Computing electrostatic potentials using regularization based on the range-separated tensor format

Peter Benner [*1], **Venera Khoromskaia** [†1,2], **Boris Khoromskij** [‡2], **Cleophas Kweyu** [§1], **and Matthias Stein** [¶1]

[1]*Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany*
[2]*Max Planck Institute for Mathematics in the Sciences, Inselstr. 22-26, D-04103 Leipzig, Germany*

Abstract

In this paper, we apply the range-separated (RS) tensor format [6] for the construction of new regularization scheme for the Poisson-Boltzmann equation (PBE) describing the electrostatic potential in biomolecules. In our approach, we use the RS tensor representation to the discretized Dirac delta [21] to construct an efficient RS splitting of the PBE solution in the solute (molecular) region. The PBE then needs to be solved with a regularized source term, and thus black-box solvers can be applied. The main computational benefits are due to the localization of the modified right-hand side within the molecular region and automatic maintaining of the continuity in the Cauchy data on the interface. Moreover, this computational scheme only includes solving a single system of FDM/FEM equations for the smooth long-range (i.e., regularized) part of the collective potential represented by a low-rank RS-tensor with a controllable precision. The total potential is obtained by adding this solution to the directly precomputed rank-structured tensor representation for the short-range contribution. Enabling finer grids in PBE computations is another advantage of the proposed techniques. In the numerical experiments, we consider only the free space electrostatic potential for proof of concept. We illustrate that the classical Poisson equation (PE) model does not accurately capture the solution singularities in the numerical approximation as compared to the new approach by the RS tensor format.

Key words: The Poisson-Boltzmann equation, Coulomb potential, summation of electrostatic potentials, long-range many-particle interactions, low-rank tensor decompositions, range-separated tensor formats.

AMS Subject Classification: 65F30, 65F50, 65N35, 65F10

[*]benner@mpi-magdeburg.mpg.de

[†]vekh@mis.mpg.de

[‡]bokh@mis.mpg.de

[§]kweyu@mpi-magdeburg.mpg.de

[¶]matthias.stein@mpi-magdeburg.mpg.de

arXiv:1901.09864v1 [math.NA] 28 Jan 2019

1   Introduction

Numerical treatment of long-range interaction potentials is a challenging task in computer modeling of multiparticle systems, for example, in calculation of electrostatics in large solvated biological systems, in protein docking, or in many particles dynamics simulations [10, 16, 27, 30, 32]. The well-known Poisson-Boltzmann equation (PBE) introduced and analyzed for example, in [15], is one of the most popular implicit solvent models for computation of the electrostatic potential in proteins. Other models include the generalized Born (GB) methods [3] and the polarizable continuum models (PCM) [2]. The PBE computes the electrostatic potential both in the protein and in the surrounding solvent, and it is widely used in protein docking, in classification problems, and for computation of the free energy of biomolecules in a self-consistent way.

The main difficulty in the traditional finite element method (FEM) approximation schemes for the PBE problem is related to the presence of a highly singular source term that includes a large sum of Dirac delta distributions which need to be resolved using rather coarse grids. To overcome these limitations, a number of regularization schemes for the finite element method (FEM) applied to the PBE, based on the full grid representation of all functional data, have been considered in the literature, see for example [13, 34] and references therein. Consequently, we note that the PBE theory has recently received major improvement in terms of accuracy by the introduction of solution decomposition techniques which have been developed for example, in [9, 29, 34], where the PBE is treated as an interface problem. This aims at avoiding the discontinuities in traces and fluxes at the interface between the biomolecule and the solvent and also to circumvent building the numerical approximations corresponding to the Dirac delta distributions because of the existence of an analytical expansions in the solute sub-region.

However, these techniques still face the following computational challenges. First, jumps in the interface conditions, arising due to regularization splitting of the solution, need to be incorporated to eliminate the solution discontinuity (e.g., Cauchy data) at the interface. Second, the boundary conditions have to be specified using some analytical representation of the solution of the PBE. And third, in regularization-based techniques, see for example, [34], one has to solve multiple algebraic systems for the linear and nonlinear boundary value problems before summing up the partial solutions which increases the computational costs. We provide an overview of these techniques in Section 2.

Here, we present a new approach for the regularization of the PBE by using the range-separated (RS) canonical tensor format introduced and analyzed in [6]. The RS tensor format relies on the independent grid-based low-rank tensor representation of the long- and short-range parts in the total sum of single-particle electrostatic potentials discretized on a fine 3D $n \times n \times n$ Cartesian grid Ω_n in the computational box $\Omega \subset \mathbb{R}^3$. This representation is based on the splitting of a single reference potential, defined by a radial function like $p(\|\bar{x}\|) = 1/\|\bar{x}\|$, into a sum of localized and long-range low-rank canonical tensors both represented on the computational grid Ω_n. The long-range part in the collective potential of a many-particle system is represented as a low-rank canonical tensor with the rank only logarithmically depending on the number of particles in the system. The rank reduction algorithm is performed by the canonical-to-Tucker (C2T) transform via the reduced higher order singular value decomposition (RHOSVD) [23] with a subsequent Tucker-to-canonical (T2C) decomposition.

2

The short-range contributions to a many-particle potential are parametrized by a single low-rank canonical tensor of local support. Notice that in [5] ,it was already sketched how the RS tensor formats may be utilized for calculation of the free interaction energy of protein-type systems, and the idea for regularized formulation of the PBE by using the smooth long-range part of the free space electrostatic potential was outlined.

In this paper ,we introduce the new regularization scheme for the solution of the PBE adapting the RS tensor format. It is based on a certain splitting scheme for the highly singular solution and right-hand side in PBE, by using the RS tensor decomposition of the discretized Dirac delta introduced in [21]. This approach requires only a simple modification (regularization) in the right-hand side of the PBE in the solute region, but it does not change the interface conditions and, hence, the FEM system matrix. The most singular component in the potential is recovered explicitly by the short range part in the RS tensor splitting of the free space potential. The main computational benefits are due to the localization of the modified right-hand side within the molecular region and automatic maintaining of the continuity in the Cauchy data on the interface. Furthermore, this computational scheme only includes solving a single system of FEM equations for the smooth long-range (i.e., regularized) part of the collective potential represented by a low-rank RS-tensor with a controllable precision. The total potential is obtained by adding this solution to the directly precomputed rank-structured tensor representation for the short-range contribution.

As numerical illustrations, we compute the free-space electrostatic potentials of biomolecules using the RS tensor format in the framework of splitting scheme, and compare them with the solutions calculated by the traditional FEM/FDM discretization methods for the PBE [1, 13, 15].

The rest of the paper is organized as follows. Section 2 provides a short overview of existing solution decomposition schemes for the PBE problem. Section 3 describes the principles of the rank-structured tensor approximation to the long-range electrostatic potential and sketches the RS tensor decomposition techniques [6] for the free space electrostatic potential of many particle systems. The main Section 4 explains how the application of the RS tensor format leads to the new regularization scheme for solving the PBE. Finally, Section 5 presents the numerical tests illustrating the benefits of the proposed method, and comparisons with the solutions obtained by the standard FDM-based PBE solver are provided.

2   On existing solution decomposition techniques

The PBE computes the dimensionless potential $u(\bar{x}) = e_c \psi(\bar{x})/K_B T$, which is scaled by $e_c/K_B T$, and $\psi(\bar{x})$ is the original electrostatic potential in centimeter-gram-second (cgs) units at $\bar{x} = (x, y, z) \in \mathbb{R}^3$. It is given by

$$-\nabla \cdot (\epsilon(\bar{x})\nabla u(\bar{x})) + \bar{\kappa}^2(\bar{x})\sinh(u(\bar{x})) = \frac{4\pi e_c^2}{K_B T}\sum_{i=1}^{N_m} z_i \delta(\bar{x} - \bar{x}_i), \quad \Omega \in \mathbb{R}^3, \qquad (2.1)$$

subject to

$$u(\bar{x}) = \frac{e_c^2}{K_B T}\sum_{i=1}^{N_m} \frac{z_i e^{-\bar{\kappa}(d-a_i)}}{\epsilon_s(1+\bar{\kappa}a_i)d} \quad \text{on } \partial\Omega, \quad d = \|\bar{x} - \bar{x}_i\|, \qquad (2.2)$$

3

where $K_B T$, K_B, T, and e_c are the thermal energy, the Boltzmann constant, the absolute temperature, and the electron charge, respectively, $\bar{\kappa}^2 = 8\pi e_c^2 I/1000\epsilon_s K_B T$ is a function of the ionic strength $I = 1/2 \sum_{j=1}^{m} c_j z_j^2$, where c_j and z_j are the charge and concentration of each ion. The sum of Dirac delta distributions, located at atomic centers $\bar{x}_i$, represents the molecular charge density, z_i are the point partial charges of the protein, ϵ_s is the solvent dielectric constant, a_i are the atomic radii, and N_m is the total number of point partial charges in the protein. The functions $\epsilon(\bar{x})$ and $\bar{\kappa}^2(\bar{x})$ are piecewise constant defined by

$$\epsilon(\bar{x}) = \begin{cases} \epsilon_m = 2 & \text{if } \bar{x} \in \Omega_m \\ \epsilon_s = 78.54 & \text{if } \bar{x} \in \Omega_s \end{cases}, \qquad \bar{\kappa}(\bar{x}) = \begin{cases} 0 & \text{if } \bar{x} \in \Omega_m \\ \sqrt{\epsilon_s}\bar{\kappa} & \text{if } \bar{x} \in \Omega_s \end{cases}, \qquad (2.3)$$

where Ω_m and Ω_s are the molecular and solvent regions, respectively, as shown in Figure 2.1.

In order to overcome the difficulties arising from the singularities caused by the impulsive source term, several solution decomposition techniques have been suggested in the literature. In the following, we will discuss those techniques that form the state of the art for solving the PBE. Following [34], the first solution decomposition is generally given by

$$\left.\begin{aligned} -\epsilon_m \Delta u(\bar{x}) &= C \sum_{i=1}^{N_m} z_i \delta(\bar{x} - \bar{x}_i), & \bar{x} \in \Omega_m, \\ -\epsilon_s \Delta u(\bar{x}) + \bar{\kappa}^2 \sinh(u(\bar{x})) &= 0, & \bar{x} \in \Omega_s, \\ u(s^+) = u(s^-), \quad \epsilon_s \frac{\partial u(s^+)}{\partial n(s)} &= \epsilon_m \frac{\partial u(s^-)}{\partial n(s)}, & s \in \Gamma, \\ u(s) &= g(s), & s \in \partial\Omega, \end{aligned}\right\} \qquad (2.4)$$

where $C = \frac{4\pi e_c^2}{K_B T}$, s^+ and s^- represent the grid points in the vicinity of the interface Γ in the solvent and the molecular regions, respectively. The solution $u(\bar{x})$ is decomposed as follows,

$$u(\bar{x}) = G(\bar{x}) + \tilde{\phi}(\bar{x}) + \tilde{\psi}(\bar{x}). \qquad (2.5)$$
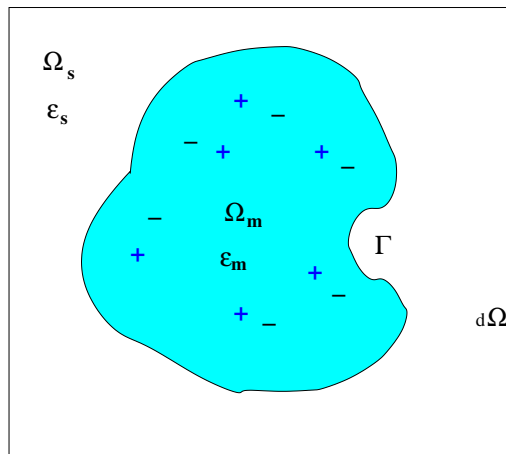
Figure 2.1: Solute and solvent regions in the computational domain for the PBE.

The corresponding components of $u(\bar{x})$ include the analytical solution $G(\bar{x})$ of the Poisson equation in the molecular domain,

$$G(\bar{x}) = \frac{\hat{C}}{\epsilon_m} \sum_{i=1}^{N_m} \frac{z_i}{\|\bar{x} - \bar{x}_i\|}, \qquad \bar{x} \in \Omega_m, \tag{2.6}$$

the solution of the linear interface boundary value problem

$$\left.\begin{aligned}
\Delta\tilde{\phi}(\bar{x}) &= 0, & \bar{x} \in \Omega_m \cup \Omega_s, \\
\tilde{\phi}(s^+) = \tilde{\phi}(s^-), \quad \epsilon_s \frac{\partial\tilde{\phi}(s^+)}{\partial n(s)} &= \epsilon_m \frac{\partial\tilde{\phi}(s^-)}{\partial n(s)} + (\epsilon_m - \epsilon_s)\frac{G(s)}{\partial n(s)}, & s \in \Gamma, \\
u(s) &= g(s) - G(s), & s \in \partial\Omega,
\end{aligned}\right\} \tag{2.7}$$

and the solution of the nonlinear interface boundary value problem

$$\left.\begin{aligned}
\Delta\tilde{\psi}(\bar{x}) &= 0, & \bar{x} \in \Omega_m, \\
-\epsilon_s\Delta\tilde{\psi}(\bar{x}) + \bar{\kappa}^2\sinh(\tilde{\psi}(\bar{x}) + \tilde{\phi}(\bar{x}) + G(\bar{x})) &= 0, & \bar{x} \in \Omega_s, \\
\tilde{\psi}(s^+) = \tilde{\psi}(s^-), \quad \epsilon_s\frac{\partial\tilde{\psi}(s^+)}{\partial n(s)} &= \epsilon_m\frac{\partial\tilde{\psi}(s^-)}{\partial n(s)}, & s \in \Gamma, \\
u(s) &= 0, & s \in \partial\Omega.
\end{aligned}\right\} \tag{2.8}$$

Secondly, we recall the solution decomposition from [29] which takes the form

$$u(\bar{x}) = \hat{u}(\bar{x}) + \tilde{u}(\bar{x}).$$

The short-range part $\hat{u}(\bar{x})$ is given by

$$\hat{u}(\bar{x}) = \begin{cases} G(\bar{x}) + u^0(\bar{x}) & \text{if} \quad \bar{x} \in \Omega_m, \\ 0 & \text{if} \quad \bar{x} \in \Omega_s, \end{cases} \tag{2.9}$$

where $u^0(\bar{x})$ is a harmonic function which compensates for the discontinuity on the interface and satisfies the following conditions

$$\left.\begin{aligned}
\Delta u^0(\bar{x}) &= 0, & \text{if} \quad \bar{x} \in \Omega_m, \\
u^0(s) &= -G(s), & s \in \Gamma.
\end{aligned}\right\} \tag{2.10}$$

The regular part $\tilde{u}(\bar{x})$ is represented by

$$\left.\begin{aligned}
-\nabla \cdot (\epsilon(\bar{x})\nabla\tilde{u}(\bar{x})) + \bar{\kappa}^2(\bar{x})\sinh(\tilde{u}(\bar{x})) &= 0, \\
[\tilde{u}(\bar{x})]_\Gamma = 0, \quad [\epsilon\nabla\tilde{u}(\bar{x}) \cdot \mathbf{n}]_\Gamma &= -\epsilon_m\nabla(G(\bar{x}) + u^0(\bar{x})) \cdot \mathbf{n}|_\Gamma.
\end{aligned}\right\} \tag{2.11}$$

Lastly, the solution decomposition in [9] is as follows: $u(\bar{x}) = G(\bar{x}) + u^r(\bar{x})$, where $G(\bar{x})$ is as in (2.6) and the regular part is given by

$$\left.\begin{aligned}
-\nabla \cdot (\epsilon\vec{\nabla}u^r) + \bar{\kappa}^2\sinh(u^r + G) &= \nabla \cdot ((\epsilon - \epsilon_m)\vec{\nabla}G), & \text{in } \Omega \\
u^r &= g - G, & \text{on } \partial\Omega.
\end{aligned}\right\} \tag{2.12}$$

5

We can also further decompose (2.12) into the linear and nonlinear components so that $u^r(\bar{x}) = u^l(\bar{x}) + u^n(\bar{x})$, where

$$\left.\begin{array}{rl} -\nabla \cdot (\epsilon \vec{\nabla} u^l) = \nabla \cdot ((\epsilon - \epsilon_m)\nabla G), & \text{in } \Omega \\ u^l = 0 & \text{on } \partial\Omega, \end{array}\right\} \tag{2.13}$$

and

$$\left.\begin{array}{rl} -\nabla \cdot (\epsilon \vec{\nabla} u^n) + \bar{\kappa}^2 \sinh(u^n + u^l + G) = 0, & \text{in } \Omega \\ u^n = g - G & \text{on } \partial\Omega. \end{array}\right\} \tag{2.14}$$

The fundamental idea in the above decomposition strategies is the pursuit of an efficient solution decomposition technique for the short- and long-range parts in a target tensor. However, all these techniques do not efficiently separate the long- and short-range components in the total potential sum, hence there is a need to apply so-called interface (or jump) conditions at the interface between the molecular region and the solvent region in order to reduce discontinuities. In what follows, we describe the RS tensor format developed in [6], applied as the main tool in our paper to modify the PBE to increase the accuracy of its numerical approximation. A more detailed description of the new solution decomposition technique is provided in Sections 4.1 and 4.2.

3 Rank-structured approximation of electrostatic potentials

Tensor-structured numerical methods are now becoming popular in scientific computing due to their intrinsic property of reducing the grid-based solution of the multidimensional problems to essentially "one-dimensional" computations. These methods evolved from bridging of the traditional rank-structured tensor formats of multilinear algebra [24, 31] with the nonlinear approximation theory based on a separable representation of multidimensional functions and operators [11, 12, 20]. One of the ingredients in the development of tensor methods was the RHOSVD which allows to reduce the rank of tensors in a canonical format by the C2T decomposition without the need to construct the full size tensor [23]. Originally, it was used for the reduction of the ranks of canonical tensors when calculating three-dimensional convolution integrals in computational quantum chemistry, see the surveys [18, 22] and the references therein.

Recently, tensor-based approaches were suggested as new methods for the calculation of multiparticle long-range interaction potentials. For a given non-local generating kernel $p(\|\bar{x}\|)$, $\bar{x} \in \mathbb{R}^3$, the calculation of the weighted sum of interaction potentials in an N-particle system, with the particle locations at $\bar{x}_\nu \in \mathbb{R}^3$, $\nu = 1, ..., N$,

$$P_0(\bar{x}) = \sum_{\nu=1}^{N} z_\nu \, p(\|\bar{x} - \bar{x}_\nu\|), \quad z_\nu \in \mathbb{R}, \quad \bar{x}_\nu, \bar{x} \in \Omega = [-b, b]^3, \tag{3.1}$$

is computationally demanding for large N. Since the generating radial basis function $p(\|\bar{x}\|)$ exhibits a slow polynomial decay in $1/\|\bar{x}\|$ as $\|\bar{x}\| \to \infty$, it follows that each individual term in (3.1) contributes essentially to the total potential at each point in the computational domain Ω. This predicts the $\mathcal{O}(N)$ complexity for a straightforward summation at every fixed space

point $\bar{x} \in \mathbb{R}^3$. Moreover, in general, the radial function $p(\|\bar{x}\|)$ has a singularity or a cusp at the origin, $\bar{x} = 0$, making its accurate grid representation problematic. An efficient numerical scheme for the grid-based calculation of $P(\bar{x})$ in multiparticle systems can be constructed by using the RS tensor format [6].

3.1   Canonical tensor approximation of the Newton kernel

First, we recall the grid-based method for the low-rank canonical representation of a spherically symmetric kernel function $p(\|\bar{x}\|)$, $\bar{x} \in \mathbb{R}^d$ for $d = 2, 3, \ldots$, by its projection onto the set of piecewise constant basis functions, see [7] for the case of the Newton kernel $p(\|\bar{x}\|) = \frac{1}{\|\bar{x}\|}$, for $x \in \mathbb{R}^3$. A single reference potential like $1/\|\bar{x}\|$ can be represented on a fine 3D $n \times n \times n$ Cartesian grid as a low-rank canonical tensor [7, 12].

In the computational domain $\Omega = [-b, b]^3$, let us introduce the uniform $n \times n \times n$ rectangular Cartesian grid Ω_n with mesh size $h = 2b/n$ (n even). Let $\{\psi_{\mathbf{i}}\}$ be a set of tensor-product piecewise constant basis functions, $\psi_{\mathbf{i}}(\bar{x}) = \prod_{\ell=1}^3 \psi_{i_\ell}^{(\ell)}(\bar{x}_\ell)$, for the 3-tuple index $\mathbf{i} = (i_1, i_2, i_3)$, $i_\ell \in I_\ell = \{1, ..., n\}$, $\ell = 1, 2, 3$. The generating kernel $p(\|\bar{x}\|)$ 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

$$\mathbf{P} := [p_{\mathbf{i}}] \in \mathbb{R}^{n \times n \times n}, \quad p_{\mathbf{i}} = \int_{\mathbb{R}^3} \psi_{\mathbf{i}}(\bar{x}) p(\|\bar{x}\|) \, d\bar{x}. \tag{3.2}$$

Then using the Laplace-Gauss transform and sinc-quadratures, the 3rd order tensor $\mathbf{P}$ can be approximated by the R-term canonical representation (see [7, 12, 17] for details),

$$\mathbf{P} \approx \mathbf{P}_R = \sum_{k=1}^R \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)} \in \mathbb{R}^{n \times n \times n}, \tag{3.3}$$

where $\mathbf{p}_k^{(\ell)} \in \mathbb{R}^n$.

Note that the reference tensor for summation of the potentials is generated in a larger computational domain, $\widetilde{\mathbf{P}}_R \in \mathbb{R}^{2n \times 2n \times 2n}$, which is necessary for application of the shift-and-windowing transforms $\mathcal{W}_\nu$, see Section 3.2 and [19] for more details.

The canonical tensor representation of the Newton kernel (3.3) has been successfully applied in computation of multidimensional operators in quantum chemistry [19, 23], where it was shown that calculations using the grid-based tensor approximations exhibit the same high accuracy level as the analytically based computation methods for the same multidimensional operators. The recent assembled tensor summation method of the long-range electrostatic potentials on large finite lattices [17] has been proved to keep the rank of the collective potential on large 3D lattices to be as small as the rank of a canonical tensor for a single Newton kernel.

3.2   Properties of the RS tensor format

Here we recall some properties of the range separated (RS) tensor format introduced in [6] for modeling of the long-range interaction potential in multiparticle systems of general type. It is based on the partitioning of the reference tensor representation of the Newton kernel into long- and short-range parts with a following assembling of the collective electrostatic potential

of a molecular system in a special way. According to the tensor canonical representation of the Newton kernel (3.3) as a sum of Gaussians, one can distinguish their supports as the short- and long-range parts,

$$\mathbf{P}_R = \mathbf{P}_{R_s} + \mathbf{P}_{R_l},$$

where

$$\mathbf{P}_{R_s} = \sum_{k \in \mathcal{K}_s} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)}, \quad \mathbf{P}_{R_l} = \sum_{k \in \mathcal{K}_l} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)}. \tag{3.4}$$

Here, $\mathcal{K}_l := \{k | k = 0, 1, \ldots, R_l\}$ and $\mathcal{K}_s := \{k | k = R_l + 1, \ldots, M\}$ are the sets of indices for the long- and short-range canonical vectors. Then the optimal splitting (3.4) is applied to the reference canonical tensor $\mathbf{P}_R$ and to its accompanying version $\widetilde{\mathbf{P}}_R = [\widetilde{p}_R(i_1, i_2, i_3)]$, $i_\ell \in \widetilde{I}_\ell$, $\ell = 1, 2, 3$, such that

$$\widetilde{\mathbf{P}}_R = \widetilde{\mathbf{P}}_{R_s} + \widetilde{\mathbf{P}}_{R_l} \in \mathbb{R}^{2n \times 2n \times 2n}.$$

The total electrostatic potential $P_0(\bar{x})$ in (3.1) is represented by a projected tensor $\mathbf{P}_0 \in \mathbb{R}^{n \times n \times n}$ that can be constructed by a direct sum of shift-and-windowing transforms of the reference tensor $\widetilde{\mathbf{P}}_R$ (see [17] for more details),

$$\mathbf{P}_0 = \sum_{\nu=1}^{N} z_\nu \mathcal{W}_\nu(\widetilde{\mathbf{P}}_R) = \sum_{\nu=1}^{N} z_\nu \mathcal{W}_\nu(\widetilde{\mathbf{P}}_{R_s} + \widetilde{\mathbf{P}}_{R_l}) =: \mathbf{P}_s + \mathbf{P}_l. \tag{3.5}$$

The shift-and-windowing transform $\mathcal{W}_\nu$ maps a reference tensor $\widetilde{\mathbf{P}}_R \in \mathbb{R}^{2n \times 2n \times 2n}$ onto its sub-tensor of smaller size $n \times n \times n$, obtained by first shifting the center of the reference tensor $\widetilde{\mathbf{P}}_R$ to the grid-point x_ν and then restricting (windowing) the result onto the computational grid Ω_n. However, the tensor representation (3.5) is non-efficient as the ranks are growing linearly in the number of particles and remain non-reducible in both canonical and Tucker tensor formats.

This problem is solved in [6] by considering the global tensor decomposition of only the "long-range part" in the tensor $\mathbf{P}_0$, defined by

$$\mathbf{P}_l = \sum_{\nu=1}^{N} z_\nu \mathcal{W}_\nu(\widetilde{\mathbf{P}}_{R_l}) = \sum_{\nu=1}^{N} z_\nu \mathcal{W}_\nu(\sum_{k \in \mathcal{K}_l} \widetilde{\mathbf{p}}_k^{(1)} \otimes \widetilde{\mathbf{p}}_k^{(2)} \otimes \widetilde{\mathbf{p}}_k^{(3)}). \tag{3.6}$$

In [6] it was proven that the canonical rank of the tensor $\mathbf{P}_l$ only logarithmically depends on the number of particles N involved in the summation. It was also shown that the rank reduction $\mathbf{P}_l \mapsto \mathbf{P}_{R_L}$ may be efficiently implemented by using the C2T and T2C algorithms, where the RHOSVD is a key ingredient [23]. As for the short-range part of the collective potential, in the RS format it is represented by a single small size tensor supplemented by a list of the particles' coordinates.

According to the definition of the RS canonical tensor format introduced in [6], for the given separation parameter $\gamma \in \mathbb{N}$, which is used for partitioning of the canonical vectors into the long and short range contributions, the RS-canonical tensor format specifies the class of d-tensors $\mathbf{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ which can be represented as a sum of a rank-R_L canonical tensor

$$\mathbf{A}_{R_L} = \sum_{k=1}^{R_L} \xi_k \mathbf{a}_k^{(1)} \otimes \cdots \otimes \mathbf{a}_k^{(d)} \in \mathbb{R}^{n_1 \times \ldots \times n_d} \tag{3.7}$$
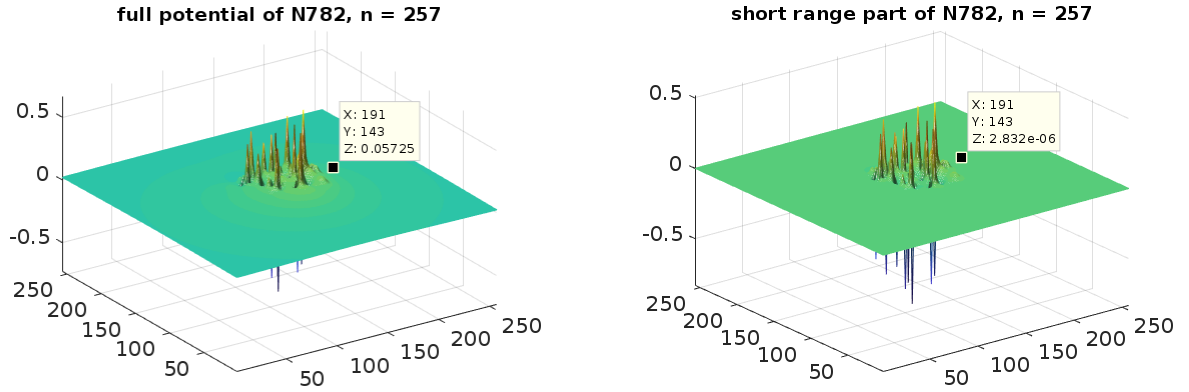
Figure 3.1: The full free space electrostatic potential of a system with 782 particles (left) and the sum of their short range contributions (right).

and a (uniform) cumulated canonical tensor

$$\widehat{\mathbf{A}}_S = \sum_{\nu=1}^{N} c_\nu \mathbf{A}_\nu \tag{3.8}$$

generated by shifts of a short reference tensor $\mathbf{A}_0$ with $\mathrm{rank}(\mathbf{A}_0) \leq R_0$. With these notations, the RS canonical tensor format reads as

$$\mathbf{A} = \sum_{k=1}^{R_L} \xi_k \mathbf{a}_k^{(1)} \otimes \cdots \otimes \mathbf{a}_k^{(d)} + \sum_{\nu=1}^{N} c_\nu \mathbf{A}_\nu, \tag{3.9}$$

where $\mathrm{diam}(\mathrm{supp}\mathbf{A}_\nu) \leq 2\gamma$.

Lemma 3.9 in [6] presents the storage cost of the RS-canonical tensor $\mathbf{A}$ in (3.9) as follows:

$$\mathrm{mem}(\mathbf{A}) \leq dR_L n + (d+1)N + dR_0\gamma.$$

Given $\mathbf{i} \in \mathcal{I} = I_1 \times \cdots \times I_d$, denote by $\overline{\mathbf{a}}_{i_\ell}^{(\ell)} \in \mathbb{R}_{\mathbb{L}}^{1 \times R}$ the row-vector with index i_ℓ in the side matrix $A^{(\ell)} \in \mathbb{R}^{n_\ell \times R_L}$ of $\mathbf{A}$, and let $\xi = (\xi_1, \ldots, \xi_d)$. Then the $\mathbf{i}$-th entry of the RS-canonical tensor $\mathbf{A} = [a_{\mathbf{i}}]$ can be calculated as a sum of long- and short-range contributions by

$$a_{\mathbf{i}} = \left(\odot_{\ell=1}^{d} \overline{\mathbf{a}}_{i_\ell}^{(\ell)} \right) \xi^T + \sum_{\nu \in \mathcal{L}(\mathbf{i})} c_\nu \mathbf{A}_\nu(\mathbf{i}),$$

at the expense $O(dR_L + 2d\gamma R_0)$. Here, $\mathcal{L}(\mathbf{i}) := \{\nu \in \{1, \ldots, N\} : \mathbf{i} \in \mathrm{supp}\mathbf{A}_\nu\}$ is the set of indices which label all the short-range tensors $\mathbf{A}_\nu$ that include the grid point $\mathbf{i}$ within their effective support [6].

The RS tensor may be represented in a Tucker tensor format as well, see [6]. RS tensors are efficient in many applications, for example for modeling of the electrostatics of many-particle systems of general type, or for modeling of scattered multidimensional data by using radial basis functions.

Next, we illustrate the performance of the canonical RS-tensor format in calculating the collective free space electrostatic potential of a model molecular system with 782 atoms, see also [6]. Summation is performed using the canonical tensor representing the reference
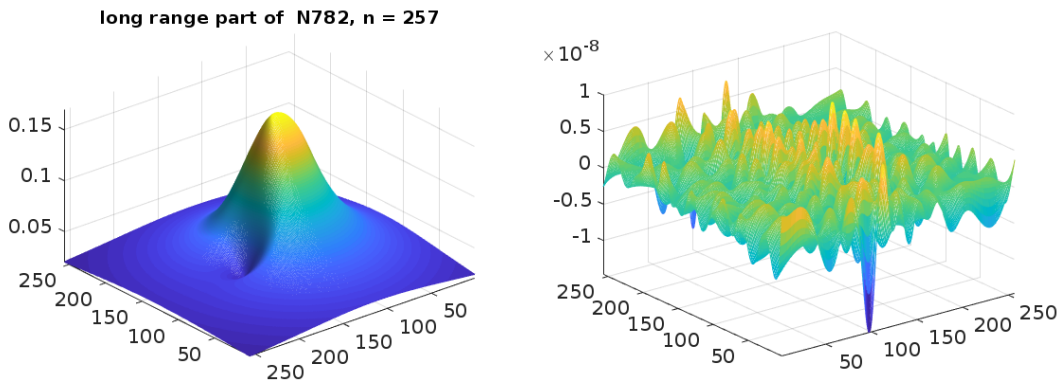
Figure 3.2: The low-rank tensor representation of the long-range part in the electrostatic potential of 782 charged particles (left) and the error of the canonical rank reduction.

Newton potential computed on the $n \times n \times n$ 3D Cartesian grid with $n = 257$ and the canonical rank $R = 29$. The RS-tensor construction is performed with $r_\ell = 14$, $R_s = 15$, simply by dividing the canonical vectors into two groups, that is from every 29 vectors of the reference tensor and for every space dimension (x, y, z), 15 sharp Gaussians are separated as the short range part, and 14 smoother Gaussians as the long-range part[1]. Then their contributions to the collective sum of potentials are calculated separately. To reduce the rank of the long-range collective sum, the C2T (and T2C) transforms using RHOSVD [23] are used, with a choice of the truncation threshold 10^{-8}. The resulting rank decreases from $R_{\ell,total} = 10948$, to $R_\ell = 382$ and only logarithmically depends on the number of particles N [6].

The left panel in Figure 3.1 shows the cross-section of the collective electrostatic potential of a molecular system at the middle of a z-plane, while the right panel shows the cross-section of only the short-range part of the collective potential $\widehat{\mathbf{A}}_S$. Notice that in the left panel showing the total potential sum, the potential at the point of the plane with $(x, y) = (191, 143)$ equals to ~ 0.057 units, while for the short range (right panel) the sum equals to $2.8 \, 10^{-6}$. The left panel in Figure 3.2 presents the cross-section of the low-rank long-range part of the collective potential at the same plane, while the right panel in this figure shows the error of the rank reduction (the same truncation threshold 10^{-8} as chosen above).

4 Application of RS tensor format for solving PBE

The RS tensor formats can be gainfully applied in computational problems which include functions with multiple local singularities or cusps, Green kernels with essentially non-local behavior, as well as in various approximation problems treated by means of radial basis functions. In what follows, we describe the new approach for the construction of computationally effective boundary/interface conditions and source terms in the Poisson-Boltzmann equation (PBE) describing the electrostatic potential of a biomolecule in gas phase and in solvent by

---

[1]Alternatively, separation of the reference tensor into the short- and long-range Gaussians may be performed by a chosen ε-truncation for the given interatomic distance.

solving the FEM/ FDM discretization of the regularized PBE. The main advantage of our approach is due to complete avoidance of the direct FEM approximation (interpolation) of the highly singular right-hand sides in the traditional formulation of the PBE and, at the same time, preventing the modification of the stiffness matrix and/or the continuity conditions across the interface.

4.1   Tensor based regularization scheme for the PBE

The traditional numerical approaches for solving the PBE are based on either multigrid [28] or domain decomposition [8] methods. Consider a solvated biomolecular system modeled by dielectrically separated domains with singular Coulomb potentials distributed in the molecular region. For schematic representation, we consider the system occupying a rectangular domain Ω with boundary $\partial\Omega$, see Figure 2.1, where the solute (molecule) region is represented by Ω_m and the solvent region by Ω_s, such that

$$\overline{\Omega} = \overline{\Omega}_m \cup \overline{\Omega}_s.$$

The linearized Poisson-Boltzmann equation takes the form, see [28],

$$-\nabla \cdot (\epsilon \nabla u) + \bar{\kappa}^2 u = \rho_f \quad \text{in} \quad \Omega, \tag{4.1}$$

where u denotes the target electrostatic potential of a protein, and

$$\rho_f = \sum_{k=1}^{N} z_k \delta(\|\bar{x} - \bar{x}_k\|), \quad z_k \in \mathbb{R},$$

is the scaled singular charge distribution supported at points $\bar{x}_k$ in Ω_m, where δ is the Dirac delta distribution, and $z_k \in \mathbb{R}$ denotes the charge located at the atomic center $\bar{x}_k$. Here $\epsilon_m = O(1) > 0$ and $\bar{\kappa} = 0$ in Ω_m, while in the solvent region Ω_s, we have $\bar{\kappa} \geq 0$ and $\epsilon_s \geq \epsilon_m$ (in some cases the ratio ϵ_s/ϵ_m could be about several tens).

The interface conditions on the interior boundary $\Gamma = \partial\Omega_m$ arise from the dielectric theory:

$$[u] = 0, \quad \left[\epsilon \frac{\partial u}{\partial n}\right] = 0 \quad \text{on} \quad \Gamma. \tag{4.2}$$

The boundary conditions on the external boundary $\partial\Omega$ can be specified depending on the particular problem setting. The simplest homogeneous Dirichlet boundary conditions

$$u_{|\partial\Omega} = 0$$

can be utilized. We shall also consider the practically interesting inhomogeneous Dirichlet boundary conditions taking the form

$$u(\bar{x})_{|\partial\Omega} = \frac{1}{4\pi\epsilon_m} \sum_{k=1}^{N} \frac{z_k e^{-\bar{\kappa}\|\bar{x} - \bar{x}_k\|}}{\|\bar{x} - \bar{x}_k\|}, \quad \bar{x} \in \partial\Omega. \tag{4.3}$$

The practically useful solution methods for the PBE are based on regularization schemes aiming at removing the singular component from the potentials in the governing equation.

Amongst others, we consider one of the most commonly used approaches based on the additive splitting of the potential in the molecular region Ω_m, see for example [28]. To that end, we first discuss the additive splitting techniques introduced in [5], based on the application of the RS tensor format,

$$u = u^r + u_0^m, \quad \text{where} \quad u_0^m = 0 \quad \text{in} \quad \Omega_s, \tag{4.4}$$

and where the singular component u_0^m satisfies the following Poisson equation in Ω_m,

$$-\epsilon_m \Delta u_0^m = \rho_f \quad \text{in} \quad \Omega_m; \quad u_0^m = 0 \quad \text{on} \quad \Gamma. \tag{4.5}$$

In this case, equation (4.1) can be transformed to an equation for the regular potential u^r:

$$-\nabla \cdot (\epsilon \nabla u^r) + \bar{\kappa}^2 u^r = 0 \quad \text{in} \quad \Omega, \tag{4.6}$$

$$[u^r] = 0, \quad \left[\epsilon \frac{\partial u^r}{\partial n}\right] = -\epsilon_m \frac{\partial u_0^m}{\partial n} \quad \text{on} \quad \Gamma.$$

To facilitate the solution of equation (4.5) with highly singular data in the right-hand side, the singular potential U in the free space was utilized, see [6],

$$-\epsilon_m \Delta U = \rho_f \quad \text{in} \quad \mathbb{R}^3, \quad |U(\bar{x})| \to 0, \quad |\bar{x}| \to \infty, \tag{4.7}$$

that can be written in the explicit form

$$U(\bar{x}) = \frac{1}{4\pi\epsilon_m} \sum_{k=1}^{N} \frac{z_k}{\|\bar{x} - \bar{x}_k\|}.$$

Introduce the characteristic (indicator) function, $\chi_{[\Omega_m]}(\bar{x})$, $\bar{x} \in \Omega$, of the domain $\Omega_m \subset \Omega$ by

$$\chi_{[\Omega_m]}(\bar{x}) = \begin{cases} 1 & \text{if } \bar{x} \in \overline{\Omega}_m \\ 0 & \text{if } \bar{x} \in \Omega_s = \Omega \setminus \overline{\Omega}_m. \end{cases} \tag{4.8}$$

Then the restriction of U onto Ω_m can be calculated by

$$u^m = \chi_{[\Omega_m]}U,$$

implying the decomposition

$$u_0^m = u^m + u^{harm},$$

where the harmonic function u^{harm} compensates the discontinuity of u^m on Γ,

$$\Delta u^{harm} = 0 \quad \text{in} \quad \Omega_m; \quad u^{harm} = -u^m = -U \quad \text{on} \quad \Gamma.$$

The advantage of this formulation is twofold:

(a) the absence of singularities in the solution u^r, and

(b) the localization of the solution splitting only in the domain Ω_m.

Grid representation of the free space singular potential U, which may include a sum of hundreds or even thousands of single Newton kernels in 3D, leads to a challenging computational problem. In our approach it can be represented on large tensor grids in Ω with controlled precision by using the RS tensor format [6] characterized by the separability constant $\gamma > 0$ which in our application can be associated with the Van Der Waals inter-atomic distance, see Section 3.2. The long-range component in the formatted parametrization remains smooth and allows global low-rank representation in Ω. We conclude with the following

Proposition 4.1 *Let the effective support of the short-range components in the reference potential* $\mathbf{P}_R$ *be chosen not larger than* $\gamma/2$*. Then the interface conditions in the regularized formulation of the PBE in (4.6) depend only on the low-rank long-range component in the free-space electrostatic potential of the atomic system. The numerical cost to build up the interface conditions on* Γ *in (4.6) does not depend on the number of particles* N*.*

Notice that here, we describe the splitting scheme in (4.5) – (4.6) just for illustration of the applicability of the RS tensor format for the solution of the PBE. This scheme requires modification of the interface conditions that is equivalent to a change of the system matrix which leads to a complicated implementation scheme. To avoid this nontrivial task, in what follows, we introduce an alternative approach, which will be discussed in the next section.

The regularization $u = u^r + u_0^m$ like in (4.5) – (4.6) benefits from the local-global separability in the low-rank RS tensor representation of the free space electrostatic potential. Notice that here, we describe the splitting scheme in (4.5) – (4.6) just for illustration of the applicability of the RS tensor format for the solution of the PBE. This scheme requires modification of the interface conditions that is equivalent to a change of the system matrix which leads to a complicated implementation scheme. To avoid this nontrivial task, in what follows, we introduce an alternative approach, which avoids the additional computation of the auxiliary harmonic function u^{harm} in the rather complicated domain Ω_m as well as the modification of the interface conditions.

4.2 The new RS tensor based splitting scheme

In this section, we present the new splitting scheme which is based on the range separated representation of the Dirac δ-distribution [21], which composes the highly singular right-hand side in the target PBE (4.1) or Poisson equation (PE) (4.5). Here, we consider the PE as proof of concept and validate the numerical results in Section 5. The idea is to modify the right-hand side ρ_f in such a way that the short-range part in the solution u can be pre-computed independently by the direct tensor decomposition of the free space potential, and the initial elliptic equation applies only to the long-range part of the total potential, satisfying the equation with the modified right-hand side by using the RS splitting of the Dirac delta, see [21] for more details. The latter is a smooth function, hence the FDM/FEM approximation error can be reduced dramatically even on relatively coarse grids in 3D.

For ease of presentation, we consider the simplest case of the single atom with unit charge located at the origin, such that the exact electrostatic potential reads $u(\bar{x}) = \frac{1}{\|x\|}$, $x \in \mathbb{R}^3$. Recall that the Newton kernel (3.3) discretized by the R-term sum of Gaussian type functions living on the $n \times n \times n$ tensor grid Ω_n is represented by a sum of short- and long-range tensors,

$$\frac{1}{\|x\|} \leadsto \mathbf{P}_R = \mathbf{P}_{R_s} + \mathbf{P}_{R_l} \in \mathbb{R}^{n \times n \times n},$$

where $\mathbf{P}_{R_s}$ and $\mathbf{P}_{R_l}$ are defined in (3.4).

Let us formally discretize the exact equation for the Newton potential, $u(\bar{x}) = \frac{1}{\|x\|}$,

$$-\Delta \frac{1}{\|x\|} = 4\pi \, \delta(\bar{x}),$$

by using the FDM/FEM Laplacian matrix A_Δ instead of Δ and via substitution of the canonical tensor decomposition $\mathbf{P}_R$ instead of $u(\bar{x}) = \frac{1}{\|x\|}$. This leads to the grid representation of the discretized Dirac delta [21]

$$\delta(\bar{x}) \rightsquigarrow \boldsymbol{\delta}_h := -\frac{1}{4\pi} A_\Delta \mathbf{P}_R,$$

that will be applied in the framework of our discretization scheme.

We remind that the 3D finite difference Laplacian matrix A_Δ, defined on the uniform rectangular grid, takes the form

$$A_\Delta = \Delta_1 \otimes I_2 \otimes I_3 + I_1 \otimes \Delta_2 \otimes I_3 + I_1 \otimes I_2 \otimes \Delta_3, \tag{4.9}$$

where $-\Delta_\ell = h_\ell^{-2}\text{tridiag}\{1, -2, 1\} \in \mathbb{R}^{n_\ell \times n_\ell}$, $\ell = 1, 2, 3$, denotes the discrete univariate Laplacian, such that the Kronecker rank of A_Δ equals to 3. Here I_ℓ, $\ell = 1, 2, 3$, is the identity matrix in the corresponding single dimension.

Now we are in the position to describe the RS tensor based splitting scheme. To that end, we use the splitting of the discretized δ-distribution into short- and long-range components in the form [21],

$$\boldsymbol{\delta}_h = \boldsymbol{\delta}_s + \boldsymbol{\delta}_l, \tag{4.10}$$

where

$$\boldsymbol{\delta}_s := -\frac{1}{4\pi} A_\Delta \mathbf{P}_{R_s}, \quad \text{and} \quad \boldsymbol{\delta}_l := -\frac{1}{4\pi} A_\Delta \mathbf{P}_{R_l}. \tag{4.11}$$

We recall that by construction, the short range part vanishes on the interface Γ, hence it satisfies the discrete Poisson equation in Ω_m with the respective right-hand side in the form $\boldsymbol{\delta}_s$ and zero boundary conditions on Γ [21]. Then we deduce that this equation can be subtracted from the full discrete linear system, such that the long-range component of the solution, $\mathbf{P}_{R_l}$, will satisfy the same linear system of equations (same interface conditions), but with a modified right-hand side corresponding to the weighted sum of the long-range tensors $\boldsymbol{\delta}_l$ only. In the simple example of the single charge, we arrive at the particular discrete Poisson equation for the long-range part in the full potential $\mathbf{P}_R$, $\mathbf{U}_l = \mathbf{P}_{R_l}$,

$$-A_\Delta \mathbf{U}_l = \boldsymbol{\delta}_l, \tag{4.12}$$

which can be solved by an appropriate iterative method.

Figure 4.1 illustrates the modified right hand side representing the long-range part of the discrete Dirac delta $\boldsymbol{\delta}_l$. It it worth noting that the FEM approximation theory can be applied to this formulation since the input data (i.e., the right-hand side) are regular enough. However, it is not the case for the initial formulation with the highly singular Dirac delta distribution in the right-hand side.
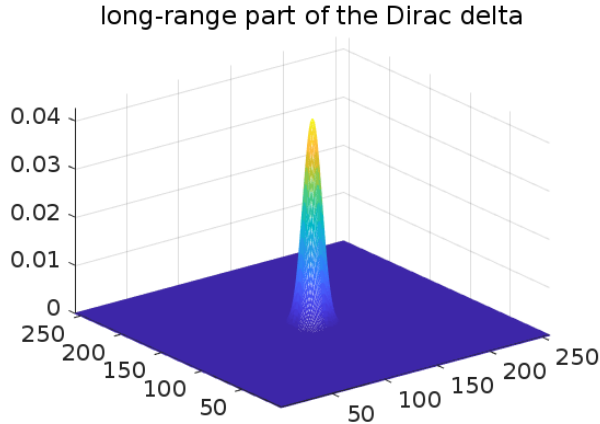
Figure 4.1: The long-range part of the Dirac delta $\boldsymbol{\delta}_l$ on an $n^{\otimes 3}$ 3D grid, $n = 256$.

This scheme can be easily extended to the case of many-atomic systems just by additive representation of the short- and long-range parts in the total free space potential,

$$- A_\Delta \mathbf{P}_{R_l} = \boldsymbol{\delta}_{R_L}, \tag{4.13}$$

where we suppose that R_L is the rank of the long-range part $\mathbf{P}_{R_l}$ of the corresponding RS tensor of type (3.9), and $\boldsymbol{\delta}_{R_L}$ is calculated as shown in (4.15).

We summarize the following benefits of the aforementioned solution decomposition scheme.

- Most important is that due to efficient splitting of the short- and long-range parts in the target tensor representing both the single Newton kernel and the total free-space potential, there is no need to modify jump conditions at the interface.

- A remarkable advantage is that the long-range part in the RS tensor decomposition of the Dirac-delta distribution [21] vanishes at the interface and, hence, the modified right-hand side generated by this long-range component remains localized in the "linear" solute region.

- The boundary conditions are obtained from the long-range part in the tensor representation of the collective electrostatic potential which reduces the computational costs involved in solving some external analytical function at the boundary.

- Only a single system of algebraic equations is solved for the smooth long-range (i.e., regularized) part of the collective potential discretized with controllable precision on a relatively coarse grid, which is then added to the directly precomputed (avoiding PDE solutions) low-rank tensor representation for the short-range contribution.

Next, we briefly comment on the general FEM approximation for the Laplacian. We recall the tensor-based scheme for evaluation of the Laplace operator in a separable basis set [19] applied for calculation of the kinetic energy part in the Fock operator.

Let the problem be posed in the finite volume box $\Omega = [-b, b]^3 \in \mathbb{R}^3$, subject to the homogeneous Dirichlet boundary conditions on $\partial\Omega$. For given discretization parameter $n \in \mathbb{N}$,

15

the equidistant $n \times n \times n$ tensor grid $\omega_{\mathbf{3},N} = \{x_\mathbf{i}\}$, $\mathbf{i} \in \mathcal{I} := \{1, ..., n\}^3$ is used, with the mesh-size $h = 2b/(n+1)$. Then the Kronecker rank-3 tensor representation of the respective Galerkin FEM stiffness matrix is given by (4.9).

Notice that the MATLAB representation of the matrix A_Δ (say, the FD matrix) can be easily described in terms of `kron` operations as follows

$$\frac{1}{h^2}A_\Delta = \texttt{kron}(\texttt{kron}(\Delta_1, I), I) + \texttt{kron}(\texttt{kron}(I, \Delta_1, I) + \texttt{kron}(\texttt{kron}(I, I), \Delta_1), \qquad (4.14)$$

applied to a long vector of size n^3 representing the Newton potential.

Then the rank-structured calculation of the "collective" right-hand side $\boldsymbol{\delta}_{R_L}$ in (4.13) is reduced to one-dimensional operations,

$$-\boldsymbol{\delta}_{R_L} = \sum_{k=1}^{R_L} \xi_k (\Delta_1 \mathbf{a}_k^{(1)} \otimes \mathbf{a}_k^{(2)} \otimes \mathbf{a}_k^{(3)} + \mathbf{a}_k^{(1)} \otimes \Delta_1 \mathbf{a}_k^{(2)} \otimes \mathbf{a}_k^{(3)} + \mathbf{a}_k^{(1)} \otimes \mathbf{a}_k^{(2)} \otimes \Delta_1 \mathbf{a}_k^{(3)}), \quad (4.15)$$

where $\mathbf{a}_k^{(\ell)}$, $\ell = 1, 2, 3$, are the canonical vectors and R_L is the canonical rank of the long-range part of the collective electrostatic free space potential of a biomolecule computed in the RS tensor format (3.9).

This is the tensor ansatz to be used as the right-hand side in the equation (4.12), which we apply in numerical experiments. With a subsequent usage of the canonical-to-full tensor transform and after reshaping a three-tensor into a long vector, $\boldsymbol{\delta}_{R_L}$ is applied in a standard PBE iterative solver as the RHS for the long-range part. Another advantage of our scheme is that the short-range part of the solution in the PBE (2.1) is obtained for free, since it is merely incorporated as the set of short-range parts of the respective Newton potentials for every particle in a biomolecule. That corresponds to a set of tensors in the second term of the collective electrostatic potential in the RS tensor format (3.9).

4.3   Discussion of the computational scheme

We summarize the main computational tasks involved in the presented tensor-based numerical scheme. The regularized PBE (RPBE) can be constructed as follows:

(a) Compute the regularized component $\boldsymbol{\delta}_{R_L}$ of the discretized Dirac delta distribution in (4.13) as described in [21];

(b) Substitute $\boldsymbol{\delta}_{R_L}$ into the right-hand side of the PBE (4.1) to obtain the following RPBE

$$-\nabla \cdot (\epsilon \nabla u^r(\bar{x})) + \bar{\kappa}_2{}^2(\bar{x})u^r(\bar{x}) = \boldsymbol{\delta}_{R_L}, \quad \text{in } \Omega, \qquad (4.16)$$

subject to the boundary condition in (2.2).

(c) Discretize the RPBE (4.16) to obtain the following linear system of equations

$$Au^r = b, \qquad (4.17)$$

which can be solved by any linear system solver.

(d) Obtain the final PBE solution u by the sum

$$u = u^r + u^s,$$

where $u^s = \mathbf{P}_{R_s}$ is the precomputed short-range component of the Newton potential sum, see (3.4).

The splitting scheme described in Section 4.2 allows to reduce the initial equation to the solution of the system with modified right-hand side by using the range-separated representation of the discretized Dirac delta. The problem is reduced to the direct tensor-based computation (without solving a PDE) of the short-range part in the collective free space electrostatic potential, see (3.4), and to the subsequent solution of the PBE equation for the long-range part only by the simple modification of the right-hand side. The advantage is that the PBE applies to the smooth part, u^r, of the total potential and hence a controllable FDM/FEM approximation error on moderate size 3D grids can be achieved.

This method can be combined with the reduced basis approach for PBE with parametric coefficients to further accelerate the numerical computations [4, 25]. This is because the modified model is affinely dependent on the parameter $\bar{\kappa}$, thereby providing a natural offline/on-line decomposition of the reduced basis method.

Finally, we notice that the important characterization of the protein molecule is given by the electrostatic solvation energy [28], which is the difference between the electrostatic free energy in the solvated state (described by the PBE) and the electrostatic free energy in the absence of solvent. Having at hand the free energy E_N (see [5], Lemma 4.1), the electrostatic solvation energy can be computed in the framework of the regularized formulation of the PBE as described above.

5   Numerical Tests

In this section, we consider only the free space electrostatic potential for the modified PE and the RS tensor format based splitting scheme. We compare the results with those of the traditional PE for various biomolecules. Notice that the PBE can be reduced to the PE by considering the case of zero ionic strength which implies that the function $\bar{\kappa}^2(\bar{x})$, hence the Boltzmann distribution term, in (2.1) is annihilated. Consequently, homogeneous dielectric constants of $\epsilon_m = \epsilon_s = 1$ are considered. We compute the electrostatic potentials using $n \times n \times n$ 3D Cartesian grids, in a box $[-b, b]^3$ with equal step size $h = \frac{2b}{n-1}$. Note that computations by the PBE/PE solver are bounded by $n = 257$, on a PC with 8GB RAM due to the storage needs of the order of $\mathcal{O}(n^3)$. Note that since the storage requirements for the tensor-structured representation of the electrostatic potential is of the order of $\mathcal{O}(n)$, see (3.3), in tensor computations of molecular potentials one can use much finer 3D grids.

First, we validate the FDM solver for the PE by comparing its solution with that of the adaptive Poisson-Boltzmann solver (APBS) using the finite difference multigrid calculations with PMG [2] option [1, 9]. Here, we consider a Born ion of unit charge, unit (Å) radius and located at the origin $((0,0,0))$. Figure 5.1 shows the single Newton kernel for the Born ion, approximated by the PE on an $n \times n$ grid surface with $n = 129$ at the cross section of the

---

[2]PMG is a Parallel algebraic Multigrid code for General semilinear elliptic equations.
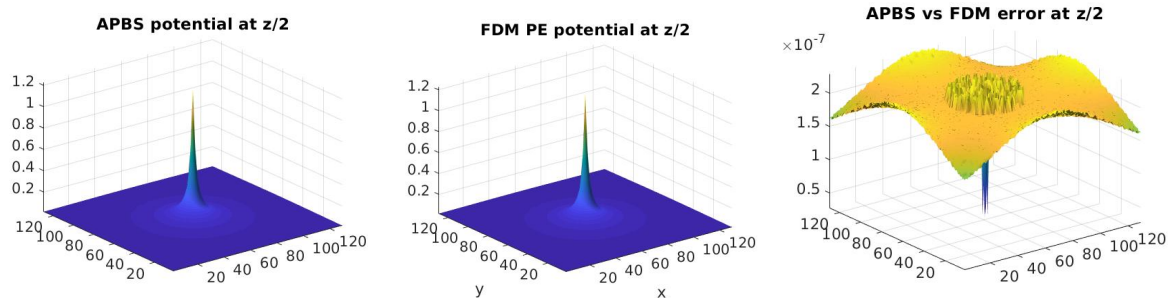
Figure 5.1: The free space potential for the Born ion computed by the APBS (left), the FDM solver (middle) and the corresponding error (right).

volume box in the middle of the z-axis computed by the APBS and the FDM solvers and the corresponding error between the two solutions. The results show that the FDM solver provides as accurate results as those of the APBS with a discrete L_2 error of $\mathcal{O}(10^{-9})$ in the full solution. Similar results for varying proteins are illustrated in [4].

Secondly, we compare the accuracy of the traditional PE model and of the PE model modified by the RS tensor format for the approximation of the single Newton kernel. Figure 5.2 shows the single Newton kernel on an $n \times n$ grid surface with $n = 129$ at the cross section of the volume box in the middle of the z-axis computed by the canonical tensor approximation obtained by sinc-quadratures and the corresponding errors by the traditional PE model and the modified PE model computed by the FDM solver. We notice that the solution of the modified PE model is of higher accuracy than that of the traditional PE model because it captures the singularities exactly, see the central region of Figure 5.2 (right).

In a similar vein, we consider the acetazolamide compound consisting of 18 atoms and determine the accuracy of the traditional PE model vis a vis the PE model modified by the RS tensor format. Figure 5.3 shows the sum of the electrostatic potentials for the acetazolamide, computed on an $n \times n$ grid surface with $n = 129$ by the canonical tensor approximation obtained by sinc-quadratures and the corresponding errors by the traditional PE model and the modified PE model computed by the FDM solver. It is clearly shown that the modified PE model provides highly accurate solutions as compared to those of the traditional PE due to the accurate treatment of the solution singularities by the RS tensor format.
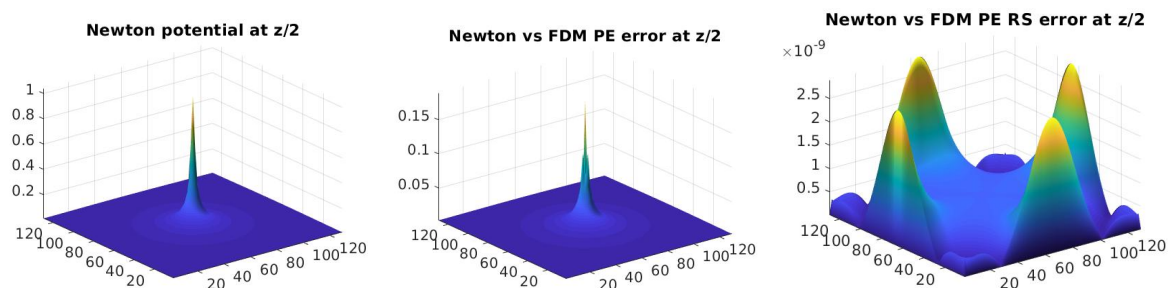
Figure 5.2: The Newton potential computed by the canonical tensor decomposition (left), the error of its computation on the same grid by using the classical PE (middle) and by the modified PE (right).

18
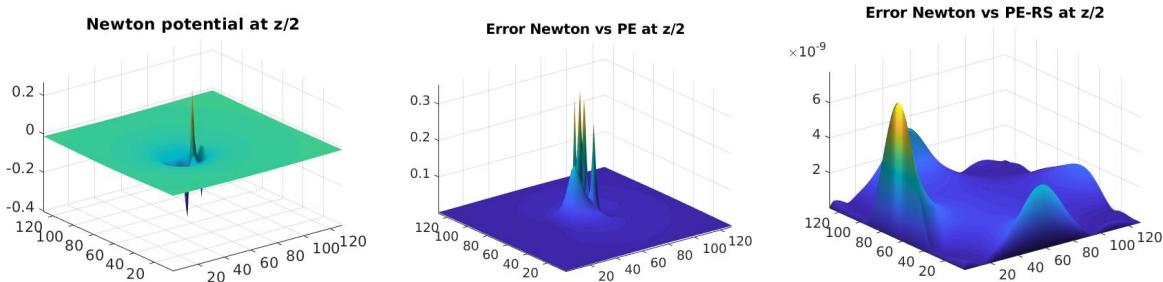
Figure 5.3: The Newton potential sums computed by the canonical tensor decomposition (left), the error of its computation on the same grid by using the classical PE (middle) and by the modified PE (right).
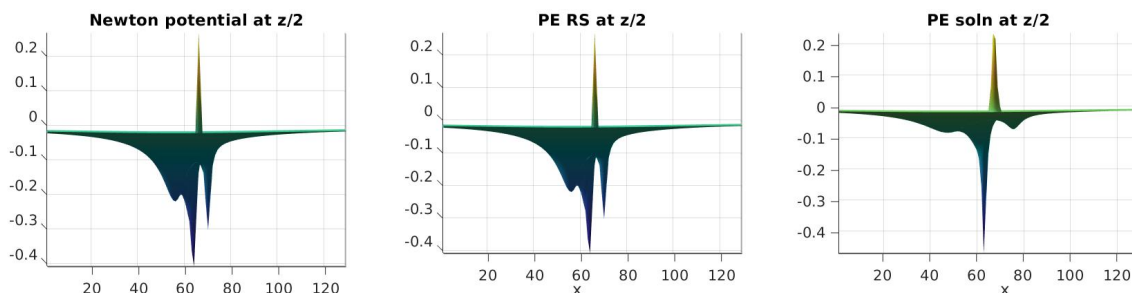

Figure 5.4: Demonstration of the solution singularities for the acetazolamide molecule captured by the canonical tensor approximation (left), by the modified PE model (middle) and by the classical PE (right).

We notice, in addition, that the classical PE model does not capture accurately the singularities in the electrostatic potential due to the numerical errors introduced by the Dirac delta distribution and partly due to the smoothing effect caused by the spline interpolation of the charges onto the grid. The modified PE model, on the other hand, is able to capture the singularities due to the independent treatment of the singularities by the RS tensor technique. This is clearly demonstrated in Figure 5.4.

5.1 Solutions to the modified Poisson equation on a sequence of fine grids

Here, we illustrate the accuracy of the modified PE by calculating the free space electrostatic potential on a sequence of fine grids and compare with the solution of the exact Newton potential determined by the canonical tensor representation. We first consider the aforementioned Born ion case and show the absolute error for the finest Cartesian grid and the discrete L_2 norm of the error for a sequence of Cartesian grids. Figure 5.5 shows the absolute error of $\mathcal{O}(10^{-11})$ obtained on a 257^3 Cartesian grid and $32\,\text{Å}$ box length. Table 5.1 shows the decay of the discrete L_2 norm of the error for a sequence of grid refinements.

Next, we consider the acetazolamide molecule with 18 atoms. This molecule is used as a ligand in the human carbonic anhydrase (hca) protein-ligand complex for the calculation of the binding energy in the adaptive Poisson-Boltzmann software (APBS) package [14] and
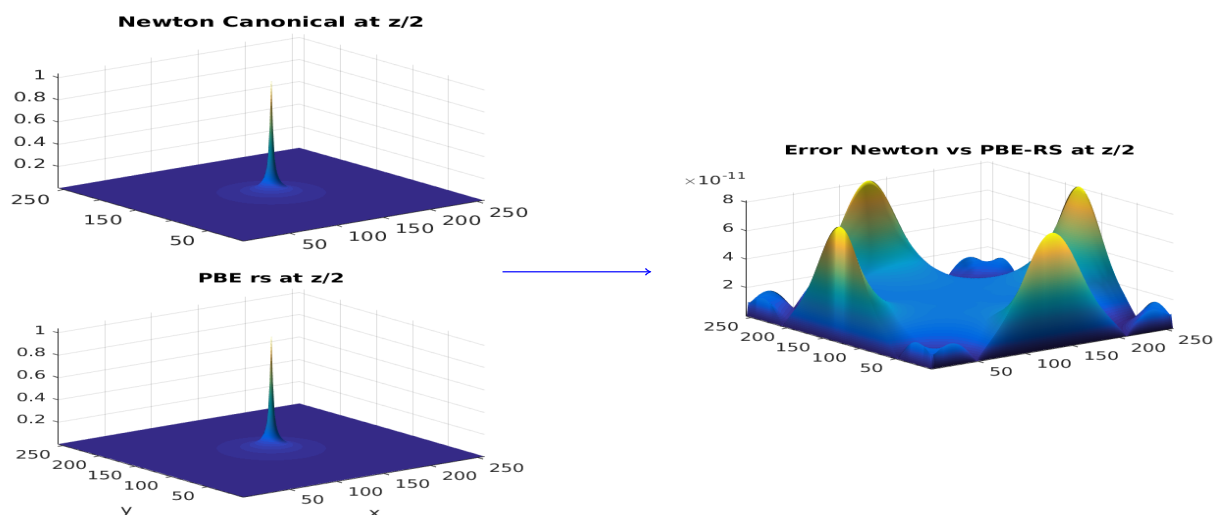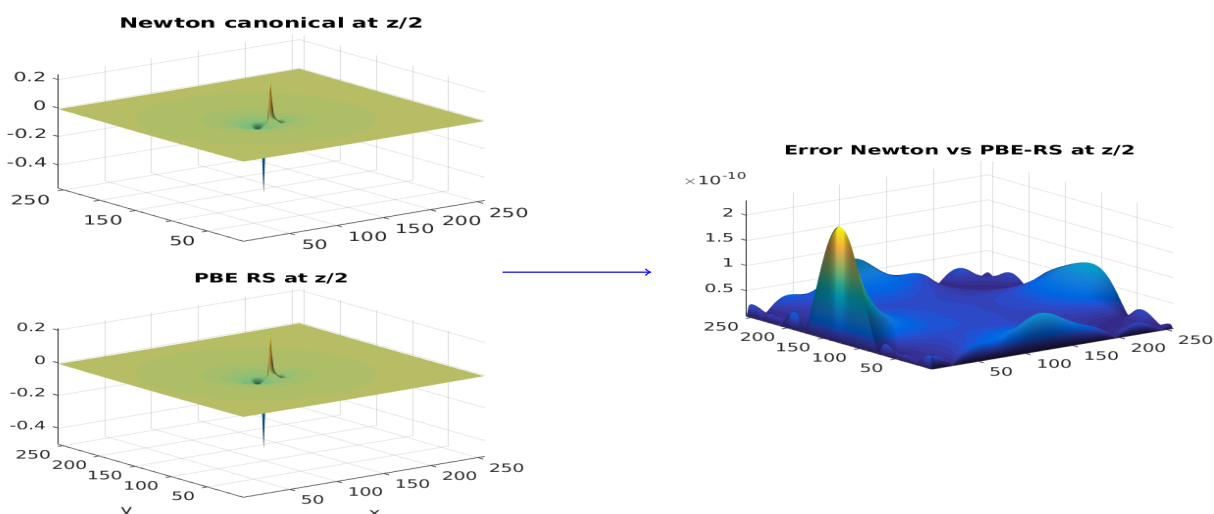
Figure 5.5: Absolute error between the solutions of the Newton potential and the modified Poisson equation for the Born ion.

n	97	129	257
Discrete L_2 norm	4.1176×10^{-7}	9.5516×10^{-8}	2.7975×10^{-9}

**Table 5.1:** The discrete L_2 norm of the error with respect to grid size for the Born ion.

the MATLAB program for biomolecular electrostatic calculations, (MPBEC) [33]. The electrostatic potential is computed as in the previous case, employing the same grid properties. The absolute error is shown in Figure 5.6 and the error behaviour with respect to mesh refinements is shown in Table 5.2.

Figure 5.6: Absolute error between the solutions of the Newton potential sums and the modified Poisson equation for the acetazolamide molecule.

n	97	129	257
Discrete L_2 norm	4.1176×10^{-7}	1.1936×10^{-7}	3.7003×10^{-9}

Table 5.2: The discrete L_2 norm of the error and the relative error with respect to grid size for the acetazolamide molecule.
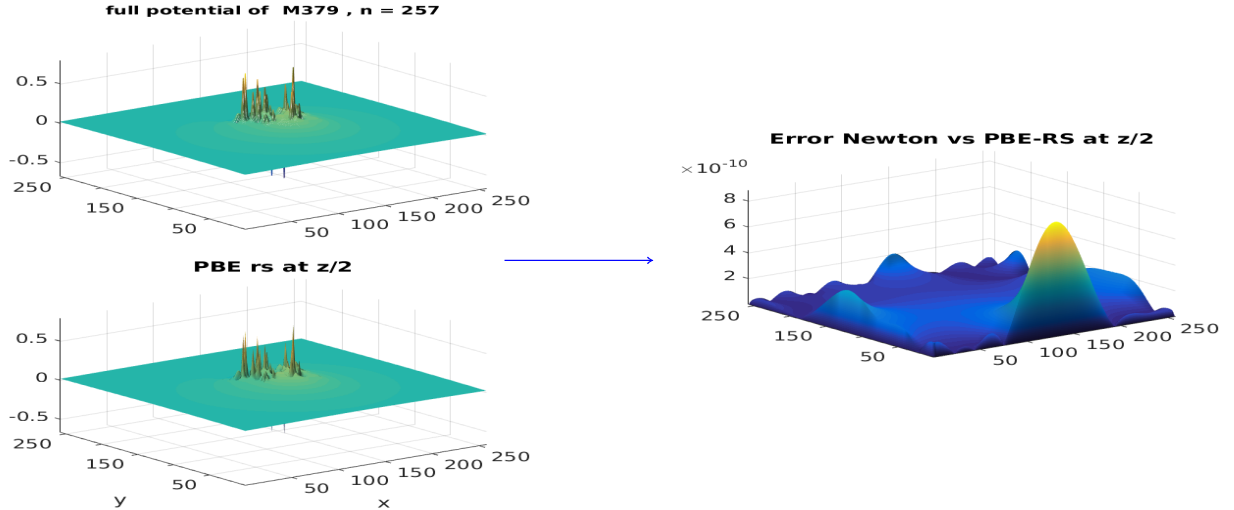
Figure 5.7: Absolute error between the solutions of the Newton potential sums and the modified Poisson equation for the protein Fasciculin 1.

Finally, we consider the protein Fasciculin 1, with 1228 atoms, an anti-acetylcholinesterase toxin from the green mamba snake venom [26]. Again, we compute the electrostatics potential as in the previous case, but with 60 Å box length and a 257^3 as the minimum Cartesian grid because of a larger molecular size. The results provided in the Figure 5.7 and Table 5.3 illustrate a similar trend of accuracy as in the previous test examples.

5.2 Accurate representation of the long-range electrostatic potential by the RS tensor

Here, we highlight the advantages of the RS tensor format in the low-rank approximation of the long-range component in the total potential sum. For this purpose, the RHOSVD within the multigrid C2T transform [23] is used which provides computation of the low-rank canonical/Tucker tensor representation of the long-range part at the asymptotic cost of $\mathcal{O}(Nn)$. Here, N is the number of charges in the molecule while n represents the grid

n	129	193	257
Discrete L_2 norm	1.2919×10^{-6}	1.7395×10^{-7}	4.3060×10^{-8}

Table 5.3: The discrete L_2 norm of the error and the relative error with respect to grid size for the protein Fasciculin 1.
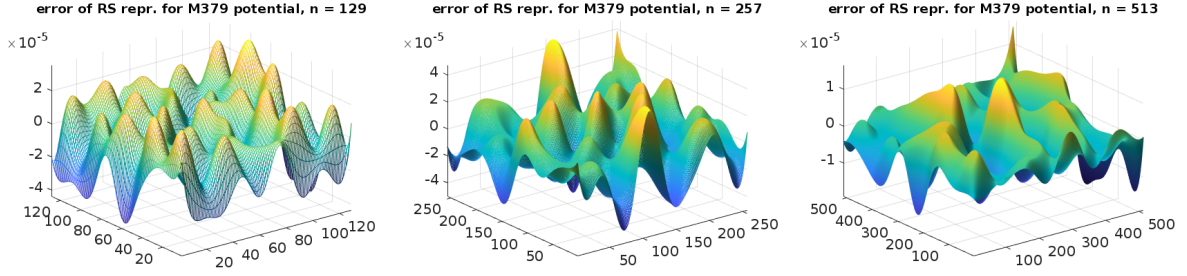
Figure 5.8: The error due to the low-rank approximation of the long-range component for the 379 atomic molecule for $n = 129^3$ (left), $n = 257^3$ (middle) and $n = 513^3$ (right) grids.
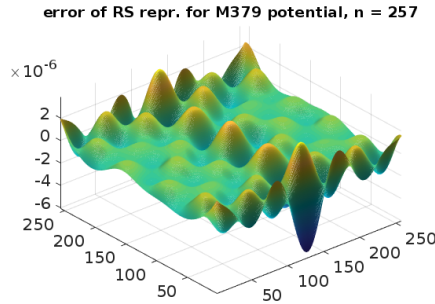
Figure 5.9: The error due to the low-rank approximation of the long-range component for the 379 atomic molecule for $n = 257^3$ grid at a lower tolerance.

dimension in a single direction. Figure 5.8 shows an error of $\mathcal{O}(10^{-5})$ of the RS tensor format approximation compared with the full size representation at various grid dimensions. These data correspond to the long-range RS-rank equal to 10, with ε-truncation threshold chosen as $\mathcal{O}(10^{-6})$ for the Newton kernel and $\mathcal{O}(10^{-7})$ for the C2T transform. Figure 5.9 shows the error obtained if we take another rank truncation criterion ε of an order less for both the Newton (i.e., $\mathcal{O}(10^{-7})$) and for the C2T transform ($\mathcal{O}(10^{-8})$), indicating that the error will be one order less too, i.e., $\mathcal{O}(10^{-6})$).

6 Conclusions

In this paper we demonstrate that the range-separated tensor format is gainfully applicable for the solution of the PBE for calculation of electrostatics in large molecular systems. The efficiency of the new tensor-based regularization scheme for the PBE is based on the exceptional properties of the grid-based RS tensor splitting of the Dirac-delta distribution. The main computational benefits are due to the localization of the modified right-hand side within the molecular region and automatic maintaining of the continuity of the Cauchy data on the interface. Another advantage is that our computational scheme only includes solving a single system of algebraic equations for the smooth long-range (i.e., regularized) part of the collective potential discretized by FDM. The total potential is obtained by adding this solution to the directly precomputed low-rank tensor representation for the short-range contribution.

The various numerical tests illustrate the main properties of the presented scheme. For example, it is clear from Figure 5.4 that the classical PE model does not accurately capture the solution singularities which emanate from the short-range component of the total target electrostatic potential in the numerical approximation. In this paper, we emphasize that this problem can be efficiently circumvented by applying the range-separated tensor format as a solution decomposition technique in order to modify the PBE/PE. In the modified PBE/PE, the Dirac-delta distribution is replaced by a smooth long-range function from (4.11). We thus only need to solve for the long-range electrostatic potential numerically and add this solution to the short-range component which is computed a priori using the canonical tensor approximation to the Newton kernel. The resultant total potential sum is of high accuracy as evident from Figures 5.5 - 5.7 and Tables 5.1 - 5.3. Finally, we summarize that the regularization scheme presented in this paper has capabilities for various generalizations which can be effectively implemented with minor changes in the RS tensor decompositions. We notice the following directions:

– The possibility of PBE computations on much finer grids is also an advantage of the proposed approach. Tensor techniques practically may be applied to finer grids compared to traditional finite element approaches;

– The regularization scheme remains verbatim in the case of the nonlinear PBE since it requires only the modification of input data for the right-hand side in the molecular region Ω_m, where the equation is linear, but this does not affect the nonlinearity domain Ω_s;

– Our approach allows the efficient calculation of electrostatics under multiple rotations of the biomolecule, which is the crucial problem in the numerical modeling of proteins.

References

[1] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Nat. Acad. Sci. U.S.A.*, 98(18):10037–10041, 2001.

[2] V. Barone, M. Cossi, and J. Tomasi. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.*, 107:3210–3221, 1997.

[3] D. Bashford and D. A. Case. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, 51:129–152, 2000.

[4] P. Benner, L. Feng, C. Kweyu, and M. Stein. Fast solution of the Poisson-Boltzmann equation with nonaffine parametrized boundary conditions using the reduced basis method. *arXiv:1705.08349*, 221, 2017.

[5] P. Benner, V. Khoromskaia, and B. N. Khoromskij. Range-separated tensor formats for numerical modeling of many-particle interaction potentials. *arXiv:1606.09218v3*, pages 1–38, 2016.

[6] P. Benner, V. Khoromskaia, and B. N. Khoromskij. Range-separated tensor format for many-particle modeling. *SIAM J. Sci. Comp.*, (2):A1034–A1062, 2018.

[7] C. Bertoglio and B. N. Khoromskij. Low-rank quadrature-based tensor approximation of the Galerkin projected Newton/Yukawa kernels. *Comp. Phys. Comm.*, 183(4):904–912, 2012.

[8] E. Cances, Y. Maday, and B. Stamm. Domain decomposition for implicit solvation models. *J. Chem. Phys.*, 139:054111, 2013.

[9] L. Chen, M. Holst, and J. Xu. The finite element approximation of the nonlinear Poisson-Boltzmann equation. *SIAM J. Numer. Anal.*, 45(6):2298–2320, 2007.

[10] M. Deserno and C. Holm. How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *J. Chem. Phys.*, 109(18):7678–7693, 1998.

[11] I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij. Hierarchical tensor-product approximation to the inverse and related operators in high-dimensional elliptic problems. *Computing*, 74:131–157, 2005.

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

[13] M. Holst, N. Baker, and F. Wang. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation: algorithms and examples. *J. Comp. Chem.*, 21:1319–1342, 2000.

[14] M. Holst and F. Saied. Multigrid solution of the Poisson-Boltzmann equation. *J. Comput. Chem.*, 14:105–113, 1993.

[15] M. J. Holst. *Multilevel methods for the Poisson-Boltzmann equation*. Ph.D. Thesis, Numerical Computing group, University of Illinois, Urbana-Champaign, IL, USA, 1994.

[16] P. H. Hünenberger and J. A. McCammon. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study. *Biophys. Chemistry*, 78:69–88, 1999.

[17] V. Khoromskaia and B. N. Khoromskij. Grid-based lattice summation of electrostatic potentials by assembled rank-structured tensor approximation. *Comp. Phys. Comm.*, 185(12), 2014.

[18] V. Khoromskaia and B. N. Khoromskij. Tensor numerical methods in quantum chemistry: from Hartree-Fock to excitation energies. *Phys. Chem. Chem. Phys.*, 17:31491 – 31509, 2015.

[19] V. Khoromskaia, B. N. Khoromskij, and D. Andrae. Fast and accurate 3D tensor calculation of the Fock operator in a general basis. *Comp. Phys. Comm.*, 183(11), 2012.

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

[21] B. N. Khoromskij. Range-separated tensor representation of the discretized multidimensional Dirac delta and elliptic operator inverse. *Preprint*, arXiv:1812.02684v1, 2018.

[22] B. N. Khoromskij. *Tensor numerical methods in scientific computing*. De Gruyter, Berlin, 2018.

[23] B. N. Khoromskij and V. Khoromskaia. Multigrid accelerated tensor approximation of function related multidimensional arrays. *SIAM J. Sci. Comp.*, 31(4):3002–3026, 2009.

[24] T. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[25] C. Kweyu, M. Hess, L. Feng, M. Stein, and P. Benner. Reduced basis method for Poisson-Boltzmann Equation. In M. Papadrakakis, V. Papadopoulos, G. Stefanou, and V. Plevris, editors, *ECCOMAS Congress 2016 - Proc. of the VII European Congress on Computational Methods in Applied Sciences and Engineering*, volume 2, pages 4187–4195, Athens, 2016. National Technical University of Athens.

[26] M. le Du, P. Marchot, P. Bougis, and J. Fontecilla-Camps. 1.9 Angstrom resolution structure of fasciculine 1, an anti-acetylcholinesterase toxin from green mamba snake venom. *J. Biol. Chem.*, 267:22122–22130, 1992.

[27] F. Lipparini, B. Stamm, E. Cances, Y. Maday, and B. Mennucci. Domain decomposition for implicit solvation models. *J. Chem. Theor. Comp.*, 9:3637–3648, 2013.

[28] B. Z. Lu, Y. C. Zhou, M. J. Holst, and J. A. McCammon. Recent progress in numerical methods for Poisson-Boltzmann equation in biophysical applications. *Commun. Comp. Phys.*, 3(5):973–1009, 2008.

[29] M. Mirzadeh, M. Theillard, A. Helgadottir, D. Boy, and F. Gibou. An adaptive, finite difference solver for the nonlinear Poisson-Boltzmann equation with applications to biomolecular computations. *Commun. Comput. Phys.*, 13(1):150–173, 2013.

[30] E. L. Pollock and J. Glosli. Comments on P(3)m, FMM and the Ewald method for large periodic Coulombic systems. *Comp. Phys. Comm.*, 95:93–110, 1996.

[31] A. Smilde, R. Bro, and P. Geladi. *Multi-way analysis with applications in the chemical sciences*. Wiley, 2004.

[32] M. Stein, R. R. Gabdoulline, and R. C. Wade. Cross-species analysis of the glycoliticmpathway by comparison of molecular interaction fields. *Molecular Biosystems*, 6:162–174, 2010.

[33] S. Vergara-Perez and M. Marucho. MPBEC, a Matlab program for biomolecular electrostatic calculations. *Comput. Phys. Commun.*, 198:179–194, 2016.

[34] D. Xie. New solution decomposition and minimization scheme for Poisson-Boltzmann equation in calculation of biomolecular electrostatics. *J Comp. Phys.*, 275:294–309, 2014.