Fast convolution solver based on far-field smooth approximation

Xin Liu [ID] [a,b], Yong Zhang [ID] [b,c,*]

[a] *Eastern Institute for Advanced Study, Eastern Institute of Technology, Ningbo, 315200, China*
[b] *Center for Applied Mathematics and KL-AAGDM, Tianjin University, Tianjin, 300072, China*
[c] *State Key Laboratory of Synthetic Biology, Tianjin University, Tianjin, 300072, China*

A B S T R A C T

The convolution potential arises in a wide variety of application areas, and its efficient and accurate evaluation encounters three challenges: singularity, nonlocality and anisotropy. We introduce a fast algorithm based on a far-field smooth approximation of the kernel, where the bounded domain Fourier transform, one of the most essential difficulties, is well approximated by the whole space Fourier transform which usually admits explicit formula. The convolution is split into a regular and singular integral, and they are well resolved by trapezoidal rule and Fourier spectral method respectively. The scheme is simplified to a discrete convolution and is implemented efficiently with Fast Fourier Transform (FFT). Importantly, the tensor generation procedure is quite simple, highly efficient and independent of the anisotropy strength. It is easy to implement and achieves spectral accuracy with nearly optimal efficiency and minimum memory requirement. Rigorous error estimates and extensive numerical investigations, together with a comprehensive comparison, showcase its superiorities for different kernels.

1. Introduction

Nonlocal potentials, which are given by a convolution of a translational invariant Green function with a fast-decaying smooth function, are common and have wide-ranging applications. Examples include the Newtonian potential in cosmology, the Poisson potential in electrostatics, plasma physics and quantum physics [1–5]. The efficient and accurate calculation of these nonlocal potentials is a prominent and vital area of research in the science and engineering community. In this paper, we focus on the evaluation of convolution-type nonlocal potential

$$\Phi(\mathbf{x}) = \int_{\mathbb{R}^d} U(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{U}(\mathbf{k})\widehat{\rho}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}\mathbf{k}, \quad \mathbf{x} \in \mathbb{R}^d, \tag{1.1}$$

where d is the ambient dimension, the density $\rho(\mathbf{x})$ is a fast-decaying smooth function, the kernel $U(\mathbf{x})$, a given radially symmetric function, is usually singular at the origin, and $\widehat{f}(\mathbf{k}) = \int_{\mathbb{R}^d} f(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}\mathbf{x}$ is Fourier transform of $f(\mathbf{x})$.

There are three major challenges for this convolution-type potential evaluation.

1. **Singularity:** The kernel and its Fourier transform are both singular, and sometimes the singularity of $\widehat{U}(\mathbf{k})$ are even stronger.
2. **Nonlocality:** The potential value at a fixed *target* point $\mathbf{x}$ depends on the density and kernel at every *source* point $\mathbf{y}$.
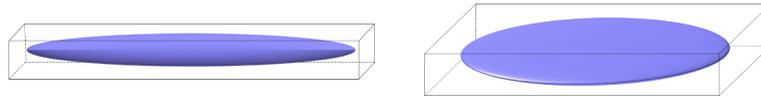

Fig. 1. Cigar-shaped (left) and pancaked-shaped (right) densities.

3. **Anisotropy:** The density may be highly anisotropic. For example, in lower-dimensional confined quantum systems [1,6], its compact support extends much shorter in one or two directions, as shown in Fig. 1 with the "cigar-shaped" and the "pancake-shaped" densities.

Since the density decays rapidly enough, it is reasonable to assume that the density is compactly supported (to a fixed precision) in a bounded domain. In computational practice, we truncate the whole space to a rectangular domain $\mathcal{D} := \prod_{j=1}^{d}[-L_j, L_j]$ and discretize it with equally spaced points in each direction. The smooth density $\rho(\mathbf{x})$ is well approximated by Fourier spectral method with spectral accuracy, achieving nearly optimal efficiency thanks to implementation using discrete Fast Fourier Transform (FFT) [7].

The numerical problem is to compute the convolution Φ on uniform grid from discrete density values given on the same grid. The scheme is expected to achieve spectral accuracy with great efficiency, and capable of dealing with highly anisotropic density case.

Over the past decade, several fast and accurate methods have been developed based on Fourier spectral method. The *NonUniform Fast Fourier Transform* (NUFFT)-based method [8–10], the first accurate fast algorithm, was implemented via the NUFFT algorithm. Later, the *Gaussian-Summation method* (GauSum) [11], emerged as the first purely FFT-based algorithm, dealt with the singularity using a summation-of-Gaussian (SOG) approximation of the kernel that is done away from the origin. In 2016, a much simpler method, the *kernel truncation method* (KTM), is proposed by Vico *et al.* [12], where they remove the singularity by truncating the kernel within a radially symmetric domain. In KTM, one has to zero-pad the density by a factor, that is no small than $\sqrt{d}+1$ [13–15], in each spatial direction so as to capture the unpleasant oscillations brought by the discontinuous kernel truncation. Unfortunately, both memory requirement and computation costs scale linearly with the anisotropy strength, and it places a huge burden in higher space dimension, especially the three-dimension problem.

Following the kernel truncation line, Greengard *et al.* [16] proposed the *anisotropic truncated kernel method* (ATKM), where the kernel was effectively truncated within an anisotropic double-sized rectangular geometry. The computation of the Fourier transform, defined on the anisotropic domain, is even more challenging and successfully precomputed via the **near-field** approximation SOG that is performed much closer to the origin compared with GauSum. Even though GauSum and ATKM achieve nearly optimal efficiency with minimum memory requirement for both isotropic and anisotropic cases, they are technically challenging in implementation. Moreover, in practice, there are cases where high-accuracy computation is of great importance or even unavoidable, for example, the fine structure of vortices and the dynamics in context of the Bose-Einstein condensates [1,17], therefore, it renders the necessity of new computational methodology.

It is worthy to point out that once the kernel is approximated by $U^\varepsilon(\mathbf{x})$ at the far field with great accuracy, the Fourier transform of residual kernel $(U - U^\varepsilon)$ is quite close to its counterpart that is defined on bounded domain. It is more favorable if the **far-field** approximation U^ε is smooth. As is noticed in classical Ewald summation method [18–21], the Coulomb kernel is split into far-field and local parts as follows

$$U(r) := \frac{1}{r} = \frac{1}{r}\,\mathrm{Erf}\left(\frac{r}{\varepsilon}\right) + \frac{1}{r}\,\mathrm{Erfc}\left(\frac{r}{\varepsilon}\right) := U^\varepsilon(r) + (U - U^\varepsilon)(r), \qquad (1.2)$$

where

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t \quad \text{and} \quad \mathrm{Erfc}(x) = 1 - \mathrm{Erf}(x)$$

are the error and complementary error functions and ε is a positive parameter to be chosen later. The far-field part U^ε is smooth and accurately represents the kernel $1/r$ at far field, because $\mathrm{Erf}(r/\varepsilon) \approx 1$ with more than 16 digits accurate when $r \geq 6\varepsilon$. The remaining near-field part $(U - U^\varepsilon)$, though singular, is compactly supported.

With such far-field smooth approximation U^ε, we split the potential into a regular integral $\Phi^R(\mathbf{x})$ and a singular integral $\Phi^S(\mathbf{x})$ as follows

$$\Phi(\mathbf{x}) = \int_{\mathbb{R}^d} U^\varepsilon(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y} + \int_{\mathbb{R}^d} (U - U^\varepsilon)(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y}$$

$$:= \Phi^R(\mathbf{x}) + \Phi^S(\mathbf{x}). \qquad (1.3)$$

Both integrals are well-defined simply because integrand of $\Phi^R(\mathbf{x})$ is smooth and $(U - U^\varepsilon) \in L^1(\mathbb{R}^d)$.

The regular integral $\Phi^R(\mathbf{x})$ is well resolved by trapezoidal rule, and the resulting summation is implemented with nearly optimal efficiency thanks to FFT. We choose to integrate the singular integral potential $\Phi^S(\mathbf{x})$ by switching to Fourier space following exactly the same way of ATKM [16]. Fortunately, the required bounded domain Fourier transform of $(U - U^\varepsilon)$ can be replaced by the whole space Fourier transform, which admits analytical explicit formula, with high accuracy, thus waiving the use of technically complicated quadrature. Such methodology can be easily extended to anisotropic density case, and it achieves an anisotropic strength-independent memory requirement and computation complexity. Moreover, the numerical quadrature of each integral can be reduced to discrete

convolutions, therefore, our method is finally simplified into one discrete convolution by combining them two. The tensor generation procedure is very simple, highly efficient (involving only FFT) and independent of the anisotropy strength.

Apart from the Coulomb kernel shown above, our method is applicable to a large class of radially symmetric kernels as long as they grow no faster than exponential functions, that is,

$$\int_{R_0}^{\infty} |U(r)| \, r^{d-2} \, e^{-r^2/\delta^2} \mathrm{d}r < \infty, \quad \delta > 0,$$

where $R_0 := \min_{j=1,\cdots,d} \{2L_j\}$. The far-field smooth approximation (**FSA**) function U^ε only needs to satisfy the following two properties

$$(a) \ U^\varepsilon(\mathbf{x}) \ \text{is smooth and radially symmetric.}$$
$$(b) \ \int_{R_0}^{\infty} |(U - U^\varepsilon)(r)| r^{d-1} \mathrm{d}r \le \varepsilon_{\text{tol}}, \ \text{where} \ \varepsilon_{\text{tol}} \ll 1 \ \text{is the desired error tolerance.} \tag{1.4}$$

Such approximation can be realized with help of some special functions, including the error and smooth window functions [22], and we shall provide details later in Section 4.

The paper is organized as follows. In Section 2, we present a detailed description of the regular and singular integral evaluation, the discrete convolution structure and its extension to anisotropic case. Error estimates are given in Section 3. In Section 4, we propose a simple way to construct far-field smooth approximation. Extensive numerical results are shown in Section 5 to illustrate the performance in terms of accuracy and efficiency. Some conclusions are drawn in Section 6.

2. Numerical method

In this section, we first focus on computation of the nonlocal potential generated by isotropic density, and discuss the anisotropic case later. For simplicity, we assume that the density function ρ is compactly supported in a square domain $\mathbf{R}_L = [-L, L]^d$, which is discretized with N equally spaced grid points in each spatial direction. We fix $h = 2L/N$ as the mesh size, and the uniform mesh grid set is denoted

$$\mathcal{T} := \left\{ (x_1, x_2, \cdots, x_d) \middle| \ x_j \in \{h\ell, \ \ell = -N/2, \cdots, N/2 - 1\}, \ j = 1, \cdots, d \right\}. \tag{2.1}$$

Following the previous discussion, we shall illustrate the computation of $\Phi^R(\mathbf{x})$ and $\Phi^S(\mathbf{x})$, and present a detailed analysis in terms of accuracy and efficiency.

2.1. Evaluation of the regular integral $\Phi^R(\mathbf{x})$

Due to the compact support of the density, we have

$$\Phi^R(\mathbf{x}) = \int_{\mathbb{R}^d} U^\varepsilon(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y} \approx \int_{\mathbf{R}_L} U^\varepsilon(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y}, \quad \mathbf{x} \in \mathbf{R}_L. \tag{2.2}$$

The above integral is well approximated by applying the trapezoidal rule quadrature [7], and the resulting summation is reduced to a discrete convolution of tensor and the discrete density. The discrete convolution structure can be efficiently accelerated using FFT within $\mathcal{O}(d(2N)^d \log(2N))$ float operations [14]. For simplicity, we only present the detailed scheme for 2D case and extension to 3D case is straightforward. Define the index set

$$\mathcal{I}_N = \left\{ (n, m) \in \mathbb{Z}^2 \mid -N/2 \le n, m \le N/2 - 1 \right\}. \tag{2.3}$$

To be exact, we obtain

$$\begin{aligned}
\Phi^R(x_n, y_m) &\approx h^2 \sum_{(n', m') \in \mathcal{I}_N} U^\varepsilon\big((n - n')h, \ (m - m')h\big) \, \rho(x_{n'}, y_{m'}) \\
&:= \sum_{(n', m') \in \mathcal{I}_N} T^{(1)}_{n-n', m-m'} \, \rho_{n', m'},
\end{aligned} \tag{2.4}$$

where $\rho_{n', m'} := \rho(x_{n'}, y_{m'})$ and the tensor $T^{(1)}_{n,m}$ is given explicitly as

$$T^{(1)}_{n,m} = h^2 \, U^\varepsilon(nh, mh).$$

2.2. Evaluation of the singular integral $\Phi^S(\mathbf{x})$

To compute $\Phi^S(\mathbf{x})$ within $\mathbf{R}_L$, we first reformulate it as

$$\begin{aligned}
\Phi^S(\mathbf{x}) &:= \int_{\mathbb{R}^d} (U - U^\varepsilon)(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y} \approx \int_{\mathbf{R}_L} (U - U^\varepsilon)(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\mathrm{d}\mathbf{y} \\
&= \int_{\mathbf{x} + \mathbf{R}_L} (U - U^\varepsilon)(\mathbf{y})\rho(\mathbf{x} - \mathbf{y})\mathrm{d}\mathbf{y} = \int_{\mathbf{R}_{2L}} (U - U^\varepsilon)(\mathbf{y})\rho(\mathbf{x} - \mathbf{y})\mathrm{d}\mathbf{y}.
\end{aligned} \tag{2.5}$$

The last identity holds true because the density is compactly supported in $\mathbf{R}_L$. To be specific, for any $\mathbf{x} \in \mathbf{R}_L$, we have

$$\mathbf{y} \in \mathbf{R}_{2L} \setminus (\mathbf{x} + \mathbf{R}_L) \implies \mathbf{x} - \mathbf{y} \notin \mathbf{R}_L \implies \rho(\mathbf{x} - \mathbf{y}) = 0.$$

To integrate Eqn. (2.5), one needs to obtain a good approximation of $\rho(\mathbf{x})$ over $\mathbf{R}_{3L}$, because, for any $\mathbf{x} \in \mathbf{R}_L$ and $\mathbf{y} \in \mathbf{R}_{2L}$, we have $\mathbf{x} - \mathbf{y} \in \mathbf{R}_{3L}$. It is natural to first zero-pad the density from $\mathbf{R}_L$ to $\mathbf{R}_{3L}$ and then construct a Fourier series approximation therein. Thanks to periodicity of Fourier basis, a two-fold zero-padding to $\mathbf{R}_{2L}$ is sufficient to guarantee spectral accuracy [11,14], and it reads as follows

$$\rho_N(\mathbf{z}) := \sum_{\mathbf{k} \in \Lambda} \widetilde{\rho}_{\mathbf{k}} \, e^{i\mathbf{k} \cdot \mathbf{z}}, \quad \mathbf{z} \in \mathbf{R}_{2L}, \tag{2.6}$$

where $\Lambda = \{\mathbf{k} := \frac{\pi}{2L}(k_1, \cdots, k_d) \in \frac{\pi}{2L}\mathbb{Z}^d \,\big|\, k_j = -N, \cdots, N-1, \; j = 1, \cdots, d\}$ denotes the Fourier mesh grid. The Fourier coefficients are defined as

$$\widetilde{\rho}_{\mathbf{k}} = \frac{1}{(2N)^d} \sum_{\mathbf{z}_{\mathbf{p}} \in \mathcal{T}} \rho(\mathbf{z}_{\mathbf{p}}) e^{-i\mathbf{z}_{\mathbf{p}} \cdot \mathbf{k}}, \quad \mathbf{k} \in \Lambda. \tag{2.7}$$

As is shown earlier, the periodic extension of $\rho_N(\mathbf{z})$ is also a spectral approximation over $\mathbf{R}_{3L}$, therefore, after substituting ρ_N for ρ in (2.5), we obtain

$$\begin{aligned}
\Phi^S(\mathbf{x}) &\approx \int_{\mathbf{R}_{2L}} (U - U^\varepsilon)(\mathbf{y}) \rho_N(\mathbf{x} - \mathbf{y}) \mathrm{d}\mathbf{y}, \\
&= \sum_{\mathbf{k} \in \Lambda} \left[\int_{\mathbf{R}_{2L}} (U - U^\varepsilon)(\mathbf{y}) e^{-i\mathbf{k} \cdot \mathbf{y}} \mathrm{d}\mathbf{y} \right] \widetilde{\rho}_{\mathbf{k}} \, e^{i\mathbf{k} \cdot \mathbf{x}}, \\
&:= \sum_{\mathbf{k} \in \Lambda} W(\mathbf{k}) \, \widetilde{\rho}_{\mathbf{k}} \, e^{i\mathbf{k} \cdot \mathbf{x}}, \quad \mathbf{x} \in \mathbf{R}_L,
\end{aligned} \tag{2.8}$$

where $W(\mathbf{k})$ denotes the Fourier transform of $(U - U^\varepsilon)$ over $\mathbf{R}_{2L}$ and is given explicitly

$$W(\mathbf{k}) = \int_{\mathbf{R}_{2L}} (U - U^\varepsilon)(\mathbf{y}) e^{-i\mathbf{k} \cdot \mathbf{y}} \mathrm{d}\mathbf{y}. \tag{2.9}$$

Usually, the above Fourier transform does not admit explicit analytical expressions, therefore, one has to design appropriate numerical quadrature. For example, Greengard *et al.* [16] proposes a fast and accurate quadrature by utilizing the Gaussian-sum approximation [23]. In fact, we can approximate $W(\mathbf{k})$ by the Fourier transform of $(U - U^\varepsilon)$ due to the second property of Eqn. (1.4). That is,

$$W(\mathbf{k}) \approx \int_{\mathbb{R}^d} (U - U^\varepsilon)(\mathbf{y}) \, e^{-i\mathbf{k} \cdot \mathbf{y}} \mathrm{d}\mathbf{y}. \tag{2.10}$$

With a suitable parameter ε, we can control the approximation error as small as any prescribed precision. The above whole space integral is well-defined since $(U - U^\varepsilon) \in L^1(\mathbb{R}^d)$. Since $(U - U^\varepsilon)$ is radially symmetric, its Fourier transform (2.10) is also radially symmetric [12] and given explicitly below

$$(\widehat{U - U^\varepsilon})(\mathbf{k}) = \begin{cases} 2\pi \int_0^\infty (U - U^\varepsilon)(r) \, \mathrm{J}_0(kr) \, r \, \mathrm{d}r, & d = 2, \\ 4\pi \int_0^\infty (U - U^\varepsilon)(r) \, \frac{\sin(kr)}{kr} \, r^2 \, \mathrm{d}r, & d = 3, \end{cases} \tag{2.11}$$

where $k = |\mathbf{k}|$ and $\mathrm{J}_0(r)$ is the Bessel function of first kind with index 0. For common kernels, including Poisson, Coulomb and Biharmonic kernels, Eqn. (2.11) has analytical expressions. For a more general kernel, one may resort to numerical integration, e.g., the Gauss-Kronrod quadrature, so to obtain an accurate approximation of the Fourier transform.

Remark 1 (Parameter choice of ε for isotropic case). The parameter ε is chosen to satisfy condition (b) of Eqn. (1.4). Roughly speaking, $\varepsilon < R_0/5.85$ $(R_0/8.65)$ yields about 16 (34) digits of accuracy for 3D Coulomb potential. Numerical experience suggests that we can set $\varepsilon = 1$ to achieve 34 digits for $L = 8$ $(R_0 = 2L = 16)$. A detailed derivation is provided in Appendix B.

The quadrature (2.8) is given on uniform mesh grid and can be rewritten as a discrete convolution of a tensor and the discrete density. The tensor is actually the inverse discrete Fourier transform of vector $\{W(\mathbf{k})\}_{\mathbf{k} \in \Lambda} \in \mathbb{C}^{(2N)^d}$. To be exact, let us take the 2D case as an example. Define the Fourier modes $\mu_p = \frac{\pi p}{2L}$ and $\mu_q = \frac{\pi q}{2L}$. Plugging Eqn. (2.7) into Eqn. (2.8) and switching the summation order, the Φ^S on a uniform grid can be rewritten as

$$\begin{aligned}
\Phi^S(x_n, y_m) &\approx \sum_{(p,q) \in \mathcal{I}_{2N}} W(\mu_p, \mu_q) \, \widetilde{\rho}_{\mathbf{k}} \, e^{\frac{2\pi i}{2N}(pn + qm)} \\
&= \sum_{(n',m') \in \mathcal{I}_N} \left[\frac{1}{(2N)^2} \sum_{(p,q) \in \mathcal{I}_{2N}} W(\mu_p, \mu_q) \, e^{\frac{2\pi i}{2N}\left[p(n-n') + q(m-m') \right]} \right] \rho_{n',m'} \\
&:= \sum_{(n',m') \in \mathcal{I}_N} T^{(2)}_{n-n', m-m'} \, \rho_{n',m'},
\end{aligned} \tag{2.12}$$

where the tensor $T^{(2)}_{n,m}$ can be computed out *once for all* within $\mathcal{O}(4N^2 \log(4N^2))$ float operations using iFFT.

2.3. Discrete tensor structure

As pointed out in [12], the KTM can be rewritten as a discrete convolution of a tensor and the discrete density. In fact, for any real-valued kernel function, the potential evaluation algorithm discretized on uniform grid can be rewritten as such a discrete convolution, including the aforementioned NUFFT [8,10], ATKM [16] and GauSum method [11]. Once the tensor T has been generated, all algorithms share the same efficiency. The differences lie in the memory requirement and efficiency of the tensor generation procedure, and whether they are dependent on anisotropy strength.

In our method, both the regular integral potential Φ^R and the singular integral potential Φ^S can be written as discrete convolutions (i.e., (2.4) and (2.12)). Then it can be simplified to a single discrete convolution as follows

$$\Phi_N(x_n, y_m) := \sum_{(n',m') \in \mathcal{I}_N} T_{n-n', m-m'} \, \rho_{n',m'}, \tag{2.13}$$

where the tensor $T_{n,m} = T_{n,m}^{(1)} + T_{n,m}^{(2)}$ is given explicitly as

$$T_{n,m} = h^2 U^\varepsilon(nh, mh) + \frac{1}{(2N)^2} \sum_{(p,q) \in \mathcal{I}_{2N}} W(\mu_p, \mu_q) e^{\frac{2\pi i}{2N}(pn+qm)}. \tag{2.14}$$

The tensor generation procedure is simple and efficient, and it involves purely iFFT transform on vectors of length $(2N)^d$. Note that the tensor T is axis-symmetric, i.e., $T_{-n,m} = T_{n,m} = T_{n,-m}$ due to the symmetry of U and U^ε, therefore, we can further reduce the memory requirement to $1/4$ of its original size. Once the tensor is available, the discrete convolution (2.13) can be implemented using purely FFT/iFFT and pointwise multiplication on vectors of length $(2N)^d$. For clarity, we present detailed step-by-step algorithm using the 1D case in Algorithm 1, and extension to 2D/3D case is straightforward. We refer the reader to [12,14] for more details.

Algorithm 1 Fast implementation of one-dimensional discrete convolution.

Input: Given $\{T_j\}_{j=-N}^{N-1}$ and $\{f_j\}_{j=-N/2}^{N/2-1}$.

Output: Compute $\widehat{T} := \text{FFT}(\widetilde{T})$ with $\widetilde{T} = [T_0, T_1, \ldots, T_{N-1}, T_{-N}, \ldots, T_{-1}]$.

1: Compute $\widehat{F} := \text{FFT}(F)$ with $F = [f_{-N/2}, \ldots, f_{N/2-1}, 0, \ldots, 0] \in \mathbb{R}^{2N}$.

2: Compute $\widehat{\Phi} := \widehat{T}\widehat{F}$ by pointwise multiplication.

3: Compute $\widetilde{\Phi} = \text{iFFT}(\widehat{\Phi})$, and set $\Phi_N = \widetilde{\Phi}(1:N)$.

2.4. Anisotropic density

We assume that the density is compactly supported in an anisotropic rectangle $\mathbf{R}_L^\gamma := \prod_{j=1}^d [-L\gamma_j, L\gamma_j]$, which is also the domain of interest. We define the *anisotropy vector* by $\gamma = (\gamma_1, \cdots, \gamma_d)$. The magnitudes of the γ_j reflect the anisotropy strength along the j-th direction. Without loss of generality, we take $\gamma_1 = 1$, $\gamma_j \leq 1$ for $j = 2, \cdots, d$ and define the anisotropy strength as $\gamma_f := \prod_{j=1}^d \gamma_j^{-1}$. For simplicity of presentation, the density is sampled on a uniform mesh grid with the same number of grid points in each spatial direction (denoted by N) with mesh grid $h_j = 2L\gamma_j/N$.

We adapt it to the anisotropic density case and derive a similar discrete convolution structure. The difference is that the rectangular domain becomes anisotropic as $\mathbf{R}_{2L}^\gamma$ and the parameter R_0 is now

$$R_0 = \min_{j=1,\cdots,d} \{2L\gamma_j\}.$$

To be specific, the discrete convolution tensor for $d = 2$ reads as

$$T_{n,m} = h_1 h_2 U^\varepsilon(nh_1, mh_2) + \frac{1}{(2N)^2} \sum_{(p,q) \in \mathcal{I}_{2N}} W\left(\frac{\pi p}{2L\gamma_1}, \frac{\pi q}{2L\gamma_2}\right) e^{\frac{2\pi i}{2N}(pn+qm)}, \tag{2.15}$$

where $W(\mathbf{k})$, the Fourier transform of $(U - U^\varepsilon)$ over $\mathbf{R}_{2L}^\gamma$, is given explicitly

$$W(\mathbf{k}) = \int_{\mathbf{R}_{2L}^\gamma} (U - U^\varepsilon)(\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{y}} d\mathbf{y}. \tag{2.16}$$

Similar to the isotropic case, $W(\mathbf{k})$ can be approximated by the Fourier transform of $(U - U^\varepsilon)$ due to the second property of Eqn. (1.4) with a suitable ε.

Remark 2 (Parameter choice of ε for anisotropic case). The choice of parameter ε follows exactly the same principles as shown in Remark 1. Roughly speaking, $\varepsilon < 2L\min\{\gamma_j\}/5.85$ ($2L\min\{\gamma_j\}/8.65$) yields about 16 (34) digits of accuracy for 3D Coulomb potential. The parameter ε decreases as the anisotropy strength increases. Numerical experience suggests that we can set $\varepsilon = 0.4$ to achieve 34 digits of accuracy for 3D Coulomb potential with anisotropy strength $\gamma = (1, 1, \gamma_3)$ for $\gamma_3 = 1, 1/2, 1/4, 1/8$ and $L = 12$, where the value of ε is determined by the smallest $\gamma_3 = 1/8$.
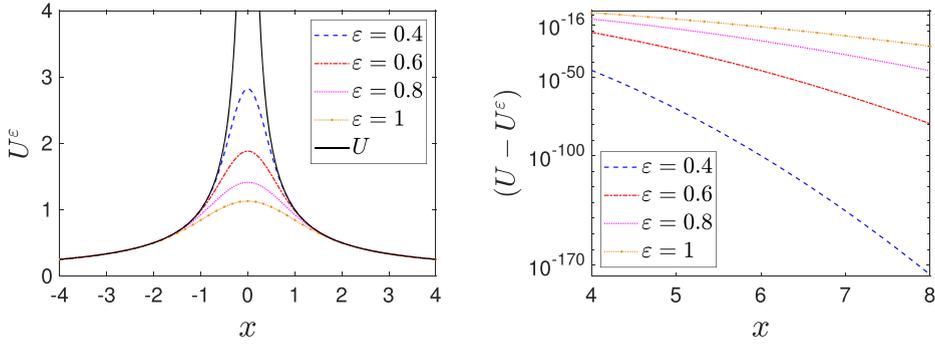
Fig. 2. Slice plot of U^ε (left) and the error $(U - U^\varepsilon)$ (right) for Coulomb kernel $U = 1/r$.

For the anisotropic case, our method remains effective and can be reformulated as a single discrete convolution. The tensor generation follows the same procedure as in the isotropic case and involves only FFT/iFFT transforms on vectors of length $(2N)^d$. Once the convolution tensor is computed in the pre-computation step, the subsequent operations involves FFT/iFFT transforms and pointwise multiplications on vectors of length $(2N)^d$. Consequently, we conclude that our method effectively avoids the additional cost associated with the anisotropy strength.

Remark 3 (Comparison of the tensor generation procedure). The memory requirement and computational costs for our method (denoted by FSA hereafter), KTM, GauSum and ATKM are $\mathcal{O}(SN^d)$ and $\mathcal{O}(SN^d \log(SN^d))$ respectively, with the only differences lying in the zero-padding factor S. For FSA, GauSum and ATKM, $S = 2^d$ for both isotropic and anisotropic cases. While, in KTM, we take $S = 3^d$ for isotropic case and

$$S = \prod_{j=1}^{d} S_j, \quad S_j := \lceil 1 + \gamma_j^{-1}(1 + \prod_{k=2}^{d} \gamma_k^2)^{1/2} \rceil \tag{2.17}$$

for anisotropic case [14], where $\lceil \cdot \rceil$ is the rounding up function. For example, in the tensor generation procedure, a double-precision computation on a 256^3 grid requires a minimum 1 GB memory for our method while the KTM requires around 3.4 GB, with the reduction factor being around **71**%. In addition, the implementation of our method is as simple as that of KTM.

Remark 4 (Parallelizability). Our method is based on FFTs and pointwise multiplications. Therefore, it can be parallelized in a straightforward manner, and its parallelizability is essentially the same as that of the FFT. Parallel implementations, such as multi-threaded FFTs (OpenMP-enabled FFTW/MKL), distributed-memory FFT libraries (MPI-based), GPU-accelerated FFTs, can further enhance the efficiency of our method.

3. Error estimates

In this subsection, we choose to present the error estimates, which mainly consist of errors coming from trapezoidal rule discretization, the density approximation, and Fourier transform of the kernel on rectangular domain. For more details, we refer the readers to Appendix A.

Theorem 1. *For smooth density $\rho(\mathbf{x})$ that is compactly supported in $\mathbf{R}_L^\gamma$, the following estimates*

$$\|\Phi - \Phi_N\|_{L^\infty(\mathbf{R}_L^\gamma)} \lesssim N^{-m} + \varepsilon_{\text{tol}},$$

hold true for any positive integer $m > 0$.

4. Far-field smooth approximation

Notice that the smooth approximation is done far away from the origin, i.e., $[R_0, \infty)$, and $\text{Erf}(r/\varepsilon) \approx 1$ with more than 16 (34) digits of accuracy for $r \geq 6\varepsilon$ (8.7ε), as shown in Fig. 2 for the Coulomb kernel, we can define the approximation as $U^\varepsilon(r) := U(r)\,\text{Erf}(r/\varepsilon)$. Such approximation satisfies the requirement (b) of Eqn. (1.4) and a rigorous derivation is provided in Theorem 2 below. It is easy to implement in practice.

Theorem 2 (Far-field approximation). *For d-dimensional radially symmetric kernel $U(r)$ that satisfies $\int_{R_0}^{\infty} |U(r)|\, r^{d-2}\, e^{-r^2/\delta^2}\, \mathrm{d}r < \infty$ ($\delta > 0$) and its approximation*

$$U^\varepsilon(r) = U(r)\,\text{Erf}\left(\frac{r}{\varepsilon}\right), \tag{4.1}$$

where error function $\text{Erf}(r) = \frac{2}{\sqrt{\pi}} \int_0^r e^{-t^2}\, \mathrm{d}t$, there exists a positive constant $\varepsilon_0 > 0$ such that following estimate

$$\int_{R_0}^{\infty} |(U - U^\varepsilon)(r)|\, r^{d-1}\, \mathrm{d}r \leq \varepsilon_{\text{tol}}, \quad \forall\, 0 < \varepsilon < \varepsilon_0 \tag{4.2}$$

Table 1

Far-field smooth approximation and the corresponding Fourier transform for common kernels.

	$U(r)$	$U^\varepsilon(r)$	$(\widehat{U - U^\varepsilon})(k)$
Poisson	$2D$: $\frac{-1}{2\pi}\ln(r)$	$\frac{-1}{2\pi}\left[\ln(r) + \frac{1}{2}\,\mathrm{E}_1(\frac{r^2}{\varepsilon^2})\right]$	$\frac{1}{k^2}\left[1 - e^{-\frac{1}{4}k^2\varepsilon^2}\right]$
Coulomb	$2D$: $\frac{1}{2\pi r}$	$\frac{1}{2\pi r}\,\mathrm{Erf}(\frac{r}{\varepsilon})$	$\frac{1}{k}\,\mathrm{Erf}(\frac{k\varepsilon}{2})$
	$3D$: $\frac{1}{4\pi r}$	$\frac{1}{4\pi r}\,\mathrm{Erf}(\frac{r}{\varepsilon})$	$\frac{1}{k^2}\left[1 - e^{-\frac{1}{4}k^2\varepsilon^2}\right]$
Biharmonic	$2D$: $\frac{-r^2}{8\pi}\left[\ln(r) - 1\right]$	$\frac{-r^2}{8\pi}\left[\ln(r) + \frac{1}{2}\,\mathrm{E}_1(\frac{r^2}{\varepsilon^2}) - 1\right]$	$\frac{-16 + e^{-\frac{1}{4}k^2\varepsilon^2}(16+4k^2\varepsilon^2+k^4\varepsilon^4)}{16k^4}$
	$3D$: $\frac{r}{8\pi}$	$\frac{r}{8\pi}\,\mathrm{Erf}(\frac{r}{\varepsilon})$	$\frac{-8 + e^{-\frac{1}{4}k^2\varepsilon^2}(8+2k^2\varepsilon^2+k^4\varepsilon^4)}{8k^4}$

holds true.

Proof. Using asymptotics of complementary error function $\mathrm{Erfc}(x) = 1 - \mathrm{Erf}(x)$, we obtain

$$\mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\mathrm{d}t = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2+x^2-x^2}\mathrm{d}t = \frac{2}{\sqrt{\pi}}e^{-x^2}\int_x^\infty e^{-(t-x)(t+x)}\mathrm{d}t$$

$$\leq \frac{2}{\sqrt{\pi}}e^{-x^2}\int_x^\infty e^{-2x(t-x)}\mathrm{d}t = \frac{1}{\sqrt{\pi}}\frac{1}{x}e^{-x^2}.$$

Then, we have

$$\int_{R_0}^\infty |(U - U^\varepsilon)(r)|r^{d-1}\mathrm{d}r = \int_{R_0}^\infty |U(r)|r^{d-1}\,\mathrm{Erfc}\,(r/\varepsilon)\mathrm{d}r \leq \frac{\varepsilon}{\sqrt{\pi}}\int_{R_0}^\infty |U(r)|r^{d-2}e^{-\frac{r^2}{\varepsilon^2}}\mathrm{d}r.$$

Function $f(\varepsilon) := \int_{R_0}^\infty |U(r)|\,r^{d-2}\,e^{-r^2/\varepsilon^2}\,\mathrm{d}r$ is monotone increasing and bounded. Using dominated convergence theorem, we obtain

$$\lim_{\varepsilon \to 0} f(\varepsilon) = \lim_{k\to\infty} f(\varepsilon_k) := \lim_{k\to\infty} f(\delta/k) = \int_{R_0}^\infty |U(r)|\,r^{d-2}\lim_{k\to\infty}e^{-\frac{r^2 k^2}{\delta^2}}\,\mathrm{d}r = 0,$$

Therefore, there exists $\varepsilon_0 > 0$ such that estimate (4.2) holds true for any $\varepsilon < \varepsilon_0$. $\quad\square$

As for the **smoothness** requirement, i.e., condition (a) of Eqn. (1.4), the approximation (4.1) is not necessarily smooth for all kernels. For non-smooth cases, we can first set $U^\varepsilon(r)$ to be 0 for $r \approx 0$ and $U(r)\,\mathrm{Erf}(r/\varepsilon)$ for $r \geq R_0$, then connect the two regions smoothly via a smooth vanishing window function [22]. To be specific, we set

$$U^\varepsilon(r) = U(r)\,\mathrm{Erf}\left(\frac{r}{\varepsilon}\right)\left[1 - \xi(r)\right], \tag{4.3}$$

where $\xi(r) = 0$ if formula (4.1) is smooth, otherwise, set $\xi(r)$ to be a smooth vanishing window function. Such construction mechanism works for almost all kernels, but it may not be the simplest one for certain kernels. For example, for the 2D Poisson kernel $U(r) = -\frac{1}{2\pi}\ln(r)$, we choose the following approximation function

$$U^\varepsilon(r) = -\frac{1}{2\pi}\left[\ln(r) + \frac{1}{2}\,\mathrm{E}_1\left(\frac{r^2}{\varepsilon^2}\right)\right], \tag{4.4}$$

where $\mathrm{E}_1(r) := \int_r^\infty t^{-1}e^{-t}\mathrm{d}t$ is the exponential integral function [24].

With the above-mentioned smooth approximations, the Fourier transform of $(U - U^\varepsilon)$ is reduced to one-dimensional integral (2.11). Here we summarize the corresponding analytical expressions for different common kernels in Table 1, from which we can observe that the corresponding Fourier transform is smooth and non-oscillatory, as against the annoying oscillatory Fourier transform in KTM.

The parameter ε is chosen to satisfy condition (b) of Eqn. (1.4), and it depends on kernel U, far-field approximation U^ε, R_0 and $\varepsilon_{\mathrm{tol}}$. A detailed derivation is provided in Appendix B, using which we can set $\varepsilon = 1$ (0.4) for the isotropic (anisotropic) density case for the 2D/3D Coulomb and 2D Poisson potentials when $\varepsilon_{\mathrm{tol}} = 10^{-16}$ and $L = 12$.

Remark 5. In computational practice, we shall choose a relatively large ε for better numerical accuracy. Even though the integrand of regular integral (2.2) is smooth for any fixed ε on the continuous level, for a smaller ε, it requires a finer mesh so to capture the sharp variations near the singularity on the discrete level. The quantitative relation can be obtained from the error estimate of the trapezoidal rule, which varies with the kernel. Roughly speaking, the relation is approximately $h < \varepsilon$. This is because the m-th derivative of U^ε typically scales like ε^{-m}, and the trapezoidal rule error behaves as $Ch^m\varepsilon^{-m}$ with $m \in \mathbb{Z}^+$.

5. Numerical results

In this section, we shall investigate the accuracy (in both double and quadruple precision) and efficiency for different nonlocal potentials evaluation with isotropic and anisotropic densities. The computational domain $\mathbf{R}_L^r$ is discretized uniformly in each spatial

direction with mesh size h_j, and we define mesh size vector as $\mathbf{h} = (h_1, \cdots, h_d)$. For simplicity, we shall use h to denote the mesh size if all the mesh sizes are equal.

The numerical error is measured in following norm

$$\mathcal{E} := \frac{\|\Phi - \Phi_{\mathbf{h}}\|_{l^\infty}}{\|\Phi\|_{l^\infty}} = \frac{\max_{\mathbf{x} \in \mathcal{T}_{\mathbf{h}}} |\Phi(\mathbf{x}) - \Phi_{\mathbf{h}}(\mathbf{x})|}{\max_{\mathbf{x} \in \mathcal{T}_{\mathbf{h}}} |\Phi(\mathbf{x})|},$$

where $\Phi_{\mathbf{h}}$ is the numerical solution on mesh grid $\mathcal{T}_{\mathbf{h}}$ and $\Phi(\mathbf{x})$ is the exact solution. In the following examples, the potential $\Phi(\mathbf{x})$ can be computed analytically.

5.1. The Coulomb potentials in 2D/3D

Example 1. Here, we consider the 2D/3D Coulomb potentials with the following two types of source density

- **Case I**: Isotropic/anisotropic density

$$\rho(\mathbf{x}) = \begin{cases} e^{-(x^2+y^2/\gamma_2^2)/\sigma^2}, & d = 2, \\ e^{-(x^2+y^2+z^2/\gamma_3^2)/\sigma^2}, & d = 3. \end{cases}$$

The corresponding 3D Coulomb potential can be computed analytically as [11]

$$\Phi(\mathbf{x}) = \begin{cases} \frac{\sigma^3 \sqrt{\pi}}{4|\mathbf{x}|} \operatorname{Erf}(\frac{|\mathbf{x}|}{\sigma}), & \gamma_3 = 1, \\ \frac{\gamma_3 \sigma^2}{4} \int_0^\infty \frac{e^{-\frac{x^2+y^2}{\sigma^2(t+1)}} e^{-\frac{z^2}{\sigma^2(t+\gamma_3^2)}}}{(t+1)\sqrt{t+\gamma_3^2}} \, dt, & \gamma_3 \neq 1. \end{cases}$$

and the 2D Coulomb potential is given analytically as [11]

$$\Phi(\mathbf{x}) = \begin{cases} \frac{\sqrt{\pi}\sigma}{2} \operatorname{I}_0(\frac{|\mathbf{x}|^2}{2\sigma^2}) e^{-\frac{|\mathbf{x}|^2}{2\sigma^2}}, & \gamma_2 = 1, \\ \frac{\gamma_2 \sigma}{\sqrt{\pi}} \int_0^\infty \frac{e^{-\frac{x^2}{\sigma^2(t^2+1)}} e^{-\frac{y^2}{\sigma^2(t^2+\gamma_2^2)}}}{\sqrt{t^2+1}\sqrt{t^2+\gamma_2^2}} \, dt, & \gamma_2 \neq 1, \end{cases}$$

where $\operatorname{I}_0(x)$ is the modified Bessel function of the first kind [24].

- **Case II**: Shifted density $\rho(\mathbf{x}) = \rho_0(\mathbf{x}) + \rho_0(\mathbf{x} - \mathbf{x}_0)$, where $\rho_0(\mathbf{x})$ is generated by taking the Laplacian of the potential $\Phi_0(\mathbf{x}) = e^{-(x^2+y^2+z^2/\gamma_3^2)/\sigma^2}$, i.e.,

$$\rho_0(\mathbf{x}) = -\Delta\Phi_0(\mathbf{x}) = \Phi_0(\mathbf{x})\left(-\frac{4x^2}{\sigma^4} - \frac{4y^2}{\sigma^4} - \frac{4z^2}{\gamma_3^4\sigma^4} + \frac{4}{\sigma^2} + \frac{2}{\gamma_3^2\sigma^2}\right).$$

The corresponding 3D Coulomb potential analytically reads as

$$\Phi(\mathbf{x}) = \Phi_0(\mathbf{x}) + \Phi_0(\mathbf{x} - \mathbf{x}_0).$$

Table 2 presents errors of the 2D/3D Coulomb potentials in double and quadruple precision for isotropic and anisotropic densities. For isotropic case, i.e., $\gamma_2 = 1$ or $\gamma_3 = 1$, we consider Case I with $L = 8$, $\sigma = \sqrt{0.8}$ and $\varepsilon = 1$. For anisotropic case, we consider Case I with $L = 8$, $\sigma = \sqrt{1.2}$, $\varepsilon = 0.5$, $\mathbf{h} = \gamma/4$ and Case II with $L = 12$, $\sigma = \sqrt{0.8}$, $\varepsilon = 0.4$, $\mathbf{x}_0 = (1, 1, 0)$ and $\mathbf{h} = \gamma/8$, where $\gamma = (1, \gamma)$ for $d = 2$ and $\gamma = (1, 1, \gamma)$ for $d = 3$.

From Table 2, we can conclude that our method achieves spectral accuracy for both isotropic and anisotropic cases. Note that it can reach 10^{-34} for quadruple-precision version when the tensor T in (2.13) is generated with high precision. For ATKM, one needs to develop a high-precision black-box version. This is possible but technically challenging and involves more Gaussian terms to reach a smaller tolerance.

Example 2. Comparison with KTM and ATKM. To compare the performance of FSA, KTM and ATKM, we choose densities in Case II of Example 1 with $\sigma = 0.5$. The computation is split into two parts: the pre-computation part (**Precomp**): computation of the tensor $\widehat{T}_{\mathbf{k}}$ for KTM and FSA or $\widehat{U}_R$ for ATKM [16], and the execution part (**Execution**). The algorithms were implemented in FORTRAN and run on a single core of a 3.00GHz Intel(R) Xeon(R) Gold 6248R CPU with a 36 MB cache in Ubuntu GNU/Linux with the Intel compiler ifort.

Table 2
Errors of the 2D/3D Coulomb potentials for isotropic and anisotropic densities in Example 1.

Isotropic density					
		$h = 1$	$h = 1/2$	$h = 1/4$	$h = 1/8$
2D	double	1.3856E-02	2.9648E-08	2.8012E-16	5.6025E-16
3D	double	2.0681E-02	2.5036E-06	5.5511E-16	6.9389E-16
	quad	2.0681E-02	2.5036E-06	4.8161E-18	2.4195E-34
Anisotropic density					
		$\gamma = 1$	$\gamma = 1/2$	$\gamma = 1/4$	$\gamma = 1/8$
2D	Case I	4.1758E-16	2.5550E-15	1.5455E-15	1.8119E-15
3D	Case I	3.7007E-16	5.3559E-15	5.1651E-15	3.9372E-15
	Case II	6.0077E-16	6.0289E-16	8.0178E-16	1.2020E-15
	Case II (quad)	6.9529E-34	6.9676E-34	1.5629E-33	2.7787E-33

Table 3
The performance of KTM and FSA for isotropic density.

$N^d = 192^3$	$\mathcal{E}$	$T_{\text{Precomp}}(s)$	$T_{\text{Execution}}(s)$
KTM	3.3502E-16	6.41	2.24
FSA	3.3394E-16	2.84	2.24

Table 4
The accuracy comparison of KTM, ATKM and FSA at a larger error level.

$N^d = 96^3$	$\gamma = 1$	$\gamma = 1/2$	$\gamma = 1/4$	$\gamma = 1/8$
KTM	5.51505E-06	5.40040E-06	4.41505E-06	3.96247E-06
ATKM	5.45321E-06	5.21110E-06	4.54769E-06	3.72487E-06
FSA	5.45318E-06	5.21106E-06	4.54756E-06	3.72418E-06

Table 3 presents the errors $\mathcal{E}$ and CPU time for KTM and FSA in isotropic density case. Table 4 presents a larger error level for KTM, ATKM and FSA methods with different values of γ. Fig. 3 shows the CPU time of pre-computation part versus different anisotropy strengths γ_f for KTM, ATKN and FSA. We do not present the execution time because the subsequent computations are the same, once the pre-computation is completed.

From Table 3 and Fig. 3, we can see that: (i) The pre-computation of FSA is the fastest for both isotropic and anisotropic cases. This is because both FSA and ATKM require FFT on vectors of length $(2^d)N^d$, whereas KTM requires FFT on vectors of length SN^d with the zero-padding factor $S > 2^d$ (taken as (2.17)) growing linearly with the anisotropy strength γ_f [14]. The complicated evaluation of kernel's Fourier transform over the rectangle domain in ATKM results in a slightly degraded efficiency. (ii) The CPU time of FSA and ATKM are independent on the anisotropy strength γ_f. In contrast, the time for KTM exhibit a linear growth with respect to γ_f. Consequently, we can conclude that FSA performs better in terms of efficiency.

5.2. The Poisson potential in 2D

Example 3. We consider the 2D Poisson potential with the following two types of source density

- **Isotropic density**: we take Gaussian $\rho(\mathbf{x}) = e^{-|\mathbf{x}|^2/\sigma^2}$. The corresponding 2D Poisson potential could be analytically integrated as [16]

$$\Phi(\mathbf{x}) = -\frac{\sigma^2}{4}\left[E_1(|\mathbf{x}|^2/\sigma^2) + 2\ln(|\mathbf{x}|)\right],$$

- **Anisotropic density**: we take

$$\rho(\mathbf{x}) = -\Delta\Phi(\mathbf{x}) = \Phi(\mathbf{x})\left(-\frac{4x^2}{\sigma^4} - \frac{4y^2}{\gamma_2^4\sigma^4} + \frac{2}{\sigma^2} + \frac{2}{\gamma_2^2\sigma^2}\right),$$

The corresponding 2D Poisson potential reads as $\Phi(\mathbf{x}) = e^{-(x^2 + y^2/\gamma_2^2)/\sigma^2}$.

Table 5 presents errors of the 2D Poisson potentials for isotropic and anisotropic densities. For isotropic case, the parameters are chosen as $L = 8$, $\sigma = \sqrt{1.2}$ and $\varepsilon = 1$. For anisotropic case, we choose $L = 10$, $\sigma = 1.2$, $\varepsilon = 0.4$ and $\mathbf{h} = (1, \gamma_2)/8$.
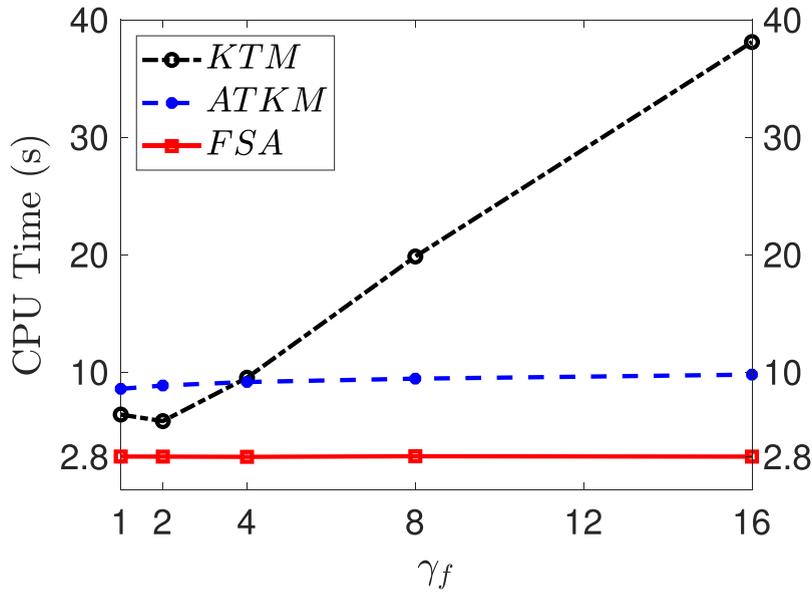
Fig. 3. Timing results of the pre-computation part versus increasing anisotropy strength γ_f with $N^d = 192^3$.

Table 5
Errors of the 2D Poisson potentials for isotropic and anisotropic densities in Example 3.

	$h = 2$	$h = 1$	$h = 1/2$	$h = 1/4$
Isotropy	2.1786E-01	1.3761E-03	5.5617E-09	4.9577E-16
	$\gamma_2 = 1$	$\gamma_2 = 1/2$	$\gamma_2 = 1/4$	$\gamma_2 = 1/8$
Anisotropy	4.5519E-16	2.2204E-16	6.2728E-16	1.5016E-15

5.3. The 3D/quasi-2D dipole-dipole interaction (DDI)

Example 4. We consider the following two cases

The 3D DDI. The kernel is given as

$$U(\mathbf{x}) = \frac{3}{4\pi} \frac{\mathbf{m} \cdot \mathbf{n} - 3(\mathbf{x} \cdot \mathbf{m})(\mathbf{x} \cdot \mathbf{n})/|\mathbf{x}|^2}{|\mathbf{x}|^3}, \quad \mathbf{x} \in \mathbb{R}^3,$$

where $\mathbf{n}, \mathbf{m} \in \mathbb{R}^3$ are unit vectors representing the dipole orientations, and the 3D potential is reformulated as [1,10]

$$\Phi(\mathbf{x}) = -(\mathbf{m} \cdot \mathbf{n})\rho(\mathbf{x}) - 3\frac{1}{4\pi|\mathbf{x}|} * (\partial_{\mathbf{nm}}\rho),$$

where $\partial_{\mathbf{m}} = \mathbf{m} \cdot \nabla$ and $\partial_{\mathbf{nm}} = \partial_{\mathbf{n}}(\partial_{\mathbf{m}})$. In fact, the potential can be calculated via the 3D Coulomb potential with source term $\partial_{\mathbf{nm}}\rho$, which can be easily computed numerically via Fourier spectral method [7]. We consider 3D DDI with the following two types of source density

- **Isotropic density**: we take Gaussian $\rho(\mathbf{x}) = e^{-|\mathbf{x}|^2/\sigma^2}$. The corresponding 3D DDI can be given explicitly as

$$\Phi(\mathbf{x}) = -(\mathbf{m} \cdot \mathbf{n})\rho(\mathbf{x}) - 3\partial_{\mathbf{mn}}\left(\frac{\sigma^2\sqrt{\pi}}{4} \frac{\mathrm{Erf}(|\mathbf{x}|/\sigma)}{|\mathbf{x}|/\sigma} \right).$$

- **Anisotropic density**: we take

$$\rho(\mathbf{x}) = \left(-\frac{4x^2}{\gamma^4\sigma^4} - \frac{4y^2}{\sigma^4} - \frac{4z^2}{\sigma^4} + \frac{4}{\sigma^2} + \frac{2}{\gamma^2\sigma^2} \right)e^{-\frac{x^2/\gamma^2+y^2+z^2}{\sigma^2}}.$$

The corresponding 3D DDI can be given explicitly as

$$\Phi(\mathbf{x}) = -(\mathbf{m} \cdot \mathbf{n})\rho(\mathbf{x}) - 3\partial_{\mathbf{mn}}\left(e^{-\frac{x^2/\gamma^2+y^2+z^2}{\sigma^2}} \right).$$

Table 6

Errors of quasi-2D/3D DDI with isotropic and anisotropic densities in Example 4.

Isotropic density				
	$h = 2$	$h = 1$	$h = 1/2$	$h = 1/4$
quasi-2D	2.0847E-01	7.4038E-03	2.2647E-07	5.0826E-15
3D	2.2087	3.3668E-02	8.5098E-07	7.5667E-15
Anisotropic density ($\gamma = 1/4$)				
	$h = 1$	$h = 1/2$	$h = 1/4$	$h = 1/8$
3D	5.4333E-02	7.3190E-04	9.8878E-11	1.8948E-16

Table 7

Errors of 3D DDI with $\mathbf{m} = \mathbf{n} = [0, 0, 1]^\top$ under different anisotropy strengths γ in Example 4.

$\gamma = 1/16$	$\gamma = 1/8$	$\gamma = 1/4$	$\gamma = 1/2$
1.7833E-16	1.8046E-16	1.8948E-16	4.5703E-16

The quasi-2D DDI. The kernel is given as [1,25]

$$U(\mathbf{x}) = -\frac{3}{2}(\partial_{\mathbf{n}_\perp \mathbf{n}_\perp} - n_3^2 \nabla_\perp^2)\frac{1}{(2\pi)^{3/2}} \int_\mathbb{R} \frac{e^{-s^2/2}}{\sqrt{|\mathbf{x}|^2 + \eta^2 s^2}} \mathrm{d}s$$

$$:= -\frac{3}{2}(\partial_{\mathbf{n}_\perp \mathbf{n}_\perp} - n_3^2 \nabla_\perp^2)\widetilde{U}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2,$$

where $\eta > 0$, $\nabla_\perp^2 = \Delta$, $\mathbf{n}_\perp = (n_1, n_2)^\top$, $\partial_{\mathbf{n}_\perp} = \mathbf{n}_\perp \cdot \nabla_\perp$ and $\partial_{\mathbf{n}_\perp \mathbf{n}_\perp} = \partial_{\mathbf{n}_\perp}(\partial_{\mathbf{n}_\perp})$. The quasi-2D DDI can be reformulated as a convolution of $\widetilde{U}$ and an effective density $\widetilde{\rho}$ as follows

$$\Phi(\mathbf{x}) = \widetilde{U} * \left[-\frac{3}{2}(\partial_{\mathbf{n}_\perp \mathbf{n}_\perp} - n_3^2 \nabla^2) \right]\rho = \widetilde{U} * \widetilde{\rho}.$$

For a Gaussian density $\rho(\mathbf{x}) = e^{-|\mathbf{x}|^2/\sigma^2}$ and dipole orientation $\mathbf{n} = (0, 0, 1)^\top$, the quasi-2D DDI are given explicitly as

$$\Phi(\mathbf{x}) = -\frac{8}{\sigma^4}e^{-\frac{r^2}{\sigma^2}}\pi \int_0^\infty \widetilde{U}(t)t e^{-\frac{t^2}{\sigma^2}}\left[(-\sigma^2 + r^2 + t^2)\,\mathrm{I}_0\left(\frac{2rt}{\sigma^2}\right) - 2rt\,\mathrm{I}_1\left(\frac{2rt}{\sigma^2}\right)\right]\mathrm{d}t,$$

where I_0 and I_1 are the modified Bessel functions of order 0 and 1 respectively.

In our computation, unless otherwise specified, the parameters are chosen $L = 12$, $\sigma = \sqrt{1.2}$ and $\varepsilon = 1$. Table 6 shows errors of the quasi-2D DDI, computed with $\eta = 1/\sqrt{32}$ and $\mathbf{n} = (0, 0, 1)^\top$ and the 3D DDI with different mesh sizes $\mathbf{h} := (\gamma h, h, h)$ for isotropic and anisotropic densities. For isotropic case, we choose $L = 8$ and dipole orientations $\mathbf{n} = (0.82778, 0.41505, -0.37751)^\top$, $\mathbf{m} = (0.3118, 0.9378, -0.15214)^\top$, while for anisotropic case, we choose $\sigma = \sqrt{0.8}$, $\varepsilon = 0.4$ and dipole orientations $\mathbf{m} = \mathbf{n} = [0, 0, 1]^\top$. Table 7 presents errors of 3D DDI with mesh size $h = 1/8$, under different anisotropy strengths γ.

From Table 6 and Table 7, it can be concluded that our method achieves spectral accuracy for both isotropic and anisotropic cases.

5.4. The Biharmonic potentials in 2D/3D

Example 5. Here, we consider the 2D/3D Biharmonic potentials with kernel given by

$$U(\mathbf{x}) = \begin{cases} -\frac{1}{8\pi}|\mathbf{x}|^2(\ln(|\mathbf{x}|) - 1), & d = 2, \\ \frac{1}{8\pi}|\mathbf{x}|, & d = 3. \end{cases}$$

We take Gaussian $\rho(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^d}e^{-|\mathbf{x}|^2/2/\sigma^2}$ and the corresponding Biharmonic potentials can be analytically integrated as

$$\Phi(\mathbf{x}) = \begin{cases} \frac{1}{8\pi}\left(r^2 + \sigma^2 e^{-\frac{r^2}{2\sigma^2}}\right) + \frac{1}{16\pi}(r^2 + 2\sigma^2)\left[\mathrm{Ei}\left(-\frac{r^2}{2\sigma^2}\right) - 2\ln(r)\right], & d = 2, \\ \frac{1}{8\pi}\left[\mathrm{Erf}\left(\frac{r}{\sqrt{2}\sigma}\right)\left(\frac{\sigma^2}{r} + r\right) + \sigma\sqrt{\frac{2}{\pi}}\,e^{-\frac{r^2}{2\sigma^2}}\right], & d = 3, \end{cases}$$

where $\mathrm{Ei}(x) = \int_{-\infty}^x e^s/s\,\mathrm{d}s$ is the exponential integral.

Table 8 presents errors of the 2D/3D Biharmonic potentials computed with $L = 12$, $\sigma = \sqrt{1.2}$ and $\varepsilon = 1$. From these results, we can see that our method achieves spectral accuracy.

Table 8
Errors of the 2D/3D Biharmonic potentials in Example 5.

	$h = 2$	$h = 1$	$h = 1/2$	$h = 1/4$
2D	2.1351E-01	2.6558E-05	5.8860E-12	1.2938E-15
3D	3.4293E-01	2.6307E-04	1.1065E-10	1.0623E-15

Table 9
Errors of the 2D/3D Yukawa potentials in Example 6.

	h	2	1	1/2	1/4
	$\lambda = 2$	1.7460E-01	4.5096E-03	4.3501E-08	5.2274E-16
2D	$\lambda = 3$	1.1428E-01	4.4972E-03	6.4647E-08	1.1345E-15
	$\lambda = 4$	7.7673E-02	3.9413E-03	8.0102E-08	2.6201E-15
	$\lambda = 2$	2.4997E-01	6.8294E-03	7.3633E-08	1.1680E-15
3D	$\lambda = 3$	1.6538E-01	6.6018E-03	1.0223E-07	1.8163E-15
	$\lambda = 4$	2.6416E-02	5.7507E-03	1.2274E-07	1.4125E-15

Table 10
Errors of the Davey-Stewartson nonlocal potential in Example 7.

h	1/2	1/4	1/8	1/16
$\mathcal{E}$	2.2057E-02	6.5495E-04	3.9396E-09	3.1099E-15

5.5. The Yukawa potentials in 2D/3D

Example 6. Here, we consider the 2D/3D Yukawa potentials with convolution kernel

$$U(\mathbf{x}) = \begin{cases} \frac{1}{2\pi} \, \mathrm{K}_0(\lambda|\mathbf{x}|), & d = 2, \\ \frac{1}{4\pi|\mathbf{x}|} e^{-\lambda|\mathbf{x}|}, & d = 3. \end{cases}$$

To this end, we consider the Gaussian source $\rho(\mathbf{x}) = e^{-|\mathbf{x}|^2/\sigma^2}$, which generates the exact potential

$$\Phi(\mathbf{x}) = \begin{cases} \frac{(\sqrt{\pi}\sigma)^3}{2} \, \frac{e^{-\lambda r + \frac{\lambda^2\sigma^2}{4}}}{4\pi r} \left[\mathrm{Erfc}\left(-\frac{r}{\sigma} + \frac{\lambda\sigma}{2} \right) - e^{2\lambda r} \, \mathrm{Erfc}\left(\frac{r}{\sigma} + \frac{\lambda\sigma}{2} \right) \right], & d = 2, \\ \int_0^\infty \mathrm{K}_0(\lambda s) \, s \, e^{-\frac{r^2+s^2}{\sigma^2}} \, \mathrm{I}_0\left(\frac{2rs}{\sigma^2} \right) \, ds, & d = 3. \end{cases}$$

Table 9 presents errors of the 2D/3D Yukawa potentials for different λ, computed with $L = 12$, $\sigma = \sqrt{1.2}$ and $\varepsilon = 1$. From these results, we can see that our method achieves spectral accuracy.

5.6. The Davey-Stewartson (DS) nonlocal potential

Example 7. In the DS equation, the nonlocal potential can be given by a convolution as follows

$$\Phi(\mathbf{x}) = -\frac{1}{2\pi} \ln(|\mathbf{x}|) * (\partial_{xx}\rho).$$

For a Gaussian density $\rho(\mathbf{x}) = \pi e^{-\pi^2|\mathbf{x}|}$, the DS nonlocal potential is given explicitly, in polar coordinates, as [11]

$$\Phi(r, \theta) = -\left[\frac{\pi}{2} + \frac{\cos(2\theta)}{2\pi r^2} \left(1 + \pi^2 r^2 - e^{\pi^2 r^2} \right) \right] e^{-\pi^2 r^2}.$$

Table 10 shows errors of the Davey-Stewartson nonlocal potential computed with $L = 8$ and $\varepsilon = 1$. From these results, we can see that our method achieves spectral accuracy.

5.7. The Helmholtz potential

Example 8. Here, we consider the 3D Helmholtz potentials with convolution kernel

$$U(\mathbf{x}) = \frac{e^{i\lambda r}}{4\pi r} \quad \text{with} \quad r = |\mathbf{x}|,$$

Table 11

Errors of the Helmholtz nonlocal potential in Example 8.

h	1	1/2	1/4	1/8
$\lambda = 1$	3.6014E-02	4.0497E-06	1.0253E-13	6.3870E-16
$\lambda = 5$	1.3074E-01	1.1282E-04	8.4923E-12	1.4303E-15
$\lambda = 10$	1.4465	2.6800E-01	1.5045E-09	2.3156E-14

with the Gaussian density $\rho(\mathbf{x}) = \frac{1}{\sigma^3 (2\pi)^{3/2}} e^{-\frac{|\mathbf{x}|^2}{2\sigma^2}}$, and the exact potential reads as [12]

$$\Phi(\mathbf{x}) = \frac{1}{4\pi r} e^{-\frac{\sigma^2 \lambda^2}{2}} \operatorname{Re}\left[e^{-i\lambda r} \operatorname{Erf}\left(\frac{r - \sigma^2 i\lambda}{\sqrt{2}\sigma} \right) \right].$$

Table 11 shows errors of the Helmholtz nonlocal potential computed with $L = 8$, $\sigma = \sqrt{0.4}$ and $\varepsilon = 1$. From these results, we can see that our method achieves spectral accuracy. We also remark that a larger λ makes the kernel more oscillatory, thus requiring a smaller mesh size h for adequate resolution.

6. Conclusions

Based on a far-field smooth approximation of the kernel, we presented a simple spectral fast algorithm with nearly optimal memory requirement $\mathcal{O}(2^d N^d)$ and computational cost $\mathcal{O}(2^d N^d \log(2^d N^d))$ for calculating nonlocal potentials in both isotropic and anisotropic cases. The smoothed kernel U^ε approximates the kernel at the far field with great accuracy, and it leads to an explicit radial symmetric substitute for the Fourier transform of $(U - U^\varepsilon)$ on bounded domain. We split the potential into two parts, and each quadrature has discrete convolution structure. By combing both convolution structures, our method can be simplified to a single discrete convolution with explicit tensor formulation, which can be accelerated by FFT on a double-sized vector. It is worthy to emphasize that the tensor generation is very simple, efficient and independent of the anisotropy strength. The performance superiorities of our method were showcased with common potentials, including the Coulomb, Poisson, Biharmonic, Yukawa, Helmholtz, Davey-Stewartson nonlocal potentials and DDI.

CRediT authorship contribution statement

Xin Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation; **Yong Zhang:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Xin Liu and Yong Zhang reports financial support was provided by the National Natural Science Foundation of China No. 12271400, the National Key R&D Program of China No. 2024YFA1012803 and basic research fund of Tianjin University under grant 2025XJ21-0010. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Error estimates

For the regular integral $\Phi^R(\mathbf{x})$, the error arises from the trapezoidal rule, which is spectrally accurate because the integrand is smooth and decays exponentially [7]. For the singular integral $\Phi^S(\mathbf{x})$, we have

$$|E_s(\mathbf{x})| = \left| \int_{\mathbf{R}_{2L}^\gamma} (U - U^\varepsilon)(\mathbf{y}) \rho(\mathbf{x} - \mathbf{y}) d\mathbf{y} - \sum_{\mathbf{k} \in \Lambda} \widehat{(U - U^\varepsilon)}(\mathbf{k}) \widetilde{\rho}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} \right|$$

$$\leq \Big| \int_{\mathbf{R}_{2L}^{\gamma}} (U - U^{\varepsilon})(\mathbf{y})(\rho - \rho_N)(\mathbf{x} - \mathbf{y}) \mathrm{d}\mathbf{y} \Big| + \Big| \sum_{\mathbf{k} \in \Lambda} \Big(W - (\widehat{U - U^{\varepsilon}}) \Big)(\mathbf{k}) \widetilde{\rho}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} \Big| := I_1 + I_2.$$

The error I_1 is of spectral accuracy because $(U - U^{\varepsilon}) \in L^1(\mathbb{R}^d)$ [14]. For I_2, we obtain

$$I_2 \leq \max_{\mathbf{k} \in \Lambda} \Big| \int_{\mathbb{R}^d \setminus \mathbf{R}_{2L}^{\gamma}} (U - U^{\varepsilon})(\mathbf{y}) e^{-i\mathbf{k} \cdot \mathbf{y}} \mathrm{d}\mathbf{y} \Big| \sum_{\mathbf{k} \in \Lambda} |\widetilde{\rho}_{\mathbf{k}}| \lesssim \int_{\mathbb{R}^d \setminus \mathbf{R}_{2L}^{\gamma}} |(U - U^{\varepsilon})(\mathbf{y})| \mathrm{d}\mathbf{y}$$

$$\leq \int_{\mathbb{R}^d \setminus \mathbf{B}_{R_0}} |(U - U^{\varepsilon})(\mathbf{y})| \mathrm{d}\mathbf{y} = S^{d-1} \int_{R_0}^{\infty} |(U - U^{\varepsilon})(r)| r^{d-1} \mathrm{d}r \lesssim \varepsilon_{\text{tol}},$$

where $\mathbf{B}_{R_0}$ is a ball centered at the origin with radius R_0 and $S^{d-1} = 2\pi^{\frac{d}{2}}/\Gamma(d/2)$. Therefore, our method can achieve spectral accuracy with a suitable ε.

Appendix B. The choice of parameter ε

The parameter ε is chosen to guarantee condition (b) of Eqn. (1.4) is satisfied, and it varies with the kernel. For the 3D Coulomb kernel $U(r) = 1/(4\pi r)$, we have

$$\int_{R_0}^{\infty} |(U - U^{\varepsilon})(r)| \ r^2 \ \mathrm{d}r = \frac{1}{4\pi} \int_{R_0}^{\infty} r \, \mathrm{Erfc}\left(\frac{r}{\varepsilon}\right) \mathrm{d}r = \frac{1}{4\pi} \varepsilon^2 \int_{R_0/\varepsilon}^{\infty} s \, \mathrm{Erfc}(s) \mathrm{d}s$$

$$= R_0^2 \Big[\frac{1}{4\pi c^2} \int_c^{\infty} s \, \mathrm{Erfc}(s) \mathrm{d}s \Big] := R_0^2 f(c), \quad \text{where } c := R_0/\varepsilon.$$

We choose c such that $f(c) R_0^2 \leq \varepsilon_{\text{tol}}$, i.e., $f(c) \leq \varepsilon_{\text{tol}}/R_0^2$. Clearly, $f(c)$ is monotonically decreasing and the critical value of c depends on R_0 and ε_{tol}. For example, for $R_0 = 24$, we can set $c \geq 5.85$ (8.65), i.e., $\varepsilon \leq R_0/5.85$ ($R_0/8.65$) to achieve 16 (34) digits of accuracy. Similarly, for the 2D Coulomb and Poisson potentials, we can set $\varepsilon \leq R_0/5.64$ and $\varepsilon \leq R_0/5.75$ respectively to guarantee 16 digits of accuracy.

References

[1] W. Bao, Y. Cai, Mathematical theory and numerical methods for Bose-Einstein condensation, Kinet. Relat. Mod. 6 (2013) 1–135.
[2] A. Cerioni, L. Genovese, A. Mirone, V. Sole, Efficient and accurate solver of the three-dimensional screened and unscreened Poisson's equation with generic boundary conditions, J. Chem. Phys. 137 (2012) 134108.
[3] T. Lahaye, C. Menotti, L. Santos, M. Lewenstein, T. Pfau, The physics of dipolar bosonic quantum gases, Rep. Prog. Phys. 72 (2009) 126401.
[4] S. Yi, L. You, Trapped atomic condensates with anisotropic interactions, Phys. Rev. A 61 (2000) 041604.
[5] S. Yi, L. You, Trapped condensates of atoms with dipole interactions, Phys. Rev. A 63 (2001) 053607.
[6] W. Bao, H. Jian, N.J. Mauser, Y. Zhang, Dimension reduction of the Schrödinger equation with Coulomb and anisotropic confining potentials, SIAM J. Appl. Math. 73 (2013) 2100–2123.
[7] J. Shen, T. Tang, L. Wang, Spectral Methods: Algorithms, Analysis and Applications, Springer, 2011.
[8] W. Bao, S. Jiang, Q. Tang, Y. Zhang, Computing the ground state and dynamics of the nonlinear Schrödinger equation with nonlocal interactions via the nonuniform FFT, J. Comput. Phys. 296 (2015) 72–89.
[9] W. Bao, Q. Tang, Y. Zhang, Accurate and efficient numerical methods for computing ground states and dynamics of dipolar Bose-Einstein condensates via the nonuniform FFT, Commun. Comput. Phys. 19 (2016) 1141–1166.
[10] S. Jiang, L. Greengard, W. Bao, Fast and accurate evaluation of nonlocal Coulomb and dipole-dipole interactions via the nonuniform FFT, SIAM J. Sci. Comput. 36 (2014) 777–794.
[11] L. Exl, N.J. Mauser, Y. Zhang, Accurate and efficient computation of nonlocal potentials based on Gaussian-sum approximation, J. Comput. Phys. 327 (2016) 629–642.
[12] F. Vico, L. Greengard, M. Ferrando, Fast convolution with free-space Green's functions, J. Comput. Phys. 323 (2016) 191–203.
[13] M.R. Jarvis, I.D. White, R.W. Godby, M.C. Payne, Supercell technique for total-energy calculations of finite charged and polar systems, Phys. Rev. B 56 (1997) 14972–14978.
[14] X. Liu, Q. Tang, S. Zhang, Y. Zhang, On optimal zero-padding of kernel truncation method, SIAM J. Sci. Comput. 46 (2024) A23–A49.
[15] C.A. Rozzi, D. Varsano, A. Marini, E.K.U. Gross, A. Rubio, Exact Coulomb cutoff technique for supercell calculations, Phys. Rev. B 73 (2006) 205119.
[16] L. Greengard, S. Jiang, Y. Zhang, The anisotropic truncated kernel method for convolution with free-space Green's functions, SIAM J. Sci. Comput. 40 (2018) A3733–A3754.
[17] X. Antoine, Q. Tang, Y. Zhang, A preconditioned conjugated gradient method for computing ground sates of rotating dipolar Bose-Einstein condensates via kernel truncation method for dipole-dipole interaction evaluation, Commun. Comput. Phys. 24 (2018) 966–988.
[18] J. Bagge, A.-K. Tornberg, Fast Ewald summation for Stokes flow with arbitrary periodicity, J. Comput. Phys. 493 (2023) 112473.
[19] P.P. Ewald, Die berechnung optischer und elektrostatischer gitterpotentiale, Ann. Phys. 369 (1921) 253–287.
[20] L. Greengard, S. Jiang, A dual-space multilevel kernel-splitting framework for discrete and continuous convolution, Commun. Pure Appl. Math. 78 (2025) 1086–1143.
[21] A.Y. Toukmajiand, J.A. Board, Ewald summation techniques in perspective: a survey, Comput. Phys. Commun. 95 (1996) 73–92.
[22] A.K. Tiwari, A. Pandey, J. Paul, A. Anand, A fast rapidly convergent method for approximation of convolutions with applications to wave scattering and some other problems, J. Comput. Phys. 459 (2022) 111119.
[23] G. Beylkin, L. Monzón, Approximation by exponential sums, Appl. Comput. Harmon. Anal. 19 (2005) 17–48.
[24] M. Abramowitz, I.A. Stegun, Handbook of mathematical functions, Dover, 1965.
[25] Y. Cai, M. Rosenkranz, Z. Lei, W. Bao, Mean-field regime of trapped dipolar Bose-Einstein condensates in one and two dimensions, Phys. Rev. A 82 (2010) 043623.