Fast one-dimensional convolution with general kernels using sum-of-exponential approximation

Yong Zhang[a], Chijie Zhuang[b,*], Shidong Jiang[c]

[a] *Center for Applied Mathematics, Tianjin University, Tianjin 300072, China*
[b] *Department of Electrical Engineering, Tsinghua University, Beijing 100084, China*
[c] *Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey, 07102, USA*

Abstract

Based on the recently-developed sum-of-exponential (SOE) approximation, in this article, we propose a fast algorithm to evaluate the one-dimensional convolution potential $\phi(x) = K * \rho = \int_0^1 K(x-y)\rho(y)dy$ at (non)uniformly distributed target grid points $\{x_i\}_{i=1}^M$, where the kernel $K(x)$ might be singular at the origin and the source density function $\rho(x)$ is given on a source grid $\{y_j\}_{j=1}^N$ which can be different from the target grid. It achieves an optimal accuracy, inherited from the interpolation of the density $\rho(x)$, within $O(M+N)$ operations. Using the kernel's SOE approximation $K_{\rm ES}$, the potential is split into two integrals: the exponential convolution $\phi_{\rm ES} = K_{\rm ES} * \rho$ and the local correction integral $\phi_{\rm cor} = (K - K_{\rm ES}) * \rho$. The exponential convolution is evaluated via the recurrence formula that is typical of the exponential function. The local correction integral is restricted to a small neighborhood of the target point where the kernel singularity is considered. Rigorous estimates of the optimal accuracy are provided. The algorithm is ideal for parallelization and favors easy extensions to complicated kernels. Extensive numerical results for different kernels are presented.

Keywords: one dimensional convolution, sum of exponentials, singular kernel, discrete density

1. Introduction

Pairwise interaction is common and important in computational physics and practical engineering, and it is usually long-ranged and described by a continuous/discrete convolution. For example, the electrostatic interactions of charged carriers are essential in simulating lightning, blue jet/gigantic jet in atmospheric science, or in the corona discharges around power transmission lines [12]. Such interactions are originally three dimensional and the evaluations will cost huge amount of computational resources. Reduction to lower dimensional convolutions is common and necessary for meaningful long-time simulations [17, 21]. Here we focus on the following rescaled one-dimensional convolution

$$\phi(x) = \int_0^1 K(x-y)\rho(y)dy, \quad \forall\ x \in [0,1], \tag{1.1}$$

where the kernel $K(x), x \in [-1,1]$ might be singular at the origin and $\rho(x)$ is the source density. Numerically, the density $\rho(x)$ may be given on a discrete *source* grid of N points, denoted as $\mathcal{S} := \{y_j\}_{j=0}^N$ which might be distributed nonuniformly, and the *target* grid of M points, denoted as $\mathcal{T} := \{x_i\}_{i=0}^M$, *does not* necessarily coincide with the source grid. Usually, the source grid is given as finite difference/element/volume grid and

*Corresponding author.
Email addresses: Zhang_Yong@tju.edu.cn (Yong Zhang), chijie@tsinghua.edu.cn (Chijie Zhuang), shidong.jiang@njit.edu (Shidong Jiang)

is often nonuniform, so is the target grid. In this paper, we aim to design an accurate and fast algorithm for (1.1) on general grids.

When both the source and target grids are uniformly distributed, the evaluation (1.1) boils down to a discrete convolution [10, 14, 19] and can be accelerated via the discrete Fast Fourier Transform within $O(N \log N)$ operations. However, on nonuniform grids, a simple direct summation of the resulted quadrature costs $O(MN)$ operations, and usually bottlenecks practical simulations, therefore, it is imperative to design fast algorithms for better efficiency while maintaining the accuracy. There have been several work dealing with such convolutions. In 1999, Yarvin and Rokhlin proposed an improved Fast Multipole Method [16] for the one dimensional discrete convolution with singular kernels, including x^{-1}, $\log(x)$ and $x^{-1/2}$. Subsequently, Beylkin [4, 5] designed a fast discrete convolution algorithm of complexity $O(NQ)$, where Q is the number of exponentials, for a wider range of kernels. Recently, Greengard et al [8] proposed an algorithm for finding nearly optimal SOE approximation for non-oscillatory functions, and applied the SOE to calculate high dimensional spatial volume convolutions. Similar ideas using the recurrence scheme of the exponential function has been developed and applied to many problems in various fields, e.g., the temporal convolution whose integration domain is $[0, t]$ rather than the whole interval $[0, T]$, in the context of nonreflecting boundary condition of the Schrödinger equation, wave equation, and fractional temporal derivatives (see, for example [2, 13, 18]).

To compute the potential ϕ, we first construct a sum-of-exponential approximation of the kernel

$$K_{\mathrm{ES}}(x) := \sum_{q=1}^{Q} \omega_q e^{-\alpha_q x}, \qquad \text{with } \omega_q,\ \alpha_q \in \mathbb{C}, \quad \Re(\alpha_q) > 0, \tag{1.2}$$

such that

$$\|K(x) - K_{\mathrm{ES}}(x)\|_\infty \le \varepsilon, \quad \forall\ x \in [\delta, 1], \tag{1.3}$$

with a small truncation parameter $0 < \delta < 1$ and a prescribed accuracy $0 < \varepsilon \ll 1$, which can be controlled as small as possible, e.g., $10^{-14} \sim 10^{-12}$. The number of exponentials Q often depends on the prescribed precision ε logarithmically [7, 9]. Then, using the above SOE approximation, the potential is split as

$$\begin{aligned}
\phi(x) &= \int_0^1 (K - K_{\mathrm{ES}})(x-y)\rho(y)dy + \int_0^1 K_{\mathrm{ES}}(x-y)\rho(y)dy \\
&= \left(\int_{[0,1]\cap[x-\delta,x+\delta]} + \int_{[0,1]\setminus[x-\delta,x+\delta]} \right) (K - K_{\mathrm{ES}})\,(x-y)\ \rho(y)\ dy + \phi_{\mathrm{ES}}(x) \\
&:= \phi_{\mathrm{cor}}(x) + \phi_\delta(x) + \phi_{\mathrm{ES}}(x).
\end{aligned} \tag{1.4}$$

The exponential convolution

$$\phi_{\mathrm{ES}}(x) := \int_0^1 K_{\mathrm{ES}}(x-y)\rho(y)dy = \sum_{q=1}^{Q} \omega_q \int_0^1 e^{-\alpha_q|x-y|}\rho(y)dy, \tag{1.5}$$

can be computed within $O((M+N)Q)$ operations for each exponentials via recurrence formula, and the correction integral

$$\phi_{\mathrm{cor}}(x) := \int_{[0,1]\cap[x-\delta,x+\delta]} (K - K_{\mathrm{ES}})(x-y)\rho(y)dy, \tag{1.6}$$

will be evaluated pointwisely which also costs $O(M)$ operations. The remainder integral $\phi_\delta(x)$ is negligible because

$$|\phi_\delta(x)| \le \int_{[0,1]\setminus[x-\delta,x+\delta]} |(K - K_{\mathrm{ES}}(x-y)|\,|\rho(y)|dy \le \varepsilon \|\rho\|_{L^1}. \tag{1.7}$$

Therefore, we have

$$\phi(x) \approx \phi_{\mathrm{ES}}(x) + \phi_{\mathrm{cor}}(x) + O(\varepsilon). \tag{1.8}$$

The convolution evaluation boils down to the computation of the exponential convolution $\phi_{\mathrm{ES}}(x)$ and the local correction integral $\phi_\delta(x)$ on the given discrete target grid.

The paper is organized as follows. In Section 2, we first briefly review the algorithm for finding the SOE approximation, then describe the discrete exponential convolution and local correction integral computation in details, followed by a discussion on the extension to asymmetric kernels and the error estimates. Extensive numerical results are reported in Section 3. Finally, some conclusions are drawn in Section 4.

2. Numerical algorithm

In this section, we briefly review the black-box algorithm for constructing the SOE approximation, and introduce the computation of the exponential convolution $\phi_{\mathrm{ES}}(x_i)$ and local correction integral $\phi_{\mathrm{cor}}(x_i)$ in details. A discussion on the extension to asymmetric kernels and the error estimate are provided in the last two subsections. Note that the density $\rho(x)$ is given on a discrete N-point source grid $\mathcal{S}$, and we aim to evaluate the potential on M-point *target* grid $\mathcal{T}$ which does not necessarily coincide with the *source* grid.

As described earlier, the evaluation is composed of three parts, i.e., the SOE approximation, the exponential convolution and the local correction integral. For sake of better readability, we first present the algorithm below, then add details in later subsections.

Algorithm 1 Generalized SOE based fast convolution algorithm.

Require: Given a precision requirement ε, the computational grid $\{y_j\}_{j=0}^N$, the convolution kernel K, and the discrete density ρ_j, compute the potential $\phi(x_i)$ defined in (1.1).

1: Precomputation stage: select appropriate δ and construct the SOE approximation of K on $[\delta, 1]$.
2: Compute the potential $\phi_{\mathrm{ES}}(x_i)$ as described in subsection (2.2).
3: Compute the potential $\phi_{\mathrm{cor}}(x_i)$ as shown in subsection (2.3).
4: Sum up $\phi_{\mathrm{ES}}(x_i)$ and $\phi_{\mathrm{cor}}(x_i)$ to achieve $\phi(x_i)$.

2.1. Black-box algorithm for the SOE approximation

The algorithm was first described in [8], and consists of two steps. First, an accurate but inefficient SOE approximation is obtained via a least squares procedure. That is, we select a large set of the exponential nodes in such a way that the associated set of exponential functions are sufficient to represent the given function, determine an oversampled grid on the given interval for the given function, and calculate the weights of the SOE approximation by matching the function values on this discrete grid with the SOE approximation in the sense of least squares. The least squares problem is simultaneously overdetermined and rank-deficient, but can be solved in a backward stable fashion, namely, via an SVD based algorithm. This will lead to a solution with a prescribed residual error. Second, a standard algorithm in model reduction is applied to reduce the number of exponentials to achieve a nearly optimal SOE approximation. For details, we refer the readers to [11]. Related SOE codes are now available at Shidong JIANG's homepage https://web.njit.edu/ jiang/pub.html.

2.2. Fast exponential convolution via recurrence: $\phi_{\mathrm{ES}}(x_i)$

We notice that it is straightforward to parallelize the Q-term exponential convolution (1.5), and it is sufficient to describe the algorithm for one-exponential convolution, i.e., the convolution of $e^{-\alpha x}$ with the density. It is clear that the convolution with the exponential function can be viewed as the solution to the ordinary differential equation $u' = -\alpha u + \rho$. Thus, such convolution can be computed in linear time using a

simple marching scheme. Alternatively, using the recurrence formula that stems from *translation invariant* property of the exponential [4, 16], the potential at $x_i, i = 1, \ldots, M$ can be rewritten as follows

$$\begin{aligned} I(x_i) &:= \int_0^1 e^{-\alpha|x_i-y|}\rho(y)dy = \sum_{j=1}^{N}\int_{y_{j-1}}^{y_j} e^{-\alpha|x_i-y|}\rho(y)dy \\ &= \sum_{y_j\le x_i}\int_{y_{j-1}}^{y_j} e^{-\alpha(x_i-y)}\rho(y)dy + \int_{y_{l_0-1}}^{y_{l_0}} e^{-\alpha|x_i-y|}\rho(y)dy + \sum_{x_i<y_j}\int_{y_j}^{y_{j+1}} e^{-\alpha(y-x_i)}\rho(y)dy \\ &:= P_i + L_i + J_i, \end{aligned}$$

where P_i, L_i and J_i are the first, second and last integral respectively and index $l_0 \in \mathbb{N}$ is determined such that x_i lies in $[y_{l_0-1}, y_{l_0}]$.

We shall treat the above three terms respectively. For the third term J_i, we have

$$\begin{aligned} J_i &= e^{-\alpha(x_{i+1}-x_i)}\sum_{x_{i+1}<y_j}\int_{y_j}^{y_{j+1}} e^{-\alpha(y-x_{i+1})}\rho(y)dy + \sum_{x_i<y_j\le x_{i+1}}\int_{y_j}^{y_{j+1}} e^{-\alpha(y-x_i)}\rho(y)dy \\ &= e^{-\alpha(x_{i+1}-x_i)}J_{i+1} + \sum_{x_i<y_j\le x_{i+1}}\int_{y_j}^{y_{j+1}} e^{-\alpha(y-x_i)}\rho(y)dy := e^{-\alpha(x_{i+1}-x_i)}J_{i+1} + J_i^C. \end{aligned} \tag{2.9}$$

Once J_i^C is available for each i, all the J_i can be computed recursively from J_M to J_1. In fact, all the J_i^C can be calculated explicitly. Specially, consider a linear interpolation of $\rho(y)$ over $[y_j, y_{j+1}]$, we have

$$J_i^C = \sum_{x_i<y_j\le x_{i+1}}\int_{y_j}^{y_{j+1}} e^{-\alpha(y-x_i)}\left(\rho_{j+1}\frac{y-y_j}{y_{j+1}-y_j} + \rho_j\frac{y_{j+1}-y}{y_{j+1}-y_j}\right)dy = \sum_{x_i<y_j\le x_{i+1}}(\rho_j\Gamma_j + \rho_{j+1}\Gamma_{j+1}),$$

where the local coefficients Γ_j, Γ_{j+1} have analytical formula, depending on y_j, y_{j+1}, x_i and the exponent α, and we omit it here for brevity. High order polynomial interpolation leads to similar local multiplications. Therefore, one can compute all the J_i within $O(M + N)$ operations. Similarly, we can calculate the first integral P_i recursively within $O(M + N)$ operations. All the integral L_i can be computed within $O(M)$ operations.

Collecting all the Q-term exponential (1.2), we get $\phi_{\rm ES}(x_i)$. The overall computation cost is $O((M+N)Q)$ with a storage requirement $O(M+N+Q)$.

2.3. Fast local correction: $\phi_{\rm cor}(x_i)$

As is seen that SOE well approximates $K(x)$ for larger x, which results in a numerically compact supported integrand in (1.6), the effective integration interval is restricted to a small neighborhood of x_i, i.e.,

$$\phi_{\rm cor}(x_i) = \int_{[0,1]\cap[x_i-\delta,x_i+\delta]}(K - K_{\rm ES})(x_i - y)\rho(y)dy. \tag{2.10}$$

The effective interval may span several elements, and the element number depends on δ and grid adaptivity. For example, for x_i, one will go through $[y_{i_0}, y_{i_0+1}], \ldots, [y_{i_1-1}, y_{i_1}]$ where y_{i_0} is the largest point that is no greater than $x_i - \delta$ and y_{i_1} is the smallest point which is no less than $x_i + \delta$.

To integrate over the effective intervals, we first interpolate the density $\rho(x)$ using local polynomials, then integrate the SOE and kernel part respectively. The SOE integration is computed the same way as shown above, and the kernel part is reduced to convolution with polynomials. The kernel integration may have analytical expressions, e.g., convolution with the power function $1/x^\alpha$ results in polynomial-form solutions. While for a more general kernel, it is usually not possible, or at least it takes great efforts to find

explicit solutions. In such cases, it is feasible to replace $K(x)$ by its generalized m-th order Taylor series approximation, i.e.,

$$K(x) \approx K_m(x) := C_0(x) + C_1\ x + C_2\ x^2 + \ldots + C_m\ x^m, \quad \text{for } |x| < \delta, \tag{2.11}$$

where $C_0(x)$ is the leading order asymptotic of the kernel, and might be singular. With such substitution, the kernel contribution boils down to a polynomial-polynomial convolution, and it is polynomial-form solution too. Discrepancy function $E^m_{\rm cor}(x_i) := \int_{[0,1]\cap[x_i-\delta,x_i+\delta]}(K-K_m)(x_i-y)\rho(y)dy$ is estimated as follows:

$$\begin{aligned} |E^m_{\rm cor}(x_i)| &\leq \Big|\int_{x_i-\delta}^{x_i+\delta}(K-K_m)(x_i-y)\rho(y)dy\Big| \\ &\leq \int_{-\delta}^{\delta}\Big|(K-K_m)(y)\Big|\ \ \big|\rho(x_i-y)\big|dy \\ &\leq C\ \delta^{m+1}\int_{-\delta}^{\delta}\big|\rho(x_i-y)\big|dy \leq C\ \delta^{m+1}\|\rho\|_{L^1}, \end{aligned}$$

and it immediately implies $\|E^m_{\rm cor}\|_\infty \leq C\delta^{m+1}\|\rho\|_{L^1}$. The integer m and parameter δ are chosen such that $\|E^m_{\rm cor}\|$ is small enough. For weak singular and nearly singular kernels, it is important to have some *a-priori* asymptotics around the singularity point.

The number of effective intervals for each x_i depends on δ and the grid adaptively, and is basically $O(1)$, therefore the total cost of computing all the $\phi_{\rm cor}(x_i)$ is $O(M)$. Summing up $\phi_{\rm ES}(x_i)$ and $\phi_{\rm cor}(x_i)$, we have the potential $\phi(x_i)$. In total, the overall computation cost is $O((M+N)Q) + O(M)$, and the maximum storage requirement is $O(M+N+Q)$.

Remark 2.1. *We remark here that it is easy to parallelize the second step, which takes up most of the computational efforts, leading to a further reduction on the computational time if needed.*

2.4. Adaptation to asymmetric kernels

The above algorithm works only for symmetric kernels, i.e., $K(-x) = K(x)$, and cannot be applied directly to the asymmetric case. For asymmetric kernels, we should modify the algorithm as follows. First, instead of approximating the kernel over $[\delta, 1]$ by a single SOE approximation, we construct two different SOE approximations over $[\delta, 1]$ and $[-1, -\delta]$, denoted as $K^+_{\rm ES}$ and $K^-_{\rm ES}$ with numbers of exponentials Q^+ and Q^- respectively.

For the exponential convolution (2.2), in order to utilize the recurrence formula, we shall adopt $K^+_{\rm ES}(x)$ for positive x, $K^-_{\rm ES}$ for negative x. The overall computation cost is $O((M+N)(Q^+ + Q^-))$ with a storage requirement $O(M+N+Q^+ + Q^-)$. For the local correction integral (2.3), only the convolution of SOE and the polynomials needs to updated, where the kernel is now switched as $K^+_{\rm ES}$ and $K^-_{\rm ES}$.

2.5. Error estimates

The convolution is decomposed as $\phi(x) = \phi_{\rm ES}(x) + \phi_{\rm cor}(x) + \phi_\delta(x)$ as shown in (1.4). According to (1.7), we do not need to compute ϕ_δ. Therefore, the numerical evaluation $\phi_h(x)$ consists of two integrals

$$\phi_h(x) = \phi^h_{\rm ES}(x) + \phi^h_{\rm cor}(x), \tag{2.12}$$

where $\phi^h_{\rm ES}, \phi^h_{\rm cor}$ are approximations of $\phi_{\rm ES}, \phi_{\rm cor}$ by replacing the density ρ with its approximation $\rho_h(x)$, and $h = \max_{0\leq j<N}\ (x_{j+1} - x_j)$ denotes the largest mesh size of the discrete grid.

For $\rho(x) \in C^2([0,1])$, assuming $\rho_h(x)$ is a linear interpolation approximation of $\rho(x)$, by standard numerical analysis, we know that the interpolation error is

$$|\rho(x) - \rho_h(x)| \leq C\|\rho^{(2)}\|_\infty (y_{j+1} - y_j)^2 \leq C\|\rho^{(2)}\|_\infty\ h^2, \forall\ x \in [y_j, y_{j+1}],\ j = 0, \ldots, N-1.$$

First, the error estimate of the exponential convolution reads as follows

$$\begin{aligned}
|\phi_{\rm ES}(x) - \phi^h_{\rm ES}(x)| &= \left| \sum_{q=1}^{Q} \omega_q \sum_{j=0}^{N-1} \int_{y_j}^{y_{j+1}} e^{-\alpha_q|x-y|}(\rho(y) - \rho_h(y))dy \right| \\
&\leq Ch^2\|\rho^{(2)}\|_\infty \left(\sum_{q=1}^{Q} |\omega_q| \sum_{j=0}^{N-1} \int_{y_j}^{y_{j+1}} e^{-\alpha_q|x-y|}dy \right) \\
&\leq Ch^2\|\rho^{(2)}\|_\infty \left(\sum_{q=1}^{Q} |\omega_q| \right) \leq C(K,\delta,\varepsilon)\ h^2\|\rho^{(2)}\|_\infty,
\end{aligned}$$

where $C(K,\delta,\varepsilon)$ is bounded and depends on the SOE approximation, and it immediately implies

$$\|\phi_{\rm ES} - \phi^h_{\rm ES}\|_\infty \leq C(K,\delta,\varepsilon)\ h^2\|\rho^{(2)}\|_\infty.$$

Second, the error estimates for the correction integral are

$$\begin{aligned}
|\phi_{\rm cor}(x) - \phi^h_{\rm cor}(x)| &= \left| \int_{[0,1]\setminus[x-\delta,x+\delta]} (K - K_{\rm ES})(x-y)\ (\rho(y) - \rho_h(y))\ dy \right| \\
&\leq Ch^2\|\rho^{(2)}\|_\infty \left| \int_{x-\delta}^{x+\delta} (K - K_{\rm ES})(x-y)dy \right| \\
&\leq Ch^2\|\rho^{(2)}\|_\infty \left(\Big| \int_{-\delta}^{\delta} K(t)dt \Big| + \Big| \int_{-\delta}^{\delta} K_{\rm ES}(t)dt \Big| \right) \\
&\leq Ch^2\|\rho^{(2)}\|_\infty \left(C(K,\delta) + \delta\ C(K,\delta,\varepsilon) \right)
\end{aligned}$$

where $C(K,\delta) = |\int_{-\delta}^{\delta} K(t)dt|$ is bounded because the convolution (1.1) is well-defined, and $C(K,\delta,\varepsilon)$ is defined as above and bounded too. Therefore, we have the following estimates

$$\|\phi_{\rm cor} - \phi^h_{\rm cor}\|_\infty \leq Ch^2\|\rho^{(2)}\|_\infty (C(K,\delta) + \delta\ C(K,\delta,\varepsilon)). \tag{2.13}$$

Finally, we have the total estimates

$$\|\phi - \phi_h\|_\infty \leq C(K,\delta,\varepsilon)h^2\|\rho^{(2)}\|_\infty + \varepsilon\|\rho\|_{L^1}, \tag{2.14}$$

where $C(K,\delta,\varepsilon)$ is a bounded constant depending on the kernel K and SOE parameters δ,ε.

3. Numerical results

To demonstrate the accuracy and efficiency, we carry out several numerical experiments. All the numerical errors are calculated in the relative maximum norm, defined as follows:

$$E_h := \frac{\|\phi - \phi_h\|_{l^\infty}}{\|\phi\|_{l^\infty}} = \frac{\max_{x_i\in\mathcal{T}_h} |\phi(x_i) - \phi_h(x_i)|}{\max_{x_i\in\mathcal{T}_h} |\phi(x_i)|}. \tag{3.15}$$

Here ϕ_h is the numerical solution and ϕ is the reference solution. When the reference solution is not available analytically, E_h is defined as $\|\phi_h - \phi_{h/2}\|_{l^\infty} := \max_{x_i\in\mathcal{T}_h} |\phi_h(x_i) - \phi_{h/2}(x_i)|$. The convergence order is calculated via the formula $\log(E_h/E_{h/2})/\log(2)$. The computational time shown here is calculated as the average of 50 evaluations without any parallel acceleration. The interval is uniformly discretized on the grid $\mathcal{T}_h := \left\{x_j = jh = j\frac{1}{N}, j = 0,\ldots,N\right\}$. For simplicity, the target points are chosen the same as source points unless otherwise stated.

The algorithm has been implemented in FORTRAN, and all reported timing results are obtained using a single 2.60GHz Intel(R) Core(TM) i7-6660U CPU with a 4-MB cache with the Intel compiler ifort and optimization level -O3.

3.1. Multiquadrics

We consider multiquadric kernels that are quite common in electrostatic computation[20, 21]. The kernel is $K(x) = \frac{1}{\sqrt{x^2+a^2}}$, $0 < a \ll 1$, and the potential $\phi(x)$ generated by some simple source functions are given explicitly as

$$\phi(x) = \begin{cases} \log\left(\frac{\sqrt{(1-x)^2+a^2}+(1-x)}{\sqrt{x^2+a^2}-x}\right), & \rho(x) = 1, \\ \sqrt{a^2+(1-x)^2} - \sqrt{a^2+x^2} + x\log\left(\frac{(1-x)+\sqrt{a^2+(1-x)^2}}{\sqrt{a^2+x^2}-x}\right), & \rho(x) = x. \end{cases} \tag{3.16}$$

In this example, the kernel parameter is chosen $a = 10^{-3}$. The local correction integral is computed exactly, thus waiving the Taylor approximation (2.11). The SOE approximation parameters are chosen as $\varepsilon = 10^{-12}, \delta = 10^{-8}$, and the number of exponentials is 139.

First, we present the accuracy and efficiency results for a linear source function $\rho(x) = \frac{1}{2}(1+x)$ for different N in Table 1. The exact solution $\phi(x)$ is a linear combination of (3.16). As the source function is fully resolved by piecewise elements, the accuracy depends only on the approximation parameters ε and δ, which are sufficiently small to ensure accurate evaluation, see (2.14). The computational time scales linearly with N as expected, and it takes about 3 seconds for one million points.

Second, we choose a Gaussian density $\rho(x) = e^{-4(x-\frac{1}{2})^2}$ and adopt a linear element approximation. Table 2 displays the successive errors E_h for different mesh sizes, and the convergence rates. A second order convergence in mesh size h is clearly shown in Table 2.

Table 1: Accuracy and CPU time (in seconds) for multiquardics $K(x) = \frac{1}{\sqrt{x^2+a^2}}$ with different N.

N	10^2	10^3	10^4	10^5	10^6
E_h	3.564E-11	1.208E-10	2.183E-11	1.576E-10	3.629E-9
Time	4.190E-4	2.948E-3	2.676E-2	2.950E-1	3.686

Table 2: Second order convergence for multiquardics kernel with Gaussian density $\rho(x) = e^{-4(x-\frac{1}{2})^2}$ and $h_0 = 10^{-3}$.

h/h_0	1	1/2	1/4	1/8	1/16
E_h	5.876E-6	1.469E-6	3.675E-7	9.204E-8	2.307E-8
Rate	2.000	1.999	1.998	1.996	-

3.2. Power function

We consider the power kernel $K(x) = x^{-\alpha}$ with $0 < \alpha < 1$, which has close connections with temporal fractional derivative. The potential $\phi(x)$ generated by a linear source $\rho(x)$ is given explicitly as

$$\phi(x) = \begin{cases} \frac{1}{1-\alpha}\left(x^{1-\alpha} + (1-x)^{1-\alpha}\right), & \rho(x) = 1, \\ \frac{x^{(2-\alpha)}}{(1-\alpha)(2-\alpha)} + \frac{x(1-x)^{1-\alpha}}{1-\alpha} + \frac{(1-x)^{2-\alpha}}{2-\alpha}, & \rho(x) = x. \end{cases} \tag{3.17}$$

The SOE approximation parameters are chosen as $\varepsilon = 10^{-12}, \delta = 10^{-6}$, and the number of exponentials for different α is around 120. Similarly to the multiquardics kernel, the local correction integral is also computed exactly, thus waiving the Taylor approximation (2.11).

First, we present the number of exponentials Q and the accuracy results for a linear source function $\rho(x) = \frac{1}{2}(1+x)$ in Table 3 on uniform grids and the Chebyshev grids ($x_j = \frac{1}{2}(1-\cos(\frac{\pi j}{N}))$) for different exponents α, and the exact solution $\phi(x)$ is a linear combination of (3.17). Second, CPU time (in seconds) is shown in Table 4 for the randomly distributed grid with $\alpha = \frac{1}{2}$, from which the linear scaling with respect to N is clearly observed. Finally, we present the successive errors E_h and its convergence rate for the Gaussian density $\rho(x) = e^{-x^2}$ in Table 5, from which one can see clearly a second order convergence rate.

Table 3: Accuracy for the power kernel $K(x) = x^{-\alpha}$ with $N = 10^4$ on the uniform and Chebyshev grid.

α	0.25	0.5	0.75	0.85	0.95	0.99
Q	122	123	125	125	127	127
Uniform	1.964E-11	2.898E-10	7.606E-9	7.523E-9	1.390E-8	4.012E-9
Chebyshev	1.830E-11	3.255E-10	4.704E-9	1.108E-8	1.418E-8	4.766E-9

Table 4: CPU time (in seconds) for the power kernel $K(x) = x^{-\frac{1}{2}}$ case on randomly nonuniform grids.

$N/10^4$	1	2	4	8	16	32	64
Time	2.921E-02	5.910E-02	1.234E-01	2.828E-01	5.989E-01	1.297	2.296

Table 5: Second order convergence for power kernels with Gaussian density $\rho(x) = e^{-x^2}$ and $h_0 = 0.1$.

	h/h_0	1	1/2	1/4	1/8	1/16	1/32
$\alpha = \frac{1}{4}$	E_h	8.808E-4	2.247E-4	5.685E-5	1.431E-5	3.592E-6	9.000E-7
	Rate	1.971	1.983	1.990	1.994	1.997	-
$\alpha = \frac{1}{2}$	E_h	1.732E-3	4.600E-4	1.194E-4	3.053E-5	7.819E-6	1.937E-6
	Rate	1.913	1.946	1.967	1.965	2.013	-
$\alpha = \frac{3}{4}$	E_h	3.556E-3	1.006E-3	2.733E-4	7.295E-5	1.935E-5	4.991E-6
	Rate	1.822	1.880	1.905	1.915	1.955	-

3.3. Kernel arising from photoionization

We consider a cylindrically symmetric density, i.e., $\rho(x,y,z) = \rho(r,0,z),\ r = \sqrt{x^2+y^2}$. The original three dimensional convolution

$$\phi(\mathbf{x}) = U * \rho, \quad \text{with} \quad U(\mathbf{x}) = \frac{e^{-\lambda_1|\mathbf{x}|} - e^{-\lambda_2|\mathbf{x}|}}{|\mathbf{x}|^3}, \quad \mathbf{x} \in \mathbb{R}^3, \quad \lambda_1, \lambda_2 \in \mathbb{R}^+, \tag{3.18}$$

is also cylindrically symmetric. With a further assumption that the density is distributed uniformly over a small disk, i.e., $\frac{\partial \rho}{\partial r} = 0,\ 0 < r \le r_d$, the potential along z-axis is given below

$$\phi(z) \quad = \quad 2\pi \int_{z_1}^{z_2} \rho(z')dz' \int_0^{r_d} U\left(\sqrt{r^2+(z-z')^2}\right) rdr := 2\pi \int_{z_1}^{z_2} \rho(z')K(z-z')dz'. \tag{3.19}$$

The kernel $K(z)$ is symmetric and reads as follows

$$\begin{aligned} K(z) &= \int_0^{r_d} \frac{\exp(-\lambda_1 R) - \exp(-\lambda_2 R)}{R^3} r dr = \int_z^{R_d} \frac{\exp(-\lambda_1 R) - \exp(-\lambda_2 R)}{R^2} dR = \int_{\lambda_1}^{\lambda_2} d\lambda \int_z^{R_d} \frac{\exp(-\lambda R)}{R} dR \\ &= \frac{(e^{-\lambda_1 z} - e^{-\lambda_2 z})R_d - (e^{-\lambda_1 R_d} - e^{-\lambda_2 R_d})z}{R_d z} + \lambda_1(\Gamma_0(\lambda_1 R_d) - \Gamma_0(\lambda_1 z)) + \lambda_2(\Gamma_0(\lambda_2 z) - \Gamma_0(\lambda_2 R_d)) \end{aligned}$$

where $R = \sqrt{r^2 + z^2}$, $R_d = \sqrt{r_d^2 + z^2}$ and $\Gamma_0(z) = \Gamma(0, z) := \int_z^\infty t^{-1} e^{-t} dt$ is the incomplete gamma function[1]. The asymptotic analysis of $K(z)$ around $z = 0$ is

$$\begin{aligned} K(z) &\approx (\lambda_1 - \lambda_2)\log(z) + (\gamma_e - 1)(\lambda_1 - \lambda_2) \\ &+\lambda_1\Gamma_0(r_d\lambda_1) - \lambda_2\Gamma_0(r_d\lambda_2) + \lambda_1\log(\lambda_1) - \lambda_2\log(\lambda_2) + \frac{e^{-r_d\lambda_2} - e^{-r_d\lambda_1}}{r_d} + \frac{1}{2}(\lambda_2^2 - \lambda_1^2)z. \quad (3.20) \end{aligned}$$

The original convolution (3.19) is rescaled to the standard form

$$\widetilde{\phi}(\widetilde{x}) = 2\pi(z_1 - z_0)\int_0^1 \widetilde{K}(\widetilde{x} - \widetilde{y})\ \ \widetilde{\rho}(\widetilde{y})\ d\widetilde{y}, \qquad \widetilde{x} \in [0, 1], \tag{3.21}$$

with $\widetilde{\rho}(\widetilde{x}) = \rho(x),\ x = z_0 + (z_1 - z_0)\widetilde{x}$ and $\widetilde{K}(\widetilde{z}) = K(z),\ z = (z_1 - z_0)\widetilde{z}$.

Here we choose a set of parameters: $\lambda_1 = 5.32, \lambda_2 = 304, z_1 = 0, z_2 = 100,\ \ r_d = 0.1$ [22]. The SOE approximation parameters are set $\varepsilon = 10^{-10}, \delta = 10^{-6}$, and the number of exponentials is 174. In the local correction integral (2.10), the kernel can not be integrated simply like in the last two examples. Therefore, we use the first order Taylor series approximation for the kernel, i.e., setting $m = 1$ in (2.11), and the remainder is $C\delta^2\|\rho\|_{L^1} \approx 10^{-12}$, which is practically small enough to guarantee the second order convergence $O(h^2)$. Table 6 displays the successive errors E_h and the convergence rate for a Gaussian density $\rho(x) = e^{-4(x-1/2)^2}$, from which one can observe a second order convergence clearly.

Table 6: Second order convergence for photoionization kernel with Gaussian density $\rho(x) = e^{-4(x-\frac{1}{2})^2}$ and $h_0 = 0.005$.

h/h_0	1	1/2	1/4	1/8	1/16	1/32
E_h	2.826E-1	7.076E-2	1.770E-2	4.427E-3	1.107E-3	2.767E-4
Rate	1.998	1.999	1.999	1.999	2.000	-

4. Conclusions

We proposed an accurate and fast algorithm for the one-dimensional spatial convolution based on sum-of-exponential (SOE) approximation of the kernel. The convolution evaluation boils down to the exponential convolution $\phi_{\rm ES} = K_{\rm ES} * \rho$ and the local correction integral $\phi_{\rm cor}(x) = \int_{[0,1]\cap[x-\delta,x+\delta]}(K - K_{\rm ES})(x-y)\rho(y)dy$, and both are computed in $O(M + N)$ operations on discrete uniform/nonuniform grids. The exponential convolution is calculated via standard recurrence formulas. The accurate local correction calculation is done by approximating the density with its low order Taylor series in a small neighborhood of length δ and integrating the kernel's singularity with polynomials as accurately as possible. Our algorithm is ideal for parallelization and favors easy extensions to complicated kernel with SOE. Extensive numerical results for different kernels have shown its efficiency, accuracy and easy extension, which implies its possible application in real simulations.

References

[1] M. Abbamowitz and I. Stegun, Handbook of Mathematical Functions, Dover, 1965.

[2] B. Alpert, L. Greengard and T. Hagstrom, Rapid evaluation of nonreflecting boundary kernels for time-domain wave propagation, SIAM J. Numer. Anal. 37 (2000) 1138-1164.

[3] G. Beylkin, C. Kurzc and L. Monzón, Fast convolution with the free space Helmholtz Green's function, J. Comput. Phys. 228 (2009) 2770-2791.

[4] G. Beylkin and L. Monzón, Approximation by exponential sums, Appl. Comput. Harmon. Anal. 19 (2005) 17-48.

[5] G. Beylkin and L. Monzón, Approximation by exponential sums revisited, Appl. Comput. Harmon. Anal. 28 (2010) 131-149.

[6] K. Glover, All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds, Int. J. Control 39 (1984) 1115-1193.

[7] 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 (2006) 177–202.

[8] S. Jiang, L. Greengard and Y. Zhang, The anisotropic truncated kernel method for convolution with free-space Green's functions, SIAM J. Sci. Comput. 40(2018) A3733-A3754.

[9] S. Jiang, J. Zhang, Q. Zhang and Z. Zhang, Fast Evaluation of the Caputo Fractional Derivative and its Applications to Fractional Diffusion Equations, Commun. Comput. Phys. 21 (2017) 650-678.

[10] Q. Jiang, A high-order fast method for computing convolution integral with smooth kernel, Comput. Phys. Commun. 181 (2010) 313-316.

[11] S. Jiang and Y. Zhang, A black-box algorithm for efficient and accurate sum-of-exponential approximation of certain class of functions on an interval, in preparation.

[12] B. Lin, C. Zhuang, Z. Cai, R. Zeng and W. Bao, An efficient and accurate MPI-based parallel simulator for streamer discharges in three dimensions, J. Comput. Phys. 401 (2020) 109026.

[13] C. Lubich and A. Schädle, Fast convolution for nonreflecting boundary conditions, SIAM J. Sci. Comput. 24 (2002) 161-182.

[14] F. Vico, L. Greengard and M. Ferrando, Fast convolution with free-space Green's functions, J. Comput. Phys.,323 (2016) 191-203.

[15] K. Xu and S. Jiang, A Bootstrap method for Sum-of-Poles approximations, J. Sci. Comput. 55 (2013) 16-39.

[16] N. Yarvin and V. Rokhlin, An improved fast multipole algorithm for potential fields on the line, SIAM J. Numer. Anal. 36 (1999) 629-666.

[17] Z. Yu, Q. Li, R. Zeng, J. He, Y. Zhang, Z. Li, C. Zhuang and Y. Liao, Calculation of surface electric field on UHV transmission lines under lightning stroke, Electric power systems research 94 (2013) 79-85.

[18] Q. Zhang, J. Zhang, S. Jiang and Z. Zhang, Numerical solution to a linearized time fractional KdV equation on unbounded domains, Math. Comp. 87 (2018) 693-719.

[19] Y. Zhang, Y. Zhao and Z. Zhou, A unified structure preserving scheme for a multi-species model with a gradient flow structure and nonlocal interactions via singular kernels, arXiv:2006.16526.

[20] X. Zhou, R. Zeng, C. Zhuang, Y. Zhang and L. Liu, A fast tree algorithm for electrical field calculation in electrical discharge simulations. IEEE Transactions on Magnetics 54 (2018) 7201034.

[21] C. Zhuang and R. Zeng, A local discontinuous Galerkin method for 1.5-dimensional streamer discharge simulations, Appl. Math. Comput. 219 (2013) 9925-9934.

[22] C. Zhuang, R. Zeng, B. Zhang and J. He, A WENO scheme for simulating streamer discharge with photoionizations, IEEE Transactions on Magnetics 50 (2014) 7007904.