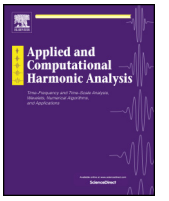
Appl. Comput. Harmon. Anal. 49 (2020) 316–327

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha



Letter to the Editor

Error analysis of an accelerated interpolative decomposition for 3D Laplace problems ☆

Xin Xing [a], Edmond Chow [b,*]

[a] *School of Mathematics, Georgia Institute of Technology, Atlanta, GA, United States of America*
[b] *School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America*

ARTICLE INFO

Article history:
Received 29 June 2018
Received in revised form 16 January 2019
Accepted 8 November 2019
Available online 14 November 2019
Communicated by Vladimir Rokhlin

MSC:
65G99

Keywords:
Interpolative decomposition
Proxy surface
Laplace kernel

ABSTRACT

In fast direct solvers for integral equations with the Laplace kernel, a hierarchical compression process needs to compute interpolative decompositions of off-diagonal block rows of the discretized integral operators. This computation can be dramatically accelerated by a technique called the proxy surface method, which is motivated by potential theory. We present a long overdue, rigorous error analysis of this acceleration technique. The analysis provides theoretical guidance for the discretization of the proxy surface used in the technique.

1. Introduction

Integral equations are of great importance in engineering and physics. One of the challenges in their numerical solution is that the discretization of integral operators yields dense matrices. For integral equations with non-oscillatory kernels from potential theory, e.g., the Laplace and Stokes kernels, fast direct solvers [1–5] have been developed based on the "data-sparse" structure of the discretized integral operators. These solvers first compute an approximant of the discretized integral operator by a hierarchical compression process, sometimes called "recursive skeletonization" [4]. This approximant is in a data-sparse format such as hierarchically semi-separable (HSS) format [6,7] so that its exact inverse can be efficiently calculated, stored, and applied.

The main bottleneck of the hierarchical compression step in these solvers is the low-rank approximation of off-diagonal block rows (and columns) of the discretized integral operator. Each off-diagonal block row

☆ Supported by National Science Foundation under grant ACI-1609842.
* Corresponding author.
E-mail addresses: xxing33@gatech.edu (X. Xing), echow@cc.gatech.edu (E. Chow).

https://doi.org/10.1016/j.acha.2019.11.003
1063-5203/

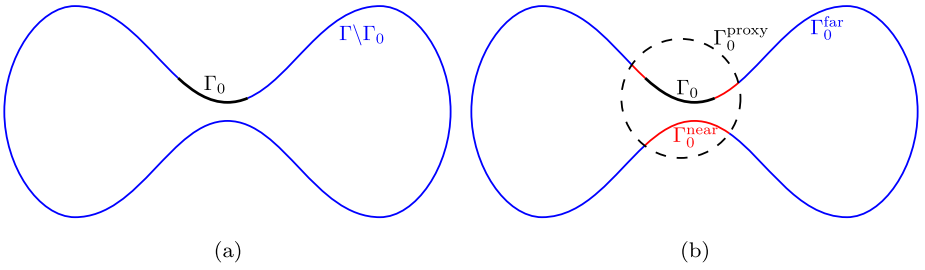


Fig. 1. Illustration of the proxy surface method for an integral equation on a contour Γ.

corresponds to the interaction between a subdomain $\Gamma_0 \subset \Gamma$ and its complement $\Gamma\backslash\Gamma_0$, where Γ denotes the domain of the integral equation. Such a block is denoted as $A_{\Gamma_0,\Gamma\backslash\Gamma_0}$ in the following discussion. Purely algebraic approaches to obtaining low-rank approximations of these blocks can lead to a prohibitive quadratic cost for the hierarchical compression. Martinsson and Rokhlin [1] introduced an acceleration technique for this problem that dramatically reduces the cost of the hierarchical compression. Similar ideas have also been used in kernel-independent FMM [8,9] and the proxy point method [10] for general kernel functions that are not from potential theory.

Fig. 1 illustrates the acceleration technique with a diagram that has been commonly used in the literature [1–3]. The acceleration technique first partitions $A_{\Gamma_0,\Gamma\backslash\Gamma_0}$ into near field interactions $A_{\Gamma_0,\Gamma_0^{\text{near}}}$ and far field interactions $A_{\Gamma_0,\Gamma_0^{\text{far}}}$ by an artificial surface Γ_0^{proxy} that encloses Γ_0. Motivated by potential theory, the technique then seeks the low-rank approximation of $A_{\Gamma_0,\Gamma_0^{\text{near}}\cup\Gamma_0^{\text{proxy}}}$ which is usually a much smaller matrix than $A_{\Gamma_0,\Gamma\backslash\Gamma_0}$. This low-rank approximation serves as an intermediate step to approximating $A_{\Gamma_0,\Gamma\backslash\Gamma_0}$ itself. The enclosing surface Γ_0^{proxy} is called a proxy surface in [4] and thus we refer to this acceleration technique as the *proxy surface method*.

The error analysis of the proxy surface method, however, is only sketched in [1] and the discretization of the proxy surface Γ_0^{proxy} is chosen heuristically in previous work [1–5]. In this paper, we provide a rigorous error analysis of the proxy surface method with the 3D Laplace kernel. The error analysis shows how the number of points needed to discretize Γ_0^{proxy} depends on the desired accuracy of the low-rank approximation and on the ratio of the radius of Γ_0^{proxy} to the radius of the subdomain Γ_0.

2. Background

Interpolative decomposition (ID, a.k.a. skeletonization) [11,12] represents or approximates a matrix $H \in \mathbb{R}^{n\times m}$ in the low-rank form UH_J, where $U \in \mathbb{R}^{n\times k}$ has bounded entries, $H_J \in \mathbb{R}^{k\times m}$ contains k rows of H, and k is the rank. The matrix H_J is called the *row skeleton*. An ID approximation defined this way is said to have precision ε_0 if the norm of each row of the error matrix $H - UH_J$ is bounded by ε_0. Using an algebraic approach, an ID approximation with a given rank or a given precision threshold can be calculated using the strong rank-revealing QR (SRRQR) decomposition [12]. The matrix U obtained from this approach can have all its entries bounded by a prespecified parameter $C_{\text{qr}} \geqslant 1$.

Consider an integral equation with the 3D Laplace kernel $K(x, y) = 1/|x - y|$

$$a(x)u(x) + \int_\Gamma K(x,y)u(y)\mathrm{d}y = f(x), \quad x \in \Gamma \subset \mathbb{R}^3, \tag{1}$$

where $a(x)$ and $f(x)$ are given functions and $u(x)$ is the unknown function to be determined. The Nyström method [13,3] for discretizing the integral equation gives

$$a(x_i)u(x_i) + \sum_{x_j\in X} K(x_i,x_j)w_j u(x_j) = f(x_i), \quad x_i \in X, \tag{2}$$

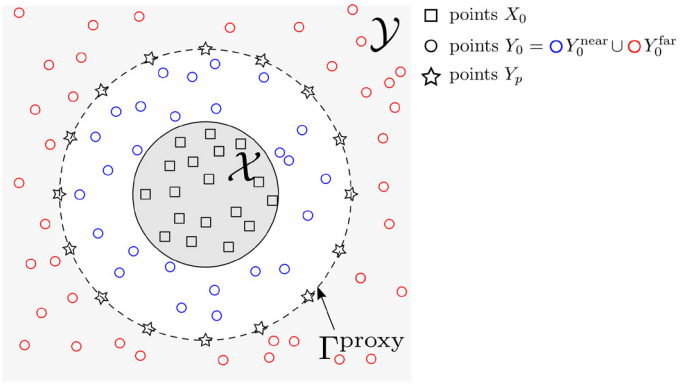


Fig. 2. Relationship between X_0, Y_0, and a proxy surface Γ^{proxy} for the submatrix $K(X_0, Y_0)$ to be approximated by the proxy surface method.

where X is the set of points used to discretize Γ and w_j is the quadrature weight at x_j. Any off-diagonal block row $A_{\Gamma_0,\Gamma/\Gamma_0}$ of the discretized integral operator in (2) has its entries defined as $A_{\Gamma_0,\Gamma/\Gamma_0}(i,j) = K(x_i, y_j)w_j$ where $X_0 = \{x_i\}$ and $Y_0 = \{y_j\}$ are subsets of X and are used to discretize Γ_0 and Γ/Γ_0, respectively.

Ignoring the quadrature weights $\{w_j\}$ in $A_{\Gamma_0,\Gamma/\Gamma_0}$, we analyze the proxy surface method for approximating the submatrix $K(X_0, Y_0) = (K(x_i, y_j))_{x_i \in X_0, y_j \in Y_0}$. Our analysis can be naturally extended to the case with quadrature weights and to the case with other discretization methods. As illustrated in Fig. 2, we assume that the points in X_0 lie in a bounded domain which we call $\mathcal{X}$, and that the points in Y_0 lie outside $\mathcal{X}$. An artificial proxy surface Γ^{proxy} encloses $\mathcal{X}$ and is discretized by a set of points Y_p. The domain outside Γ^{proxy} is called the *far field* of $\mathcal{X}$ and denoted as $\mathcal{Y}$. The set of points Y_0 is partitioned into Y_0^{near} and Y_0^{far} which are inside and outside Γ^{proxy}, respectively.

The proxy surface method calculates a low-rank approximation of the submatrix $K(X_0, Y_0)$ in the form of an ID,

$$K(X_0, Y_0) \approx UK(X_{\mathcal{S}}, Y_0) \tag{3}$$

where $X_{\mathcal{S}} \subset X_0$ is the subset of points associated with the row skeleton of the ID approximation. The proxy surface method accelerates the purely algebraic computation (based on SRRQR alone) of such an ID approximation, as shown in Algorithm 1. In practice, the set of points Y_p used to discretize Γ^{proxy} in Algorithm 1 is usually much smaller than Y_0^{far} and the purely algebraic approximation of $K(X_0, Y_0^{\text{near}} \cup Y_p)$ is much cheaper than that of $K(X_0, Y_0)$. Thus, the main idea of the method is to replace the columns $K(X_0, Y_0^{\text{far}})$ in the matrix $K(X_0, Y_0)$ by the smaller matrix $K(X_0, Y_p)$ while keeping $K(X_0, Y_0^{\text{near}})$ unchanged.

Algorithm 1 Proxy surface method.

Input: X_0, Y_0, Γ^{proxy}.
Output: U and $X_{\mathcal{S}}$ for an ID approximation $UK(X_{\mathcal{S}}, Y_0)$ of $K(X_0, Y_0)$.
Step 1: partition Y_0 into Y_0^{near} and Y_0^{far} which are inside and outside Γ^{proxy}, respectively, as shown in Fig. 2.
Step 2: select a set of points Y_p for discretizing Γ^{proxy}.
Step 3: calculate U and $X_{\mathcal{S}}$ from an ID approximation of $K(X_0, Y_0^{\text{near}} \cup Y_p)$ using SRRQR as

$$K(X_0, Y_0^{\text{near}} \cup Y_p) \approx UK(X_{\mathcal{S}}, Y_0^{\text{near}} \cup Y_p). \tag{4}$$

To simplify analysis, we first assume that the points in Y_0 only lie outside Γ^{proxy}, i.e., Y_0^{near} is empty, and thus (4) in Algorithm 1 is replaced by

$$K(X_0, Y_p) \approx UK(X_{\mathcal{S}}, Y_p). \tag{5}$$

In this paper, we focus on the error analysis of this reduced case. The general case, removing the assumption that Y_0^{near} is empty, is handled in Section 5 using the analysis of this reduced case.

The effectiveness of the proxy surface method in the reduced case is explained as follows, which has been presented previously [1,2,4], although in a different form.

Recall that $\mathcal{Y}$ is the domain outside Γ^{proxy} as shown in Fig. 2. For any fixed $x \in \mathcal{X}$, the function $K(x, y)$ is harmonic in the single variable $y \in \mathcal{Y}$ and thus $K(x, y)$ is the solution to the exterior Dirichlet problem

$$\begin{aligned} -\Delta u(y) &= 0, && y \in \mathcal{Y}, \\ u(y) &= K(x, y), && y \in \Gamma^{\text{proxy}} = \partial\mathcal{Y}. \end{aligned}$$

Using the Green's function $G(x, y)$ for the above Laplace boundary value problem, $K(x, y)$ can be represented using a double layer potential as

$$K(x, y) = -\int_{\Gamma^{\text{proxy}}} K(x, z)\frac{\partial G}{\partial \nu}(z, y)\mathrm{d}S(z), \quad y \in \mathcal{Y},$$

where ν denotes the outer normal direction of Γ^{proxy} at point z. This representation can be further discretized through numerical quadrature as

$$K(x, y) \approx K(x, Y_p)W(Y_p, y), \quad x \in \mathcal{X},\ y \in \mathcal{Y} \tag{6}$$

where Y_p here is a set of quadrature points on Γ^{proxy} and $W(Y_p, y)$ is $-\frac{\partial G}{\partial \nu}(Y_p, y)$ times corresponding quadrature weights. Substituting X_0 and Y_0 into the above discretized representation gives the approximation

$$K(X_0, Y_0) \approx K(X_0, Y_p)W(Y_p, Y_0). \tag{7}$$

Based on (7), the ID approximation error of $K(X_0, Y_0)$ in (3) can be bounded by that of $K(X_0, Y_p)$ in (5) as

$$\begin{aligned} \|K(X_0, Y_0) - UK(X_{\mathcal{S}}, Y_0)\|_F &\approx \|\left(K(X_0, Y_p) - UK(X_{\mathcal{S}}, Y_p)\right)W(Y_p, Y_0)\|_F \\ &\leqslant \|K(X_0, Y_p) - UK(X_{\mathcal{S}}, Y_p)\|_F \|W(Y_p, Y_0)\|_2. \end{aligned} \tag{8}$$

Note that the above explanation is only qualitative and not enough for an error analysis of the proxy surface method since the discretization of the integral representation in (6) and the boundedness of $\|W(Y_p, Y_0)\|_2$ in (8) have not been rigorously studied.

Furthermore, the number of points in Y_p used to discretize Γ^{proxy} is chosen heuristically in practice. Refs. [1,2] suggest using $|Y_p| \sim O(|X_0|)$. Ref. [4] claims correctly but without an explanation that, for the Laplace kernel with a fixed ratio of the radius of Γ^{proxy} to the radius of $\mathcal{X}$, the number of points needed to discretize Γ^{proxy} depends on the desired precision of the low-rank approximation.

In this paper, to theoretically justify the proxy surface method, we address the following two problems: (a) the quantitative relationship between the ID approximation error of $K(X_0, Y_p)$ in (5) and that of $K(X_0, Y_0)$ in (3); (b) how to choose the number of points in Y_p to guarantee a given precision of the ID approximation of $K(X_0, Y_0)$ in (3).

3. Main result

Denote the open ball of radius r centered at the origin in 3D as $B(0, r)$. Consider $\mathcal{X} = B(0, r_1)$, $\Gamma^{\text{proxy}} = \partial B(0, r_2)$, and $\mathcal{Y} = \mathbb{R}^3 \backslash B(0, r_2)$ with $r_2 > r_1$ as illustrated in Fig. 3. For the 3D Laplace kernel $K(x, y) =$

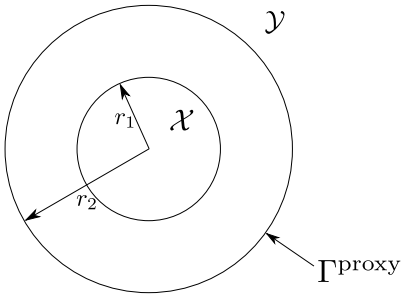


Fig. 3. 2D illustration of the 3D domains $\mathcal{X}$ and $\mathcal{Y}$ and the proxy surface Γ^{proxy}.

$1/|x-y|$, this section provides an error analysis for the proxy surface method in the reduced case that Y_0^{near} is empty and the points in Y_0 are only in $\mathcal{Y}$. The error analysis for the general case that Y_0^{near} is nonempty is discussed in Section 5 based on the results in this section.

Assume that a set of points Y_p is used to discretize Γ^{proxy} and that the submatrix $K(X_0, Y_0)$ to be approximated is associated with point sets $X_0 \subset \mathcal{X}$ and $Y_0 \subset \mathcal{Y}$. In the proxy surface method, the obtained ID approximation $K(X_0, Y_0) \approx UK(X_{\mathcal{S}}, Y_0)$ can be viewed row-by-row as

$$K(x_i, Y_0) \approx u_i^T K(X_{\mathcal{S}}, Y_0), \quad x_i \in X_0, \tag{9}$$

where u_i^T denotes the ith row of U. Since U and $X_{\mathcal{S}}$ are computed from the ID approximation $K(X_0, Y_p) \approx UK(X_{\mathcal{S}}, Y_p)$ and are not related to Y_0, the approximation in (9) can be applied to any set of points Y_0 in $\mathcal{Y}$. Thus, the error of the approximation in (9) depends on the quality of the function approximation

$$K(x_i, y) \approx u_i^T K(X_{\mathcal{S}}, y), \quad x_i \in X_0,\ y \in \mathcal{Y}.$$

Denote the error of this function approximation as

$$e_i(y) = K(x_i, y) - u_i^T K(X_{\mathcal{S}}, y), \quad x_i \in X_0,\ y \in \mathcal{Y}. \tag{10}$$

Note that the ith row of the error matrix $K(X_0, Y_0) - UK(X_{\mathcal{S}}, Y_0)$ from the ID approximation of $K(X_0, Y_0)$ in (3) is exactly $e_i(Y_0)$. Similarly, the ith row of the error in the ID approximation of $K(X_0, Y_p)$ in (5) is $e_i(Y_p)$. Later in this paper (in Theorem 1 at the end of this section), we will assume that the ID approximation of $K(X_0, Y_p)$ in (5), calculated using SRRQR, has precision $\varepsilon\sqrt{|Y_p|}$ and thus $\|e_i(Y_p)\|_2 \leqslant \varepsilon\sqrt{|Y_p|}$ for any $x_i \in X_0$.

For an arbitrary set of points Y_0 in $\mathcal{Y}$, the best upper bound for $e_i(Y_0)$ is

$$\|e_i(Y_0)\|_2 \leqslant \sqrt{|Y_0|} \max_{y \in \mathcal{Y}} |e_i(y)|, \tag{11}$$

where equality holds when $|e_i(y)|$ reaches the same maximum in $\mathcal{Y}$ for all points in Y_0. Our error analysis of the proxy surface method seeks an upper bound for $|e_i(y)|$ in the whole domain $\mathcal{Y}$ in terms of $\|e_i(Y_p)\|_2$ which itself can be bounded by specifying the precision of the SRRQR. The bulk of the analysis rests on the following proposition.

Proposition 1. *If the set of points $Y_p \subset \Gamma^{proxy} = \partial B(0, r_2)$ satisfies the condition that numerical quadrature with the points in Y_p and equal weights $\frac{4\pi r_2^2}{|Y_p|}$ is exact for polynomials on Γ^{proxy} of degree up to $2c$, then $e_i(y)$ defined in (10) for any $x_i \in X_0$ can be bounded as*

$$|e_i(y)| \leqslant (c+1)\frac{\|e_i(Y_p)\|_2}{\sqrt{|Y_p|}} + (c+2)\frac{(1 + |X_{\mathcal{S}}|\|u_i\|_\infty)}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1},\ y \in \mathcal{Y}. \tag{12}$$

Proof. For any $x \in \mathcal{X}$, the function $K(x,y)$ is harmonic in the single variable $y \in \mathcal{Y}$. Since $e_i(y)$ is a linear combination of $K(x_i,y)$ and $\{K(x_j,y) : x_j \in X_{\mathcal{S}}\}$, it is also harmonic in $\mathcal{Y}$. By the maximum principle of harmonic functions, $e_i(y)$ satisfies

$$\max_{y\in\mathcal{Y}} |e_i(y)| = \max_{y\in\Gamma^{\text{proxy}}} |e_i(y)|. \tag{13}$$

Thus, it suffices to prove the upper bound (12) for $y \in \Gamma^{\text{proxy}}$.

The multipole expansion of $K(x,y)$ with $(x,y) \in \mathcal{X} \times \Gamma^{\text{proxy}}$ is written as

$$K(x,y) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} M_l^m(x) \frac{1}{r_2^{l+1}} Y_l^m(\alpha,\beta), \tag{14}$$

where (r_2,α,β) denotes the polar coordinates of y on Γ^{proxy}, $Y_l^m(\alpha,\beta)$ is the spherical harmonic function of degree l and order m, and $\{M_l^m(x)\}$ is a set of known moment functions. Truncating the above infinite sum at index c, the remainder can be bounded as

$$\left| K(x,y) - \sum_{l=0}^{c} \sum_{m=-l}^{l} M_l^m(x) \frac{1}{r_2^{l+1}} Y_l^m(\alpha,\beta) \right| \leqslant \frac{1}{r_2 - r_1} \left(\frac{r_1}{r_2}\right)^{c+1}.$$

Substituting the multipole expansion (14) into (10), the multipole expansion for $e_i(y)$ with $y \in \Gamma^{\text{proxy}}$ can be written as

$$\begin{aligned} e_i(y) &= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(M_l^m(x_i) - u_i^T M_l^m(X_{\mathcal{S}}) \right) \frac{1}{r_2^{l+1}} Y_l^m(\alpha,\beta) \\ &= \sum_{l=0}^{c} \sum_{m=-l}^{l} E_l^m Y_l^m(\alpha,\beta) + R_c(y), \end{aligned} \tag{15}$$

where E_l^m represents the coefficients collected for $Y_l^m(\alpha,\beta)$ and the remainder $R_c(y)$ can be bounded as

$$|R_c(y)| \leqslant \frac{(1 + |X_{\mathcal{S}}| \|u_i\|_\infty)}{r_2 - r_1} \left(\frac{r_1}{r_2}\right)^{c+1} \tag{16}$$

using the triangle inequality. Since $\{Y_l^m(\alpha,\beta)\}$ is a set of orthonormal polynomial functions on the unit sphere $\mathbb{S}^2$, the coefficients E_l^m in (15) can be calculated analytically as

$$\begin{aligned} E_l^m &= \int_{\mathbb{S}^2} \left(e_i(r_2 y) - R_c(r_2 y)\right) Y_l^m(y) \mathrm{d}S(y) \\ &= \frac{1}{r_2^2} \int_{\Gamma^{\text{proxy}}} \left(e_i(y) - R_c(y)\right) Y_l^m(y) \mathrm{d}S(y), \end{aligned}$$

where $Y_l^m(y)$ is defined as $Y_l^m(\alpha,\beta)$ for any $y = (|y|,\alpha,\beta)$.

Note that $(e_i(y) - R_c(y))Y_l^m(y)$ is a polynomial on Γ^{proxy} of degree at most $c + l$. Since numerical quadrature with the points in Y_p and equal weights $\frac{4\pi r_2^2}{|Y_p|}$ is exact for polynomials on Γ^{proxy} of degree up to $2c$, E_l^m with $l \leqslant c$ can be further represented as

$$\begin{aligned} E_l^m =& \frac{1}{r_2^2} \int_{\Gamma^{\text{proxy}}} (e_i(y) - R_c(y)) Y_l^m(y) \mathrm{d}S(y) \\ =& \frac{1}{r_2^2} \sum_{y_j \in Y_p} (e_i(y_j) - R_c(y_j)) Y_l^m(y_j) \frac{4\pi r_2^2}{|Y_p|} \\ =& \frac{4\pi}{|Y_p|} \left(e_i(Y_p) - R_c(Y_p)\right)^T Y_l^m(Y_p). \end{aligned} \tag{17}$$

Substituting (17) into (15), $e_i(y)$ with $y \in \Gamma^{\text{proxy}}$ can be rewritten as

$$\begin{aligned} e_i(y) =& \frac{4\pi}{|Y_p|} \left(e_i(Y_p) - R_c(Y_p)\right)^T \left(Y_0^0(Y_p)\ Y_1^{-1}(Y_p)\ \dots\ Y_c^c(Y_p)\right) \begin{pmatrix} Y_0^0(y) \\ Y_1^{-1}(y) \\ \vdots \\ Y_c^c(y) \end{pmatrix} + R_c(y) \\ =& \frac{4\pi}{|Y_p|} \left(e_i(Y_p) - R_c(Y_p)\right)^T M\Phi(y) + R_c(y), \end{aligned} \tag{18}$$

where M and $\Phi(y)$ denote, respectively, the middle matrix and the last vector function of y in the first equation. Note that any two distinct columns of M, say $Y_{l_1}^{m_1}(Y_p)$ and $Y_{l_2}^{m_2}(Y_p)$, are orthogonal with each other, i.e.,

$$Y_{l_1}^{m_1}(Y_p)^T Y_{l_2}^{m_2}(Y_p) = \frac{|Y_p|}{4\pi} \int_{\mathbb{S}^2} Y_{l_1}^{m_1}(y) Y_{l_2}^{m_2}(y) \mathrm{d}S(y) = \frac{|Y_p|}{4\pi} \delta_{l_1,l_2} \delta_{m_1,m_2},$$

and thus the scaled matrix $\sqrt{\frac{4\pi}{|Y_p|}} M$ has orthonormal columns. Therefore, it holds that

$$\sqrt{\frac{4\pi}{|Y_p|}} \left\| M\Phi(y) \right\|_2 = \left\| \Phi(y) \right\|_2. \tag{19}$$

Meanwhile, by the addition theorem of spherical harmonics, the norm of the vector function $\Phi(y)$ at any $y \in \Gamma^{\text{proxy}}$ is

$$\left\| \Phi(y) \right\|_2 = \sqrt{\sum_{l=0}^{c} \left(\sum_{m=-l}^{l} |Y_l^m(y)|^2 \right)} = \sqrt{\sum_{l=0}^{c} \frac{2l+1}{4\pi}} = \frac{c+1}{\sqrt{4\pi}}. \tag{20}$$

Combining (16), (19), and (20) into (18), we obtain an upper bound on $e_i(y)$ as

$$\begin{aligned} |e_i(y)| \leqslant& \left| \frac{4\pi}{|Y_p|} e_i(Y_p)^T M\Phi(y) \right| + \left| \frac{4\pi}{|Y_p|} R_c(Y_p)^T M\Phi(y) \right| + |R_c(y)| \\ \leqslant& \frac{4\pi}{|Y_p|} \|e_i(Y_p)\|_2 \|M\Phi(y)\|_2 + \frac{4\pi}{|Y_p|} \|R_c(Y_p)\|_2 \|M\Phi(y)\|_2 + |R_c(y)| \\ \leqslant& (c+1) \frac{\|e_i(Y_p)\|_2}{\sqrt{|Y_p|}} + (c+1) \frac{\|R_c(Y_p)\|_2}{\sqrt{|Y_p|}} + |R_c(y)| \\ \leqslant& (c+1) \frac{\|e_i(Y_p)\|_2}{\sqrt{|Y_p|}} + (c+2) \frac{(1 + |X_{\mathcal{S}}| \|u_i\|_\infty)}{r_2 - r_1} \left(\frac{r_1}{r_2} \right)^{c+1}, \end{aligned} \tag{21}$$

using the Cauchy-Schwarz and triangle inequalities. □

Combining Proposition 1 and the inequality (11), the error bound of the proxy surface method for the ID approximation of $K(X_0, Y_0)$ in the reduced case can be stated as follows.

Theorem 1 *(Error bound for the proxy surface method). If the set of points Y_p satisfies the condition in Proposition 1 and the ID approximation of $K(X_0, Y_p)$ in (5) has precision $\varepsilon\sqrt{|Y_p|}$, i.e., $\|e_i(Y_p)\|_2 \leqslant \varepsilon\sqrt{|Y_p|}$ for each $x_i \in X_0$, the ID approximation of $K(X_0, Y_0)$ in (3), calculated by the proxy surface method, has error $e_i(Y_0)$ in the ith row bounded as*

$$\frac{\|e_i(Y_0)\|_2}{\sqrt{|Y_0|}} \leqslant (c+1)\frac{\|e_i(Y_p)\|_2}{\sqrt{|Y_p|}} + (c+2)\frac{(1+|X_{\mathcal{S}}|\|u_i\|_\infty)}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1} \tag{22}$$

$$\leqslant (c+1)\varepsilon + (c+2)\frac{(1+|X_{\mathcal{S}}|\|u_i\|_\infty)}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1}. \tag{23}$$

When there are not many points in Y_p, i.e., c is small, the error bound (22) is dominated by its second term which comes from the truncation error $R_c(y)$ of the multipole expansion of $e_i(y)$ in (15). On the other hand, observe that the second term in (22) decays exponentially with c. Thus, with a sufficiently large number of points in Y_p, the root-mean-square error $\|e_i(Y_0)\|_2/\sqrt{|Y_0|}$ of the ID approximation of $K(X_0, Y_0)$ is controlled by the first term in (22) which is proportional to the root-mean-square error $\|e_i(Y_p)\|_2/\sqrt{|Y_p|}$ of the ID approximation of $K(X_0, Y_p)$.

4. Selection of Y_p

Using the quadrature point sets provided in Ref. [14], $2c^2+2c+O(1)$ points are needed in Y_p to construct an exact quadrature for polynomials on Γ^{proxy} of degree up to $2c$. Thus, the main question in the selection of Y_p is choosing the smallest constant c that balances the precision and efficiency of the proxy surface method.

Since the error bound (23) contains $|X_{\mathcal{S}}|$ and $\|u_i\|_\infty$ which depend on the ID approximation of $K(X_0, Y_p)$, a priori estimates of these two quantities are needed for the selection of Y_p. When using SRRQR to calculate the ID approximation of $K(X_0, Y_p)$, entries of the obtained U can be bounded by a prespecified parameter $C_{\text{qr}} \geqslant 1$ and thus $\|u_i\|_\infty \leqslant C_{\text{qr}}$ for any $x_i \in X_0$. The size of $X_{\mathcal{S}}$ is an estimate of the numerical rank of $K(X_0, Y_p)$ and thus satisfies $|X_{\mathcal{S}}| \leqslant \min(|X_0|, |Y_p|)$. Plugging these estimates into (23), we obtain an a priori error bound as

$$\frac{\|e_i(Y_0)\|_2}{\sqrt{|Y_0|}} \leqslant (c+1)\varepsilon + (c+2)\frac{C_{\text{qr}}\min(|X_0|, |Y_p|)+1}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1}. \tag{24}$$

In the upper bound (24), the second term decays exponentially in c while the first term increases linearly in c. Therefore, choosing the integer c that makes the second term approximately the same as the first term, i.e.,

$$\frac{C_{\text{qr}}\min\left(|X_0|, 2c^2+2c+O(1)\right)+1}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1} \approx \varepsilon \tag{25}$$

can approximately minimize the upper bound in (24) (as a function of c). Then, the set of points Y_p from the dataset of Ref. [14] corresponding to the selected c can be used for the discretization of Γ^{proxy}. With such a selection of Y_p, the error in the ith row of the ID approximation of $K(X_0, Y_0)$ is bounded as

$$\|e_i(Y_0)\|_2 \leqslant (2c+3)\varepsilon\sqrt{|Y_0|}.$$

The condition required for Y_p in Proposition 1 is needed for a rigorous error analysis. Also, the a priori upper bound in (24) may not be tight. Thus, the above choice for the number of points in Y_p is conservative. However, the key idea conveyed by Theorem 1 and the above selection of Y_p is that, as long as the absolute distance between Γ^{proxy} and $\mathcal{X}$ is not too small, e.g., $r_2 - r_1 \geqslant 1$, the number of points needed in Y_p to guarantee the root-mean-square approximation precision of the proxy surface method only depends on ε in the precision threshold $\varepsilon\sqrt{|Y_p|}$ for the ID approximation of $K(X_0, Y_p)$ and on the ratio of the radius of Γ^{proxy} to the radius of $\mathcal{X}$.

5. Error analysis for the general case

For the general case, Y_0 is partitioned into Y_0^{near} and Y_0^{far} where the points in Y_0^{near} lie between $\mathcal{X}$ and Γ^{proxy} and the points in Y_0^{far} lie in $\mathcal{Y}$. In Algorithm 1, assume that the ID approximation $K(X_0, Y_0^{\text{near}} \cup Y_p) \approx UK(X_{\mathcal{S}}, Y_0^{\text{near}} \cup Y_p)$ calculated by the proxy surface method using SRRQR has precision ε_0. The error function $e_i(y)$ defined in (10) for any $x_i \in X_0$ with the resulting U and $X_{\mathcal{S}}$ satisfies $\|e_i(Y_0^{\text{near}} \cup Y_p)\|_2 \leqslant \varepsilon_0$. Furthermore, the ith row error vector of the defined ID approximation $K(X_0, Y_0) \approx UK(X_{\mathcal{S}}, Y_0)$ can be written as $e_i(Y_0) = e_i(Y_0^{\text{near}} \cup Y_0^{\text{far}})$.

Since the points in Y_0^{far} lie in $\mathcal{Y}$, the error bound (22) in Theorem 1 can be applied to $e_i(Y_p)$ and $e_i(Y_0^{\text{far}})$ and be written as

$$\frac{\|e_i(Y_0^{\text{far}})\|_2}{\sqrt{|Y_0^{\text{far}}|}} \leqslant (c+1)\frac{\|e_i(Y_p)\|_2}{\sqrt{|Y_p|}} + (c+2)\frac{(1+|X_{\mathcal{S}}|\|u_i\|_\infty)}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1}. \tag{26}$$

Based on this inequality and the two loose inequalities $\|e_i(Y_0^{\text{near}})\|_2 \leqslant \varepsilon_0$ and $\|e_i(Y_p)\|_2 \leqslant \varepsilon_0$ derived from $\|e_i(Y_0^{\text{near}} \cup Y_p)\|_2 \leqslant \varepsilon_0$, the row error vector $e_i(Y_0)$ can be bounded as

$$\begin{aligned}
\|e_i(Y_0)\|_2^2 &= \|e_i(Y_0^{\text{near}})\|_2^2 + \|e_i(Y_0^{\text{far}})\|_2^2 \\
&\leqslant \varepsilon_0^2 + |Y_0^{\text{far}}|\left((c+1)\frac{\varepsilon_0}{\sqrt{|Y_p|}} + (c+2)\frac{(1+|X_{\mathcal{S}}|\|u_i\|_\infty)}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1}\right)^2 \\
&\lesssim \left(1 + (c+1)^2\frac{|Y_0^{\text{far}}|}{|Y_p|}\right)\varepsilon_0^2.
\end{aligned} \tag{27}$$

The final inequality above assumes that the constant c is large enough so that the last term in (26) is negligible compared to the first term.

The error bound (27) is loose due to the use of $\|e_i(Y_0^{\text{near}})\|_2 \leqslant \varepsilon_0$ and $\|e_i(Y_p)\|_2 \leqslant \varepsilon_0$. More precise estimates of $\|e_i(Y_0^{\text{near}})\|_2$ and $\|e_i(Y_p)\|_2$ can be used to obtain a sharper error bound on $\|e_i(Y_0)\|_2$. These estimates would be related to the actual distribution of the points Y_0^{near} in the domain between $\mathcal{X}$ and Γ^{proxy}.

6. Numerical experiments

In the following tests, we verify our main error analysis results of the proxy surface method in the reduced case that the points in Y_0 all lie in domain $\mathcal{Y}$ and outside the proxy surface Γ^{proxy}. The precision threshold for the ID approximation of $K(X_0, Y_p)$ is set to $\varepsilon\sqrt{|Y_p|}$ so that $\|e_i(Y_p)\|_2 \leqslant \varepsilon\sqrt{|Y_p|}$ for each $x_i \in X_0$. The value of ε will be specified for each test below. The parameter C_{qr} for SRRQR in the ID approximation of $K(X_0, Y_p)$ is set to 2.

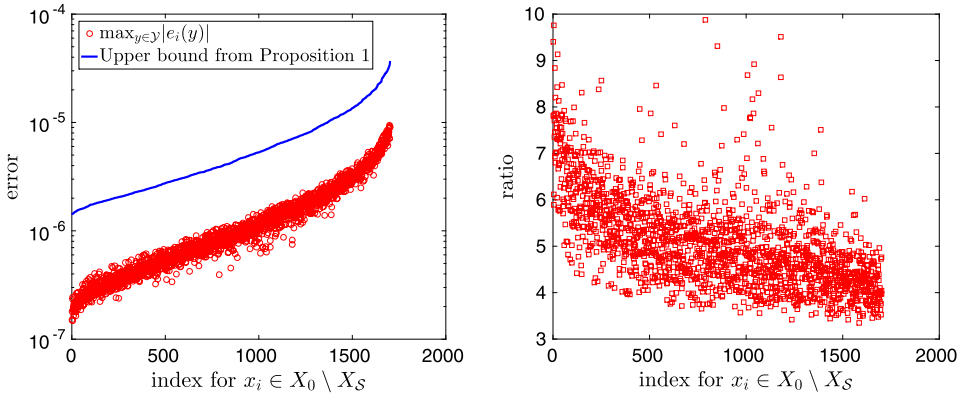


Fig. 4. Values of $\max_{y\in\mathcal{Y}} |e_i(y)|$ and its upper bound (12) for each $x_i \in X_0 \backslash X_{\mathcal{S}}$ with (a) values of the two quantities and (b) ratio of the upper bound to $\max_{y\in\mathcal{Y}} |e_i(y)|$. Indices for $x_i \in X_0 \backslash X_{\mathcal{S}}$ are sorted so that the upper bounds are in ascending order.

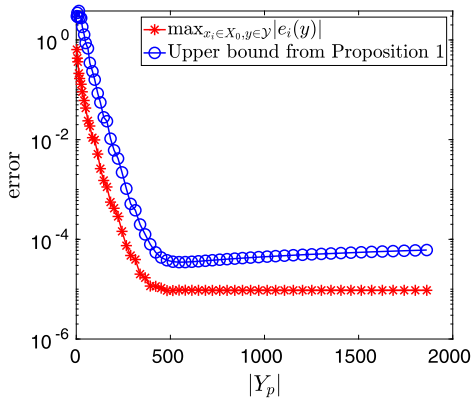


Fig. 5. Values of $\max_{x_i\in X_0, y\in\mathcal{Y}} |e_i(y)|$ and its upper bound (28) for different constants c and corresponding different point sets Y_p selected from Ref. [14]. The precision threshold $\varepsilon\sqrt{|Y_p|}$ with fixed $\varepsilon = 10^{-6}$ is used in the ID approximation of $K(X_0, Y_p)$ for different Y_p.

6.1. Error bound for $e_i(y)$ in Proposition 1

Consider $\mathcal{X} = B(0,1)$, $\mathcal{Y} = \mathbb{R}^3 \backslash B(0,2)$, and $\varepsilon = 10^{-6}$. The constant c estimated by (25) is 30 and the corresponding set of points Y_p selected from Ref. [14] has 1862 points. We randomly and uniformly selected 2000 points in $\mathcal{X}$ for X_0. The ID approximation $K(X_0, Y_p) \approx UK(X_{\mathcal{S}}, Y_p)$ calculated by SRRQR with precision $\varepsilon\sqrt{|Y_p|}$ has $X_{\mathcal{S}}$ with 298 points. Lastly, for each $x_i \in X_0$, the error function $e_i(y) = K(x_i, y) - u_i^T K(X_{\mathcal{S}}, y)$ is defined using the obtained U and $X_{\mathcal{S}}$.

To check the error bound (12) in Proposition 1, we plot $\max_{y\in\mathcal{Y}} |e_i(y)|$ and its upper bound (12) in Fig. 4 for each $x_i \in X_0 \backslash X_{\mathcal{S}}$ (for any $x_i \in X_{\mathcal{S}}$, $e_i(y)$ is the zero function). We estimate $\max_{y\in\mathcal{Y}} |e_i(y)|$ by densely sampling $|e_i(y)|$ over Γ^{proxy}, cf. (13). As can be observed, the difference between the upper bound (12) and $\max_{y\in\mathcal{Y}} |e_i(y)|$ is usually within an order of magnitude. However, the ratio of these two quantities being always greater than 3 indicates that an even sharper upper bound may exist.

In a further numerical test of Proposition 1, we vary the constant c and the corresponding Y_p selected using Ref. [14] for the same set of points X_0 and the same ε. For different Y_p, Fig. 5 plots $\max_{x_i\in X_0, y\in\mathcal{Y}} |e_i(y)|$ and its upper bound derived from Proposition 1, i.e.,

$$\max_{x_i\in X_0, y\in\mathcal{Y}} |e_i(y)| \leqslant (c+1)\varepsilon + (c+2)\frac{1 + \max_i \|u_i\|_\infty |X_{\mathcal{S}}|}{r_2 - r_1}\left(\frac{r_1}{r_2}\right)^{c+1}. \tag{28}$$

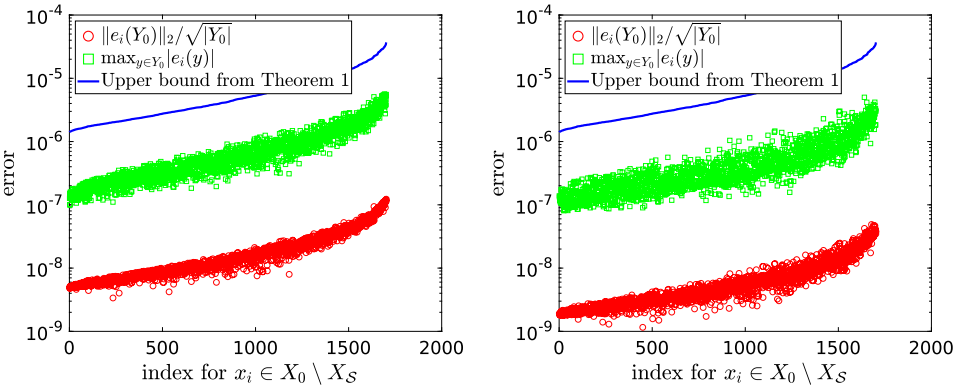


Fig. 6. Values of $\|e_i(Y_0)\|_2/\sqrt{|Y_0|}$, $\max_{y\in Y_0}|e_i(y)|$, and their shared upper bound (22) in Theorem 1 for two different Y_0. The indices x_i are sorted so that the upper bounds are in ascending order.

It can be observed that the upper bound (28) is quite tight and also identifies the knee at $|Y_p| \approx 500$ where $\max_{x_i\in X_0, y\in\mathcal{Y}}|e_i(y)|$ stops decreasing. The fact that $\max_{x_i\in X_0, y\in\mathcal{Y}}|e_i(y)|$ does not further decrease with larger $|Y_p|$ is due to the precision threshold $\varepsilon\sqrt{|Y_p|}$ used in the ID approximation of $K(X_0, Y_p)$. The knee also indicates that approximately 500 points for Y_p is sufficient to obtain the best possible precision for the proxy surface method in this specific case, i.e., $r_1 = 1$, $r_2 = 2$, and $\varepsilon = 10^{-6}$. However, the method of choosing Y_p in Section 4 gives $c = 30$ and $|Y_p| = 1862$. The main cause of this overestimation of $|Y_p|$, by comparing (28) and (25), turns out to be the looseness of $|X_S| \leqslant \min(|X_0|, |Y_p|)$ utilized in (25).

6.2. Error bound for $\|e_i(Y_0)\|_2$ in Theorem 1

The upper bound for $\|e_i(Y_0)\|_2$ in Theorem 1 simply combines the upper bound of $\max_{y\in\mathcal{Y}}|e_i(y)|$ in Proposition 1, which has been shown to be quite tight in the previous test, and the inequality (11), i.e., $\|e_i(Y_0)\|_2 \leqslant \sqrt{|Y_0|}\max_{y\in\mathcal{Y}}|e_i(y)|$. Equality in (11) is achieved when $|e_i(y)|$ reaches its maximum in $\mathcal{Y}$ at all the points in Y_0. However, for an arbitrary set of points Y_0, the inequality (11) turns out to be very loose as demonstrated below.

Using the same set of $\mathcal{X}$, $\mathcal{Y}$, ε, and $X_0 \subset \mathcal{X}$ as in the previous test, we considered the set of points Y_p associated with $c = 30$. We randomly and uniformly selected 20000 points for Y_0 in two subdomains of $\mathcal{Y}$, $B(0,4)\backslash B(0,2)$ and $B(0,8)\backslash B(0,2)$. For the proxy surface method, the root-mean-square error $\|e_i(Y_0)\|_2/\sqrt{|Y_0|}$ and the maximum entry-wise error $\max_{y\in Y_0}|e_i(y)|$ for each $x_i \in X_0\backslash X_S$ are plotted in Fig. 6 along with their shared upper bound (22) in Theorem 1.

For both subdomains of $\mathcal{Y}$ from which Y_0 is selected, $\|e_i(Y_0)\|_2/\sqrt{|Y_0|}$ is more than one order of magnitude smaller than $\max_{y\in Y_0}|e_i(y)|$. Thus, in these cases, the inequality (11) is very loose, which makes the upper bound in Theorem 1 very loose. However, as mentioned earlier, the inequality (11) is the best upper bound for $e_i(Y_0)$ if no further assumptions on Y_0 are made other than $Y_0 \subset \mathcal{Y}$.

6.3. Selection of Y_p

From Section 4, the selection of Y_p mainly depends on the domains $\mathcal{X}$ and $\mathcal{Y}$ and the value ε in the precision threshold $\varepsilon\sqrt{|Y_p|}$ for the ID approximation of $K(X_0, Y_p)$. Varying these parameters, Table 1 lists the number of points in Y_p selected by the method described in Section 4. Although our selection method is quite conservative as shown previously in Fig. 5, the results in Table 1 clearly show how the number of points chosen for Y_p is affected by the problem parameters.

Table 1
Estimated constant c and number of points selected for Y_p under different settings of radius r_1 for $\mathcal{X} = B(0, r_1)$, radius r_2 for $\mathcal{Y} = \mathbb{R}^3 \backslash B(0, r_2)$, and ε in the ID precision threshold $\varepsilon\sqrt{|Y_p|}$.

	r_1	r_2	r_1/r_2	ε	c	$\lvert Y_p\rvert$
Reference test	1	2	0.5	10^{-6}	30	1862
Different ε	1	2	0.5	10^{-4}	23	1106
	1	2	0.5	10^{-8}	38	2965
Different $\frac{r_1}{r_2}$	1	4	0.25	10^{-6}	12	314
	1	6	0.16	10^{-6}	9	181
Different $r_2 - r_1$	10	20	0.5	10^{-6}	27	1514
	100	200	0.5	10^{-6}	23	1106

7. Conclusion

The error analysis of the proxy surface method has now been established by showing the quantitative relationship (22) between the ID approximation error of $K(X_0, Y_0)$ and the ID approximation error of $K(X_0, Y_p)$. The analysis also provides an estimate of the number of points needed to discretize proxy surfaces under different problem settings. The same error analysis technique can be applied to the proxy surface method for more general data-sparse matrices with entries defined by the interactions between two compact charge distributions, e.g., matrices in the Galerkin method for integral equations and electron repulsion integral tensors with Gaussian-type basis functions in quantum chemistry.

References

[1] P.G. Martinsson, V. Rokhlin, A fast direct solver for boundary integral equations in two dimensions, J. Comput. Phys. 205 (1) (2005) 1–23.
[2] W.Y. Kong, J. Bremer, V. Rokhlin, An adaptive fast direct solver for boundary integral equations in two dimensions, Appl. Comput. Harmon. Anal. 31 (3) (2011) 346–369.
[3] A. Gillman, P.M. Young, P.-G. Martinsson, A direct solver with $O(N)$ complexity for integral equations on one-dimensional domains, Front. Math. China 7 (2) (2012) 217–247.
[4] K. Ho, L. Greengard, A fast direct solver for structured linear systems by recursive skeletonization, SIAM J. Sci. Comput. 34 (5) (2012) A2507–A2532.
[5] E. Corona, P.G. Martinsson, D. Zorin, An $O(N)$ direct solver for integral equations on the plane, Appl. Comput. Harmon. Anal. 38 (2) (2015) 284–317.
[6] S. Chandrasekaran, M. Gu, T. Pals, A fast ULV decomposition solver for hierarchically semiseparable representations, SIAM J. Matrix Anal. Appl. 28 (3) (2006) 603–622.
[7] J. Xia, S. Chandrasekaran, M. Gu, X.S. Li, Fast algorithms for hierarchically semiseparable matrices, Numer. Linear Algebra Appl. 17 (6) (2010) 953–976.
[8] L. Ying, G. Biros, D. Zorin, A kernel-independent adaptive fast multipole algorithm in two and three dimensions, J. Comput. Phys. 196 (2) (2004) 591–626.
[9] L. Ying, A kernel independent fast multipole algorithm for radial basis functions, J. Comput. Phys. 213 (2) (2006) 451–457.
[10] X. Xing, E. Chow, Interpolative decomposition via proxy points for kernel matrices, SIAM J. Matrix Anal. Appl. (2019), to appear.
[11] H. Cheng, Z. Gimbutas, P.G. Martinsson, V. Rokhlin, On the compression of low rank matrices, SIAM J. Sci. Comput. 26 (4) (2005) 1389–1404.
[12] M. Gu, S. Eisenstat, Efficient algorithms for computing a strong rank-revealing QR factorization, SIAM J. Sci. Comput. 17 (4) (1996) 848–869.
[13] K.E. Atkinson, The Numerical Solution of Integral Equations of the Second Kind, Cambridge University Press, Cambridge, 1997.
[14] R.S. Womersley, Efficient spherical designs with good geometric properties, in: Contemporary Computational Mathematics-A Celebration of the 80th Birthday of Ian Sloan, Springer, 2018, pp. 1243–1285.