucker representation with controllable rank parameter.  $\square$

The numerical illustrations below demonstrate the CP rank by RHOSVD decomposition of the long-range part  $\mathbf{G}_{N,l}$  in the multi-point tensor interpolant via Slater functions.

Now, we generate a tensor composed of a sum of Slater functions, discretized by collocation over  $n^{\otimes 3}$  representation grid with  $n = 384$ , and placed in the nodes of a sampling  $L_1 \times L_2 \times L_3$  lattice with randomly chosen weights  $c_j$  in the interval  $c_j \in [-5, 5]$  for every node. Every single Slater function is generated as a canonical tensor by using sinc-quadratures for the approximation of the related Laplace transform. Table 2 shows ranks of the long-range part of this tensor composed of Slater potentials located in the nodes of the lattices of increasing size.  $N$  indicates the number of nodes, while  $R_{ini}$  and  $R_{comp}$



are the initial and compressed canonical ranks of the resulting long-range part tensor, respectively. Tucker ranks correspond to the ranks in the canonical-to-Tucker decomposition step. Threshold values for the Slater potential generator is  $\varepsilon_N = 10^{-5}$ , while the tolerance thresholds for the rank reduction procedure are given by  $\varepsilon_{Tuck} = 10^{-3}$  and  $\varepsilon_{T2C} = 10^{-5}$ . We observe that the ranks of the long-range part of the potential increase only slightly in the size of the 3D sampling lattice,  $N$ .

**Figure 6** demonstrates the full-, short-, and long-range components of the multi-Slater tensor constructed by the weighted sum of Slater functions with *randomly* chosen weights  $c_j$  in the interval  $c_j \in [-5, 5]$ . The positions of the generating nodes are located on the  $12 \times 12 \times 4$  3D lattice. The parameters of the tensor interpolant are set up as follows:  $\lambda = 0.5$ , the representation grid is of size  $n^{\otimes 3}$  with  $n = 384$ ,  $R = 8$ ,  $R_l = 3$  and the number of samples  $N = 576$  (Figures zoom a part of the grid.). The initial CP rank of the sum of  $N_0$  interpolating Slater potentials is about 4,468. Middle and right pictures show the long- and short-range parts of the composite tensor, respectively. The initial rank of the canonical tensor representing the long-range part is equal to  $R_L = 2304$ , which is reduced by the C2C procedure *via* RHOSVD to  $R_{cc} = 71$ . The rank truncation threshold is  $\varepsilon = 10^{-3}$ .

**Figure 7** and **Table 3** demonstrate the decomposition of the multi-Slater tensor with the amplitudes  $c_j$  in the nodes  $(x_j, y_j, z_j)$

modulated by the function of the  $(x, y, z)$ -coordinates

$$F(x, y, z) = a_1 \cos(x + 2y + 4z) \exp(-a_2 \sqrt{x^2 + 2y^2 + 4z^2}), \quad (41)$$

with  $a_1 = 6$  and  $a_2 = 0.1$ , i.e.,  $c_j = F(x_j, y_j, z_j)$ .

Next, we generate a tensor composed of a sum of discretized Slater functions on a sampling lattice  $L_1 \times L_2 \times L_3$ , living on 3D representation grid of size  $n^{\otimes 3}$  with  $n = 232$ . The amplitudes of the individual Slater functions are modulated by a function of  $x, y, z$ -coordinates Equation (41) in every node of the lattice. **Table 3** shows rank of the long-range part of this multi-Slater tensor with respect to the increasing size of the lattice.  $N = L_1 L_2 L_3$  is the number of nodes, and  $R_{ini}$  and  $R_{comp}$  are the initial and compressed canonical ranks, respectively. Tucker ranks are shown at the canonical-to-Tucker decomposition step. Threshold values for the Slater potential generation is  $\varepsilon_N = 10^{-5}$ , the thresholds for the canonical-to-canonical rank reduction procedure are given by  $\varepsilon_{Tuck} = 10^{-3}$  and  $\varepsilon_{T2C} = 10^{-5}$ . **Table 3** demonstrates the very moderate increase of the reduced rank in the long-range part of the Slater potential sum on the size of the 3D sampling lattice.

**Figure 7** demonstrates the full-, long-, and short-range components of the multi-Slater tensor. Slater kernels with  $\lambda = 0.5$  and with the amplitudes modulated by the function Equation (41) of the  $(x, y, z)$ -coordinates are places on the nodes of a  $12 \times 12 \times 4$  sampling lattice, living on 3D grid of size  $n^{\otimes 3}$  with

$n = 384, R = 8, R_l = 3$ , and with the number of sampling nodes  $N = 576$ .

### 5. REPRESENTING GREEN'S KERNELS IN TENSOR FORMAT

In this section, we demonstrate how the RHOSVD can be applied for the efficient tensor decomposition of various singular radial functions composed by polynomial expansions of a few reference potentials already precomputed in the low-rank tensor format. Given the low-rank CP tensor  $\mathbf{A}$  further considered as a reference tensor, the low rank representation of the tensor-valued polynomial function

$$P(\mathbf{A}) = a_0\mathbf{I} + a_1\mathbf{A} + a_2\mathbf{A}^2 + \dots + a_n\mathbf{A}^n,$$

where the multiplication of tensors is understood in the sense of pointwise Hadamard product, can be calculated *via*  $n$ -times application of the RHOSVD by using the Horner scheme in the form

$$P(x) = a_0 + x(a_1 + x(a_2 + x(a_3 + \dots x(a_{n-1} + a_n x) \dots))).$$

Similar scheme can be also applied in the case of multivariate polynomials.

For examples considered, in this section, we make use of the discretized Slater  $e^{-\|x\|}$  and Newton  $\frac{1}{\|x\|}$ ,  $x \in \mathbb{R}^d$ , kernels as the reference tensors. The following statement was proven in [11, 51] (see also Lemma 4.4).

**Proposition 5.1.** *The discretized over  $n^{\otimes d}$ -grid radial functions  $e^{-\|x\|}$  and  $\frac{1}{\|x\|}$ ,  $x \in \mathbb{R}^d$ , included in representation of various Green kernels and fundamental solutions for elliptic operators with constant coefficients, both allow the low-rank CP tensor approximation. The corresponding rank-R representations can be calculated in  $O(dRn)$  operations without precomputing and storage of the target tensor in the full (entry-wise) format.*

Tensor decomposition for discretized singular kernels such as  $\|x\|$ ,  $\frac{1}{\|x\|^m}$ ,  $m \geq 2$ , and  $e^{-\kappa\|x\|}/\|x\|$ , can be now calculated by applying the RHOSVD to polynomial combinations of the reference potentials as in Proposition 5.1. The most important benefit of the presented techniques is the opportunity to compute the rank- $R$  tensor approximations without pre-computing and storage of the target tensor in the full format tensor.

In what follows, we present the particular examples of singular kernels in  $\mathbb{R}^d$  which can be treated by the above presented techniques. Consider the fundamental solution of the advection-diffusion operator  $\mathcal{L}_d$  with constant coefficients in  $\mathbb{R}^d$

$$\mathcal{L}_d = -\Delta + 2\bar{b} \cdot \nabla + \kappa^2, \quad \bar{b} \in \mathbb{C}^d, \quad \kappa \in \mathbb{C}.$$

If  $\kappa^2 + |\bar{b}|^2 = 0$ , then for  $d \geq 3$  it holds

$$\eta_0(x) = \frac{1}{(d-2)\omega_d} \frac{e^{(\bar{b},x)}}{\|x\|^{d-2}}, \quad \|x\| = \frac{\|x\|^2}{\|x\|},$$

where  $\omega_d$  is the surface area of the unit sphere in  $\mathbb{R}^d$ , [56–58]. Notice that the radial function  $\frac{1}{\|x\|^{d-2}}$  for  $d \geq 3$  allows the RS decomposition of the corresponding discrete tensor representation based on the sinc quadrature approximation, which implies the RS representation of the kernel function  $\eta_0(x)$ , since the function  $e^{(\bar{b},x)}$  is already separable. From computational point of view, both the CP and RS canonical decompositions of discretized kernels  $\frac{1}{\|x\|^{d-2}}$  can be computed by successive application of RHOSVD approximation to the products of canonical tensors for the discretized Newton potential  $\frac{1}{\|x\|}$ .

In the particular case  $\bar{b} = 0$ , we obtain the fundamental solution of the operator  $\mathcal{L}_3 = -\Delta + \kappa^2$  for  $d = 3$ , also known as the Yukawa (for  $\kappa \in \mathbb{R}_+$ ) or Helmholtz (for  $\kappa \in \mathbb{C}$ ) Green kernels

$$\eta_\lambda(x) = \frac{1}{4\pi} e^{-\kappa\|x\|}/\|x\|, \quad x \in \mathbb{R}^3.$$

In the case of Yukawa kernel the tensor representations by using Gaussian sums are considered in [1, 2], see also references therein.

The Helmholtz equation with  $\text{Im } \kappa > 0$  (corresponds to the diffraction potentials) arises in problems of acoustics, electro-magnetics and optics. We refer to [59] for the detailed discussion of this class of fundamental solutions. Fast algorithms for the oscillating Helmholtz kernel have been considered in [1]. However, in this case the construction of the RS tensor decomposition remains an open question.

In the case of 3D biharmonic operator  $\mathcal{L} = \Delta^2$  the fundamental solution reads as

$$p(\|x\|) = -\frac{1}{8\pi} \|x\|, \quad x \in \mathbb{R}^3.$$

The hydrodynamic potentials correspond to the classical Stokes operator

$$\nu \Delta u - \text{grad } p = f, \quad \text{div } u = 0,$$

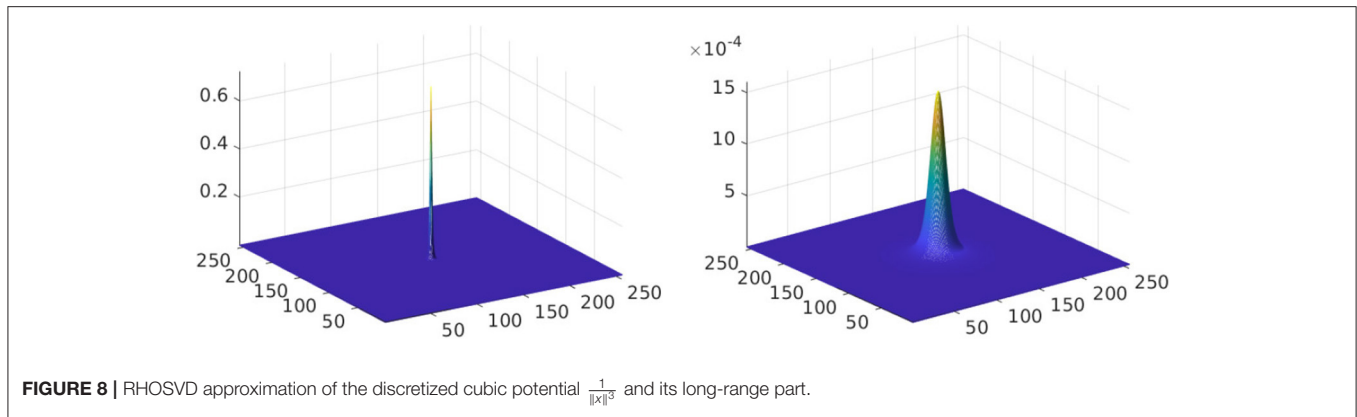
where  $u$  is the velocity field,  $p$  denotes the pressure, and  $\nu$  is the constant viscosity coefficient. The solution of the Stokes problem in  $\mathbb{R}^3$  can be expressed by the hydrodynamic potentials

$$u_k(x) = \int_{\mathbb{R}^3} \sum_{\ell=1}^3 \Psi_{k\ell}(x-y) f_\ell(y) dy, \quad p(x) = \int_{\mathbb{R}^3} \langle \Theta(x-y), f \rangle(y) dy \tag{42}$$

with the fundamental solution

$$\Psi_{k\ell}(x) = \frac{1}{8\pi\nu} \left( \frac{\delta_{k\ell}}{\|x\|} + \frac{x_k x_\ell}{\|x\|^3} \right), \quad \Theta(x) = \frac{x}{4\pi\|x\|^3}, \quad x \in \mathbb{R}^3. \tag{43}$$

The existence of the low-rank RS tensor representation for the hydrodynamic potential is based on the same argument as in Remark 5.1. In turn, in the case of biharmonic fundamental solution we use the identity



**FIGURE 8** | RHOSVD approximation of the discretized cubic potential  $\frac{1}{\|x\|^3}$  and its long-range part.

where the nominator has the separation rank equals to  $d$ . The latter representation can be also applied for calculation of the respective tensor approximations.

Here, we demonstrate how the application of RHOSVD allows to easily compute the low rank Tucker/CP approximation of the discretized singular potential  $\frac{1}{\|x\|^3}$ ,  $x \in \mathbb{R}^3$ , as well as the respective RS-representation, having at hand the RS representation of the tensor  $\mathbf{P} \in \mathbb{R}^{n \times n \times n}$  discretizing the Newton kernel. In this example, we use the discretization of  $\frac{1}{\|x\|^3}$  in the form

$$\mathbf{P}^{(3)} = \mathbf{P} \odot \mathbf{P} \odot \mathbf{P},$$

where by  $\mathbf{P}^{(3)}$  we denotes the collocation projection discretization of  $\frac{1}{\|x\|^3}$ . The low rank Tucker/CP tensor approximation to  $\mathbf{P}^{(3)}$  can be computed by the direct application of the RHOSVD to the above product type representation. The RS representation of  $\mathbf{P}^{(3)}$  is calculated based on Lemma 4.3. Given the RS-representation Equation (31) of the discretized Newton kernel,  $\mathbf{P}_R$ , we define the low rank CP approximation to the discretized singular part in the hydrodynamic potential  $\mathbf{P}^{(3)}$  by

$$\mathbf{P}_{R'}^{(3)} = \mathbf{P}_R \odot \mathbf{P}_R \odot \mathbf{P}_R.$$

In view of Lemma 4.3, the long range part of RS decomposition of  $\mathbf{P}_{R'}^{(3)}$ , can be computed by RHOSVD approximation to the following Hadamard product of tensors,

$$\mathbf{P}_{R'_l}^{(3)} = \mathbf{P}_{R_l} \odot \mathbf{P}_{R_l} \odot \mathbf{P}_{R_l}.$$

**Figure 8** visualizes the tensor  $\mathbf{P}_{R'}^{(3)}$  as well as its long range part  $\mathbf{P}_{R'_l}^{(3)}$ .

The potentials are discretized on  $n \times n \times n$  Cartesian grid with  $n = 257$ , the rank truncation threshold is chosen for  $\varepsilon = 10^{-5}$ . The CP rank of the Newton kernel is equal to  $R = 19$ , while we set  $R_l = 10$ , thus resulting in the initial ranks 6859 and  $10^3$  for RHOSVD decomposition of  $\mathbf{P}_{R'}^{(3)}$  and  $\mathbf{P}_{R'_l}^{(3)}$ , respectively. The RHOSVD decomposition reduces the large rank parameters to  $R' = 122$  (the Tucker rank is  $r = 13$ ) and  $R'_l = 58$  (the Tucker rank is  $r = 8$ ), correspondingly.

## 6. RHOSVD FOR RANK REDUCTION IN 3D ELLIPTIC PROBLEM SOLVERS

Efficient rank reduction procedure based on the RHOSVD is a prerequisite for the development of the tensor-structured solvers for the three-dimensional elliptic problem, which reduce the computational complexity to almost linear scale,  $O(nR)$ , contrary to usual  $O(n^3)$  complexity.

Assume that all input data in the governing PDE are given in the low-rank tensor form. The convenient tensor format for these problems is a canonical tensor representation of both the governing operator, and of the initial guess as well as of the right hand side. The commonly used numerical techniques are based on certain iterative schemes that include at each iterative step multiple matrix-vector and vector-vector algebraic operations each of them enlarges the tensor rank of the output in the additive or multiplicative way. It turns out that in common practice the most computationally intensive step in the rank-structured algorithms is the adaptive rank truncation, which makes the rank truncation procedure ubiquitous.

We notice that in PDE based mathematical models the total numerical complexity of the particular computational scheme, i.e., the overall cost of the rank truncation procedure is determined by the multiple of the number of calls to the rank truncation algorithm (merely the number of iterations) and the cost of a single RHOSVD transform (mainly determined by the rank parameter of the input tensor). In turn, both complexity characteristics depend on the quality of the rank-structured preconditioner so that optimization of the whole solution process is can be achieved by the trade-off between Kronecker rank of the preconditioner and the complexity of its implementation.

In the course of preconditioned iterations, the tensor ranks of the governing operator, the preconditioner and the iterand are multiplied, and, therefore, a robust rank reduction is mandatory procedure for such techniques applied to iterative solution of elliptic and pseudo-differential equations in the rank-structured tensor format.

In particular, the RHOSVD was applied to the numerical solution of PDE constrained (including the case of fractional operators) optimal control problems [36, 39], where the complexity of the order  $O(nR \log n)$  was demonstrated.



In the case of higher dimensions the rank reduction in the canonical format can be performed directly (i.e., without intermediate use of the Tucker approximation) by using the cascading ALS iteration in the CP format, see [50] concerning the tensor-structured solution of the stochastic/parametric PDEs.

## 7. CONCLUSIONS

We discuss theoretical and computational aspects of the RHOSVD served for approximation of tensors in low-rank Tucker/canonical formats, and show that this rank reduction technique is the principal ingredient in tensor-based computations for real-life problems in scientific computing and data modeling. We recall rank reduction scheme for the canonical input tensors based on RHOSVD and subsequent Tucker-to-canonical transform. We present the detailed error analysis of low rank RHOSVD approximation to the canonical tensors (possibly with large input rank), and provide the proof on the uniform bound for the relative approximation error.

We recall that the first example on application of the RHOSVD was the rank-structured computation of the 3D convolution transform with the nonlocal Newton kernel in  $\mathbb{R}^3$ , which is the basic operation in the Hartree-Fock calculations.

The RHOSVD is the basic tools for utilizing the multilinear algebra in RS tensor format, which employs the sinc-analytic tensor approximation methods applied to the important class of radial functions in  $\mathbb{R}^d$ . This enables efficient rank decompositions of tensors generated by functions with multiple local cusps or singularities by separating their short- and long-range parts. As an example, we construct the RS tensor representation of the discretized Slater function  $e^{-\lambda\|x\|}$ ,  $x \in \mathbb{R}^d$ . We then describe the RS tensor approximation to various Green's kernels obtained by combination of this function with other potentials, in particular,

with the Newton kernel providing the Yukawa potential. In this way, we introduce the concept of reproducing radial functions which pave the way for efficient RS tensor decomposition applied to a wide range of function-related multidimensional data by combining the multilinear algebra in RS tensor format with the RHOSVD rank reduction techniques.

Our next example is related to application of RHOSVD to low-rank tensor interpolation of scattered data. Our numerical tests demonstrate the efficiency of this approach on the example of multi-Slater interpolant in the case of many measurement points. We apply the RHOSVD to the data generated *via* random or function modulated amplitudes of samples and demonstrate numerically that for both cases the rank of the long-range part remains small and depends weakly on the number of samples.

Finally, we notice that the described RHOSVD algorithms have proven their efficiency in a number of recent applications, in particular, in rank reduction for the tensor-structured iterative solvers for PDE constraint optimal control problems (including fractional control), in construction of the range-separated tensor representations for calculation of the electrostatic potentials of many-particle systems (arising in protein modeling), and for numerical analysis of large scattered data in  $\mathbb{R}^d$ .

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

BK: mathematical theory, basic concepts, and manuscript preparation. VK: basic concepts, numerical simulations, and manuscript preparation. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Khoromskij BN. *Tensor Numerical Methods in Scientific Computing*. Berlin: De Gruyter Verlag (2018).
- Khoromskaia V, and Khoromskij BN. *Tensor Numerical Methods in Quantum Chemistry*. Berlin: De Gruyter Verlag (2018).
- P. Comon. "Tensor decompositions," in *Mathematics in Signal Processing V* (2002) 1–24.
- Comon P, Luciani X, and De Almeida ALF. Tensor decompositions, alternating least squares and other tales. *J Chemometr J Chemometr Soc.* (2009) 23:393–405. doi: 10.1002/CEM.1236
- Smilde A, Bro R, and Geladi P. *Multi-Way Analysis With Applications in the Chemical Sciences*. Wiley (2004).
- Cichocki A, Lee N, Oseledets I, Pan AH, Zhao Q, and Mandic DP. Tensor networks for dimensionality reduction and large-scale optimization: part 1 low-rank tensor decompositions. *Found Trends Mach Learn* (2016) 9:249–429. doi: 10.1561/22000000059
- De Lathauwer L, De Moor B, and Vandewalle J. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* (2000) 21:1253–78. doi: 10.1137/S0895479896305696
- Ten Berge JMF, and Sidiropoulos ND. On uniqueness in CANDECOMP/PARAFAC. *Psychometrika* (2002) 67:399–409.
- Sidiropoulos ND, De Lathauwer L, Fu X, Huang K, Papalexakis EE, Faloutsos C. Tensor decomposition for signal processing and machine learning. *IEEE Trans Signal Process.* (2017) 65:3551–82. doi: 10.1109/TSP.2017.2690524
- Golub GH, and Van Loan F. *Matrix Computations*. Baltimore, MD: Johns Hopkins University Press (1996).
- Hackbusch W, and Khoromskij BN. Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part I. Separable approximation of multi-variate functions. *Computing.* (2006) 76:177–202. doi: 10.1007/s00607-005-0144-0
- Khoromskij BN, and Khoromskaia V. Low rank tucker-type tensor approximation to classical potentials. *Central Eur J Math.* (2007) 5:523–50. doi: 10.2478/s11533-007-0018-0
- Marcati C, Rakhuba M, and Schwab C. Tensor rank bounds for point singularities in  $\mathbb{R}^3$ . *E-preprint arXiv:1912.07996*, (2019).
- Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. *J Math Phys.* (1927) 6:164–89.
- Harshman RA. "Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis," In: *UCLA Working Papers Phonetics*. vol. 16 (1970). 1–84.
- Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* (1966) 31:279–311.
- Oseledets IV, and Tyrtshnikov EE, Breaking the curse of dimensionality, or how to use svd in many dimensions. *SIAM J Sci Comput.* (2009) 31:3744–59. doi: 10.1137/090748330
- Oseledets IV. Tensor-train decomposition. *SIAM J Sci Comput.* (2011) 33:2295–317. doi: 10.1137/090752286

19. Hackbusch W, and Kühn S. A new scheme for the tensor representation. *J. Fourier Anal. Appl.* (2009) 15:706–22. doi: 10.1007/s00041-009-9094-9
20. Khoromskij BN.  $O(d \log N)$ -quantics approximation of  $N$ - $d$  tensors in high-dimensional numerical modeling. *Construct Approx.* (2011) 34:257–89. doi: 10.1007/S00365-011-9131-1
21. Oseledets I. Constructive representation of functions in low-rank tensor formats. *Constr. Approx.* (2013) 37:1–18. doi: 10.1007/s00365-012-9175-x
22. Kressner D, Steinlechner M, and Uschmajew A. Low-rank tensor methods with subspace correction for symmetric eigenvalue problems. *SIAM J Sci Comput.* (2014) 36:A2346–68. doi: 10.1137/130949919
23. Litvinenko A, Keyes D, Khoromskaia V, Khoromskij BN, and Matthies HG. Tucker tensor analysis of Matern functions in spatial statistics. *Comput. Meth. Appl. Math.* (2019) 19:101–22. doi: 10.1515/cmam-2018-0022
24. Rakhuba M, and Oseledets I. Fast multidimensional convolution in low-rank tensor formats via cross approximation. *SIAM J Sci Comput* (2015) 37:A565–82. doi: 10.1137/140958529
25. Uschmajew A. Local convergence of the alternating least squares algorithm for canonical tensor approximation. *SIAM J Mat Anal Appl* (2012) 33:639–52. doi: 10.1137/110843587
26. Hackbusch W and Uschmajew A. Modified iterations for data-sparse solution of linear systems. *Vietnam J Math.* (2021) 49:493–512. doi: 10.1007/s10013-021-00504-9
27. V. Khoromskaia. *Numerical Solution of the Hartree-Fock Equation by Multilevel Tensor-structured methods*. Berlin: TU Berlin (2010).
28. Khoromskaia V, and Khoromskij BN. Grid-based lattice summation of electrostatic potentials by assembled rank-structured tensor approximation. *Comp. Phys. Commun.* (2014) 185:3162–74. doi: 10.1016/j.cpc.2014.08.015
29. Khoromskij BN, Khoromskaia V, and Flad H-J. Numerical solution of the hartree-fock equation in multilevel tensor-structured format. *SIAM J. Sci. Comput.* (2011) 33:45–65. doi: 10.1137/090777372
30. Dolgov SV, Khoromskij BN, and Oseledets I. Fast solution of multi-dimensional parabolic problems in the TT/QTT formats with initial application to the Fokker-Planck equation. *SIAM J. Sci. Comput.* (2012) 34:A3016–38. doi: 10.1137/120864210
31. Kazeev, M. Khammash, M. Nip, and Ch. Schwab. Direct solution of the chemical master equation using quantized tensor trains. *PLoS Comput Biol.* (2014) 10:e1003359. doi: 10.1371/journal.pcbi.1003359
32. Benner P, Khoromskaia V, and Khoromskij BN. Range-separated tensor format for many-particle modeling. *SIAM J Sci Comput.* (2018) 40:A1034-062. doi: 10.1137/16M1098930
33. Benner P, Khoromskaia V, Khoromskij BN, Kweyu C, and Stein M. Regularization of Poisson–Boltzmann type equations with singular source terms using the range-separated tensor format. *SIAM J Sci Comput.* (2021) 43:A415-45. doi: 10.1137/19M1281435
34. Khoromskij BN. Range-separated tensor decomposition of the discretized Dirac delta and elliptic operator inverse. *J Comput Phys.* (2020) 401:108998. doi: 10.1016/j.jcp.2019.108998
35. Kweyu C, Khoromskaia V, Khoromskij B, Stein M, and Benner P. Solution decomposition for the nonlinear Poisson–Boltzmann equation using the range-separated tensor format. *arXiv preprint arXiv:2109.14073.* (2021).
36. Heidel G, Khoromskaia V, Khoromskij BN, and Schulz V. Tensor product method for fast solution of optimal control problems with fractional multidimensional Laplacian in constraints. *J Comput Phys.* (2021) 424:109865. doi: 10.1016/j.jcp.2020.109865
37. Dolgov S, Kalise D, and Kunisch KK. Tensor decomposition methods for high-dimensional Hamilton–Jacobi–Bellman equations. *SIAM J. Sci. Comput.* (2021) 43:A1625–50. doi: 10.1137/19M1305136
38. Dolgov S, and Pearson JW. Preconditioners and tensor product solvers for optimal control problems from chemotaxis. *SIAM J. Sci. Comput.* (2019) 41:B1228–53. doi: 10.1137/18M1198041
39. Schmitt B, Khoromskij BN, Khoromskaia V, and Schulz V. Tensor method for optimal control problems constrained by fractional three-dimensional elliptic operator with variable coefficients. *Numer. Lin Algeb Appl.* (2021) 1–24:e2404. doi: 10.1002/nla.2404
40. Bachmayr M, Schneider R, and Uschmajew A. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Found Comput Math.* (2016) 16:1423–72. doi: 10.1007/s10208-016-9317-9
41. Boiveau T, Ehlacher V, Ern A, and Nouy A. Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods. *ESAIM Math Model Numer Anal* (2019) 53:635-58. doi: 10.1051/m2an/2018073
42. Espig M, Hackbusch W, Litvinenko A, Matthies HG, Zander E. Post-processing of high-dimensional data. (2019) *E-Preprint arXiv:1906.05669.*
43. Ch. Lubich, T. Rohwedder, R. Schneider and B. Vandereycken. Dynamical approximation of hierarchical Tucker and tensor-train tensors. *SIAM J Matrix Anal. Appl.* (2013) 34:470–94. doi: 10.1137/120885723
44. A. Litvinenko, Y. Marzouk, H. G. Matthies, M. Scavino, A. Spantini. Computing f-divergences and distances of high-dimensional probability density functions–low-rank tensor approximations. *E-preprint arXiv:2111.07164.* (2021).
45. Grasedyck L. Hierarchical singular value decomposition of tensors. *SIAM. J. Matrix Anal. Appl.* (2010) 31:2029. doi: 10.1137/090764189
46. Khoromskij BN, and Khoromskaia V. Multigrid tensor approximation of function related arrays. *SIAM J. Sci. Comput.* (2009) 31:3002–26. doi: 10.1137/080730408
47. Ehlacher V, Grigori L, Lombardi D, and Song H. Adaptive hierarchical subtensor partitioning for tensor compression. *SIAM J. Sci. Comput.* (2021) 43:A139–63. doi: 10.1137/19M128689X
48. Kressner D, and Uschmajew A. On low-rank approximability of solutions to high-dimensional operator equations and eigenvalue problems. *Lin Algeb Appl.* (2016) 493:556–72. doi: 10.1016/J.LAA.2015.12.016
49. Oseledets IV, Rakhuba MV, and Uschmajew A. Alternating least squares as moving subspace correction. *SIAM J Numer Anal.* (2018) 56:3459–79. doi: 10.1137/17M1148712
50. Khoromskij BN, and Schwab C. Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs. *SIAM J. Sci. Comput.* (2011) 33:364–85. doi: 10.1137/100785715
51. Khoromskij BN. Structured rank- $(r_1, \dots, r_d)$  decomposition of function-related tensors in  $\mathbb{R}^d$ . *Comput. Meth. Appl. Math.* (2006) 6:194–220. doi: 10.2478/cmam-2006-0010
52. Stenger F. *Numerical Methods Based on Sinc and Analytic Functions*. Springer-Verlag (1993).
53. Braess D. *Nonlinear Approximation Theory*. Berlin: Springer-Verlag (1986).
54. Khoromskaia V, and Khoromskij BN. Prospects of Tensor-Based Numerical Modeling of the Collective Electrostatics in Many-Particle Systems. *Comput Math Math Phys.* (2021) 61:864–86. doi: 10.1134/S0965542521050110
55. Matérn B. *Spatial Variation, Vol. 36 of Lecture Notes in Statistics*. 2nd Edn. Berlin: Springer-Verlag (1986).
56. Abramowitz M, and Stegun IA. *Handbook of Mathematical Functions*. New York, NY: Dover Publ., (1968).
57. Sauter SA, and Schwab C. *Boundary Element Methods*. Springer (2011).
58. Hsiao GC, and Wendland WL. *Boundary Integral Equations*. Berlin: Springer (2008).
59. Maz'ya V, and Schmidt G. Approximate approximations. *Math Surv Monograph* (2007) 141:349.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Khoromskaia and Khoromskij. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

### 8. APPENDIX: PROOFS OF THEOREM 2.3 AND LEMMA 2.5

Proof of Theorem 2.3.

*Proof:* Using the contracted product representations of  $\mathbf{A} \in \mathcal{C}_{R,n}$  and  $\mathbf{A}_{(r)}^0 \in \mathcal{T}_r$ , and introducing the  $\ell$ -mode residual

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

$$\{\Delta^{(\ell)}\}_v = \sum_{k=r_\ell+1}^n \sigma_{\ell,k} \mathbf{z}_k^{(\ell)} v_{k,v}^\ell, \quad v = 1, \dots, R,$$

with notations  $V_0^{(\ell)} = [\mathbf{v}_1^{(\ell)}, \dots, \mathbf{v}_{r_\ell}^{(\ell)}]^T$ ,  $\mathbf{v}_k^{(\ell)} = \{v_{k,v}^\ell\}_{v=1}^R \in \mathbb{R}^R$ , we arrive at the following expansion for the approximation error in the form

$$\mathbf{A} - \mathbf{A}_{(r)}^0 = \boldsymbol{\xi} \times_1 U^{(1)} \times_2 \dots \times_d U^{(d)}$$

$$- \boldsymbol{\xi} \times_1 W^{(1)} \times_2 \dots \times_d W^{(d)} := \sum_{\ell=1}^d \mathbf{B}_\ell,$$

where

$$\mathbf{B}_\ell = \boldsymbol{\xi} \times_1 U^{(1)} \dots \times_{\ell-1} U^{(\ell-1)} \times_\ell \Delta^{(\ell)}$$

$$\times_{\ell+1} W^{(\ell+1)} \dots \times_d W^{(d)}$$

$$= \sum_{v=1}^R \xi_v \left[ \mathbf{u}_v^{(\ell)} \dots \times_{\ell-1} \mathbf{u}_v^{(\ell-1)} \times_\ell \{\Delta^{(\ell)}\}_v \right.$$

$$\left. \times_{\ell+1} \sum_{k=1}^{r_{\ell+1}} \sigma_{\ell+1,k} \mathbf{z}_k^{(\ell+1)} v_{k,v}^{\ell+1} \dots \times_d \sum_{k=1}^{r_d} \sigma_{d,k} \mathbf{z}_k^{(d)} v_{k,v}^d \right].$$

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

$$\|\mathbf{A} - \mathbf{A}_{(r)}^0\| \leq \sum_{\ell=1}^d \|\mathbf{B}_\ell\|,$$

providing the estimate (in view of  $\|\mathbf{u}_v^{(\ell)}\| = 1, \ell = 1, \dots, d, v = 1, \dots, R$ )

$$\|\mathbf{B}_\ell\| \leq \sum_{v=1}^R |\xi_v| \left( \sum_{k=r_\ell+1}^n \sigma_{\ell,k}^2 (v_{k,v}^\ell)^2 \right)^{1/2}$$

$$\times \left( \sum_{k=1}^{r_{\ell+1}} \sigma_{\ell+1,k}^2 (v_{k,v}^{\ell+1})^2 \right)^{1/2} \dots \left( \sum_{k=1}^{r_d} \sigma_{d,k}^2 (v_{k,v}^d)^2 \right)^{1/2}.$$

Furthermore, since  $U^{(\ell)}$  has normalized columns, i.e.,  $\|\mathbf{u}_v^{(\ell)}\| = \left\| \sum_{k=1}^n \sigma_{\ell,k} \mathbf{z}_k^{(\ell)} v_{k,v}^\ell \right\| = 1, \ell = 1, \dots, d$ , we obtain

$\sum_{k=1}^n \sigma_{\ell,k}^2 (v_{k,v}^\ell)^2 = 1$  for  $\ell = 1, \dots, d, v = 1, \dots, R$ . Now the error estimate follows

$$\|\mathbf{A} - \mathbf{A}_{(r)}^0\| \leq \sum_{\ell=1}^d \sum_{v=1}^R |\xi_v| \left( \sum_{k=r_\ell+1}^n \sigma_{\ell,k}^2 (v_{k,v}^\ell)^2 \right)^{1/2}$$

$$\leq \sum_{\ell=1}^d \left( \sum_{v=1}^R \xi_v^2 \right)^{1/2} \left( \sum_{v=1}^R \sum_{k=r_\ell+1}^n \sigma_{\ell,k}^2 (v_{k,v}^\ell)^2 \right)^{1/2}$$

$$= \sum_{\ell=1}^d \|\boldsymbol{\xi}\| \left( \sum_{k=r_\ell+1}^n \sigma_{\ell,k}^2 \sum_{v=1}^R (v_{k,v}^\ell)^2 \right)^{1/2}$$

$$= \|\boldsymbol{\xi}\| \sum_{\ell=1}^d \left( \sum_{k=r_\ell+1}^n \sigma_{\ell,k}^2 \right)^{1/2}.$$

The case  $R < n$  can be analyzed along the same line. □

Proof of Lemma 2.5.

*Proof:* (A) The canonical vectors  $\mathbf{y}_k^{(\ell)}$  of any test element on the left-hand side of (11),

$$\mathbf{Z} = \sum_{k=1}^R \lambda_k \mathbf{y}_k^{(1)} \otimes \dots \otimes \mathbf{y}_k^{(d)} \in \mathcal{C}_{R,n}, \tag{A1}$$

can be chosen in  $\text{span}\{\mathbf{v}_1^{(\ell)}, \dots, \mathbf{v}_{r_\ell}^{(\ell)}\}$ , that means

$$\mathbf{y}_k^{(\ell)} = \sum_{m=1}^{r_\ell} \mu_{k,m}^{(\ell)} \mathbf{v}_m^{(\ell)}, \quad k = 1, \dots, R, \ell = 1, \dots, d. \tag{A2}$$

Indeed, assuming

$$\mathbf{y}_k^{(\ell)} = \sum_{m=1}^{r_\ell} \mu_{k,m}^{(\ell)} \mathbf{v}_m^{(\ell)} + \mathbf{e}_k^{(\ell)} \quad \text{with} \quad \mathbf{e}_k^{(\ell)} \perp \text{span}\{\mathbf{v}_1^{(\ell)}, \dots, \mathbf{v}_{r_\ell}^{(\ell)}\},$$

we conclude that  $\mathbf{e}_k^{(\ell)}$  does not effect the cost function in (11) because of the orthogonality of  $V^{(\ell)}$ . Hence, setting  $\mathbf{e}_k^{(\ell)} = 0$ , and plugging (A2) in (A1), we arrive at the desired Tucker decomposition of  $\mathbf{Z}, \mathbf{Z} = \boldsymbol{\beta}_z \times_1 V^{(1)} \times_2 \dots \times_d V^{(d)}, \boldsymbol{\beta}_z \in \mathcal{C}_{R,r}$ . This implies

$$\|\mathbf{A} - \mathbf{Z}\|^2 = \|(\boldsymbol{\beta}_z - \boldsymbol{\beta}) \times_1 V^{(1)} \times_2 \dots \times_d V^{(d)}\|^2$$

$$= \|\boldsymbol{\beta} - \boldsymbol{\beta}_z\|^2 \geq \min_{\boldsymbol{\mu} \in \mathcal{C}_{R,r}} \|\boldsymbol{\beta} - \boldsymbol{\mu}\|^2.$$

On the other hand, we have

$$\begin{aligned} \min_{\mathbf{Z} \in \mathcal{C}_{R,n}} \|\mathbf{A} - \mathbf{Z}\|^2 &\leq \min_{\boldsymbol{\beta}_z \in \mathcal{C}_{R,r}} \|(\boldsymbol{\beta} - \boldsymbol{\beta}_z) \times_1 V^{(1)} \times_2 \dots \times_d V^{(d)}\|^2 \\ &= \min_{\boldsymbol{\mu} \in \mathcal{C}_{R,r}} \|\boldsymbol{\beta} - \boldsymbol{\mu}\|^2. \end{aligned}$$

This proves (11).

(B) Likewise, for any minimizer  $\mathbf{A}_{(R)} \in \mathcal{C}_{R,n}$  in the right-hand side in (11), one obtains

$$\mathbf{A}_{(R)} = \boldsymbol{\beta}_{(R)} \times_1 V^{(1)} \times_2 V^{(2)} \dots \times_d V^{(d)}$$

with the respective rank- $R$  core tensor  $\boldsymbol{\beta}_{(R)} = \sum_{k=1}^R \lambda_k \mathbf{u}_k^{(1)} \otimes \dots \otimes \mathbf{u}_k^{(d)} \in \mathcal{C}_{R,r}$ . Here  $\mathbf{u}_k^{(\ell)} = \{\mu_{k,m_\ell}^{(\ell)}\}_{m_\ell=1}^{r_\ell} \in \mathbb{R}^{r_\ell}$ , are calculated by plugging representation (A2) in (A1), and then by changing the order of summation,

$$\begin{aligned} \mathbf{A}_{(R)} &= \sum_{k=1}^R \lambda_k \mathbf{y}_k^{(1)} \otimes \dots \otimes \mathbf{y}_k^{(d)} \\ &= \sum_{k=1}^R \lambda_k \left( \sum_{m_1=1}^{r_1} \mu_{k,m_1}^{(1)} \mathbf{v}_{m_1}^{(1)} \right) \otimes \dots \otimes \left( \sum_{m_d=1}^{r_d} \mu_{k,m_d}^{(d)} \mathbf{v}_{m_d}^{(d)} \right) \\ &= \sum_{m_1=1}^{r_1} \dots \sum_{m_d=1}^{r_d} \left\{ \sum_{k=1}^R \lambda_k \prod_{\ell=1}^d \mu_{k,m_\ell}^{(\ell)} \right\} \mathbf{v}_{m_1}^{(1)} \otimes \dots \otimes \mathbf{v}_{m_d}^{(d)}. \end{aligned}$$

Now (12) implies that

$$\|\mathbf{A} - \mathbf{A}_R\| = \|\boldsymbol{\beta} - \boldsymbol{\beta}_{(R)}\|,$$

since the  $\ell$ -mode multiplication with the orthogonal side matrices  $V^{(\ell)}$  does not change the cost functional. Inspection of the left-hand side in (11) indicates that the latter equation ensures that  $\boldsymbol{\beta}_{(R)}$  is, in fact, the minimizer of the right-hand side in (11).  $\square$