A fast direct solver for scattering problems involving elongated structures

P.G. Martinsson, V. Rokhlin *

University of Colorado at Boulder, Applied Mathematics, 526 UCB, Boulder, CO 800277, United States

Abstract

A fast solver is presented for the solution of scattering problems in which the scatterer is a relatively thin, elongated object. The scheme presented here is a version of an algorithm previously published by the authors, and is based on the observation that under certain conditions and with certain modifications, the scheme will retain its O(n) CPU time estimate independently of the size of the scatterer in wavelengths. The performance of the scheme is illustrated with numerical examples.

© 2006 Published by Elsevier Inc.

Keywords: Direct solvers; Scattering problems

1. Introduction

A standard approach to the numerical solution of large-scale scattering problems is to reduce the problem to a set of integral equations on the boundary of the scatterer (or scatterers), discretize the integral equations via an appropriate numerical scheme (usually, Nyström or Galerkin) and to solve the resulting system of linear algebraic equations iteratively; GMRES appears to be the preferred iterative solver in this environment. While in many situations the existing methods are quite satisfactory, the iterative methods tend to display slow convergence whenever the condition number of the underlying scattering problem is high (*i.e.* when the system is close to a resonance). Furthermore, when the problem to be solved involves a single scattering structure irradiated from many directions (a frequently encountered situation), iterative methods obtain only limited advantage from the resulting multiple right-hand sides.

In this paper, we present an algorithm for the direct inversion of integral operators associated with scattering problems involving thin, elongated scatterers. The computational cost of the scheme is O(N), where N is the number of nodes in the discretization of the surface of the scatterer (which in this case tends to be proportional to the length of the scatterer in wavelengths). The algorithm of this paper should be viewed as a development of

---

* Corresponding author. Tel.: +1 303 898 4852.
  E-mail address: martinss@colorado.edu (V. Rokhlin).

the direct inversion scheme found in [12]. A class of problems very close to the one addressed here is treated in [14], where an $O(N\log^2 N)$ algorithm is presented; [14] is in turn an extension of [8].

This paper is structured as follows: Section 2 defines a model problem and lists some known techniques for discretizing and solving such problems. Section 3 contains the analytical apparatus we use in the construction of the numerical scheme of this paper, and in Section 4, we describe the "fast" algorithm. Section 5 illustrates the performance of the scheme via several numerical examples involving scatterers that are up to 3000 wavelengths in size. Finally, Section 6 contains generalizations and conclusions.

2. Preliminaries

In this section, we describe a scattering problem that will serve as our model, and reduce the model problem to an integral equation of the second kind on the boundary of the scatterer. We also summarize some of the standard techniques for the numerical solution of the obtained integral equation.

2.1. Integral equations resulting from scattering problems

For definitiveness, we will be discussing the solution of the exterior Dirichlet problem. Specifically, given a complex wavenumber k (which is assumed to have a non-negative imaginary part) and a subset Ω of $\mathbb{R}^2$ bounded by a sufficiently smooth Jordan curve Γ, we seek the function $u_{\text{out}} : \mathbb{R}^2 \setminus \Omega \to \mathbb{C}$ satisfying the Helmholtz equation

$$-\Delta u_{\text{out}} - k^2 u_{\text{out}} = 0, \tag{2.1}$$

subject to the radiation condition

$$\lim_{t \to \infty} \frac{t u_{\text{out}}(tx)}{e^{ikt}} = c(x) \tag{2.2}$$

for any $x \in \mathbb{R}^2$ such that $\|x\| = 1$, and to the boundary condition

$$u_{\text{in}} + u_{\text{out}} = 0 \tag{2.3}$$

on Γ, with $u_{\text{in}} : \Gamma \to \mathbb{C}$ a continuous function.

In order to solve the boundary value problem (2.1)–(2.3), numerically, we make the Ansatz

$$u_{\text{out}}(x) = \int_\Gamma \left(\frac{\partial}{\partial n(y)} H_0(k|x-y|) + ik H_0(k|x-y|) \right) \sigma(y) ds(y), \tag{2.4}$$

where $n(y)$ is the outward pointing unit normal of Γ at the point y, σ is a function to be determined, and H_0 is the Hankel function of the first kind of order zero. (This choice of an Ansatz, which leads to what is called the "combined field integral equation", is discussed in detail, *inter alia*, in [7].) The function u_{out} defined by (2.4) automatically satisfies the Helmholtz Eq. (2.1) with the radiation condition (2.2), and the boundary condition (2.3) is satisfied if and only if σ satisfies the equation

$$-2i\sigma(x) + \int_\Gamma \left(\frac{\partial}{\partial n(y)} H_0(k|x-y|) + ik H_0(k|x-y|) \right) \sigma(y) ds(y) = -u_{\text{in}}(x) \tag{2.5}$$

for all $x \in \Gamma$. In effect, the process described in the preceding paragraph converts the boundary value problem (2.1)–(2.3) into the integral Eq. (2.5). The purpose of this paper is to solve the Eq. (2.5) rapidly in the case where the scatterer is a relatively thin, elongated object.

2.2. Nyström discretization of integral equations

Having formulated a scattering problem as an integral equation of the form

$$v(x) + \int_\Gamma K(x,y) v(y) ds(y) = f(x), \quad \text{for } x \in \Gamma \tag{2.6}$$

(as discussed in Section 2.1), the next step is to approximate the continuum Eq. (2.6) by a system of N linear algebraic equations. There are several ways of doing so; throughout this paper, we have utilized a version of the Nyström method. This technique has been described in detail elsewhere (see, *e.g.* [2,7]), and here we simply observe that the end result of the process is a set of N linear algebraic equations of the form

$$(I + A)v^{(N)} = f^{(N)}, \tag{2.7}$$

where A is a matrix that approximates the integral operator. It is important to note that when the Nyström method is used, the entries of the matrix A are in a certain sense samples of the kernel function $K(x, y)$. To be precise, if the contour is discretized into the points $\{x_i\}_{i=1}^{N}$, and if we let a_{ij} denote the ijth entry of A, then for some numbers $(\varphi_i)_{i=1}^{N}$, $(\psi_i)_{i=1}^{N}$,

$$a_{ij} = \varphi_i K(x_i, x_j)\psi_j, \quad \text{when } |i - j| \geqslant m. \tag{2.8}$$

In Eq. (2.8), the matrix elements in a diagonal band of width m are excluded since the kernel function is typically singular near the diagonal and the quadrature weights must then be adjusted to retain high order accuracy, see, for example [11]. (Typically, $m \leqslant 10$.)

2.3. Iterative solvers for the linear systems resulting from scattering problems

The system of linear algebraic equations (2.7) is commonly solved using an iterative solver such as GMRES in conjunction with a fast matrix-vector multiplication algorithm such as the Fast Multipole Method that computes a matrix-vector product in O(N) operations. If the iterative solver needs N_{iter} iterations to converge to some specified accuracy, then the entire solution process requires O($N_{\text{iter}} N$) floating point operations. In many environments, N_{iter} is small and this approach works very well.

The principal drawback of iterative methods in the context of high-frequency scattering problems is that such problems are almost always close to being resonant, in which case a large number of iterations may be needed (since the system (2.7) gets ill-conditioned), which in turn slows down the solution process. Moreover, it is common that a sequence of scattering problems (corresponding to a collection of different right-hand sides) needs to be solved with the same geometry, which is a situation that favors direct methods over iterative ones.

2.4. Direct solvers for the linear systems resulting from scattering problems

If Gaussian elimination or a similar scheme were used to solve the system (2.7), the number of floating point operations required would be O(N^3), which would severely limit the size of problems that could be solved in practice. However, several schemes for the acceleration of the direct solution of linear systems arising from integral equations have been proposed over the last decade, see *e.g.* [4,10,12]. Most such schemes rely on rank-deficiencies in the off-diagonal blocks of the matrix, which means that they do not perform asymptotically fast for high-frequency problems (we say that a method is "fast" if its computational speed is O($N\log^q N$) for some integer q). An algorithm presented in [5] reduces the computational complexity to O($N^{3/2}$) for two-dimensional Lippmann–Schwinger problems (and would yield O(N^2) complexity in three dimensions), but to the authors' best knowledge, there does not exist a "fast" algorithm (in the sense described above) for the direct solution of high frequency problems on general domains.

However, for the special case of scattering from elongated objects (see Section 4 for a precise definition), it was demonstrated in [14] that the system (2.7) can be solved directly in O($N\log^2 N$) operations (an earlier result in this direction was given in [8]). The explanation for this is to be found in the observation that for elongated scatterers, the rank of interaction between two separated pieces of the scatterer scales as the *logarithm* of the size of the pieces (measured in wave lengths), rather than scaling linearly with the size of the pieces as would be the case for a general scatterer. Versions of this observation are formulated in, *e.g.* [3,8,9,14].

In Section 3, we formalize and extend some of the observations found in [3,8,9,14]. The result will enable the conversion of many fast inversion schemes devised for non-oscillatory problems (such as, *e.g.* those presented in [4,10,12]) to fast algorithms for high-frequency scattering problems on elongated scatterers. We describe such a modification of the scheme of [12] in Section 4, and illustrate the performance of the resulting scheme on some practical examples in Section 5.

2.5. An interpolation result

The following lemma (that will be used in Section 3) states that for any J-dimensional linear space X of continuous functions on a given compact set, there exists a set of J interpolation points such that any function in X can be interpolated *exactly* from its values at those J points. Moreover, none of the interpolation coefficients are large. Lemma 1 is a special case of Theorem 3.2 of [13].

Lemma 1. *Suppose that Ω is a compact set in $\mathbb{R}^2$, J is a positive integer, and that $\chi_1, \ldots, \chi_J$ are continuous complex-valued functions on Ω. Then there exist J points $y_1, \ldots, y_J$ in Ω and J functions $\varphi_1, \ldots, \varphi_J$ on Ω such that, for $j = 1, \ldots, J$ and $y \in \Omega$,*

$$\chi_j(y) = \sum_{q=1}^{J} \chi_j(y_q)\varphi_q(y) \tag{2.9}$$

and

$$|\varphi_j(y)| \leqslant 1. \tag{2.10}$$

3. Low-rank approximations

In this section, we prove certain estimates regarding the ranks of interaction between different parts of an elongated scatterer. The starting point is Theorem 2 which can be found (in a somewhat different form) in [14]. It states that within precision ε, the rank of interaction between the two boxes shown in Fig. 1, of lengths L and fixed widths, is $O(\log(kL)|\log \varepsilon|^2)$, as $L \to \infty$ and $\varepsilon \to 0$. A proof of Theorem 2 can be found in Appendix A to this paper. Theorems 3–5 are the purpose of this section; each of them can be viewed as a modification of Theorem 2. Theorems 3–5 will be used in the construction of low-rank factorizations of off-diagonal blocks in the matrices to be inverted.

In the following theorem (as elsewhere in this paper), H_0 denotes the Hankel function of the first kind of order zero.

Theorem 2. *Suppose that L, W, k, ε, S are real numbers such that $L > 0$, $W \geqslant 0$, $k > 0$, $0 < \varepsilon < 1/2$, $S > 0$, and $\Omega_S = [-L, 0] \times [-W/2, W/1]$, $\Omega_T = [S, S + L] \times [-W/2, W/2]$ are two boxes in $\mathbb{R}^2$ (see Fig. 1).*

Then there exists an integer J, functions $\psi_j : \Omega_S \to \mathscr{C}$, $\chi_j : \Omega_T \to \mathscr{C}$, with $j = 1, 2, \ldots, J$, and a real number c that does not depend on either L or ε such that if

$$S \geqslant c|\log \varepsilon|/k, \tag{3.1}$$

then

$$H_0(k|x - y|) = \sum_{j=1}^{J} \psi_j(x)\chi_j(y) + E(x, y), \quad \text{for all } x \in \Omega_T, \ y \in \Omega_S, \tag{3.2}$$

where

$$|E(x, y)| \leqslant \varepsilon. \tag{3.3}$$

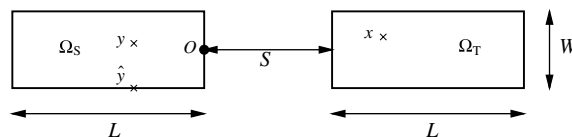*Furthermore, there exists a number C such that, as $L \to \infty$, and $\varepsilon \to 0$,*

Fig. 1. The boxes Ω_S and Ω_T contain source and target points, respectively. A "typical" source point y and a "typical" target point x are marked with crosses. The point $\hat{y}$ is the vertical projection of y onto the bottom boundary. The origin in the coordinate system used is labeled O.

$$J \leqslant C \log(kL) |\log \varepsilon|^2. \tag{3.4}$$

Finally, the functions ψ_j and χ_j are smooth and bounded uniformly in L and ε.

Remark 1. In Theorem 2, there is a requirement that the boxes must be separated by a distance of $O(|\log \varepsilon|/k)$. The dependence of ε in this relation can for practical purposes be ignored. If $\varepsilon > 10^{-15}$, then a separation $S = 1/k$ works well. The reason is that the sum (A.13) in the proof converges very fast; for double precision accuracy, no more than 15 terms are ever needed. Another consequence of this fact is that the square on the $|\log \varepsilon|$ factor in (3.4) is generally not noticeable in practical calculations.

Remark 2. Theorem 2 does not specify how the interaction rank J depends on the width W of the boxes. In fact, it grows quite rapidly with W, and the theorem is generally not practically useful for W larger than a few wavelengths. Some actual values of J for various combinations of kL, kW, kS and ε, are given in Section 5.1.

Theorem 3 sharpens Theorem 2 by stating that the functions $\psi_j(x)$ in the expansion (3.2) can in fact be chosen to take the form $\psi_j(x) = H_0(k|x - y_j|)$ for some points $y_j \in \Omega_S$. This assertion is proved by applying the interpolation result of Lemma 1 to the expansion (3.2).

Theorem 3. *Let L, W, k, S, ε, Ω_S, Ω_T and J be as in Theorem 2. Then there exist points $\{y_j\}_{j=1}^J$ in Ω_S, and functions $\{\varphi_j\}_{j=1}^J$, such that,*

$$H_0(k|x - y|) = \sum_{j=1}^J H_0(k|x - y_j|) \varphi_j(y) + E(x, y) \quad \text{for } x \in \Omega_T, \ y \in \Omega_S, \tag{3.5}$$

where

$$|E(x, y)| \leqslant \varepsilon (J + 1). \tag{3.6}$$

Moreover, for $j = 1, \ldots, J$ and $y \in \Omega_S$,

$$|\varphi_j(y)| \leqslant 1. \tag{3.7}$$

Proof. In order to convert the results of Theorem 2 to those of Theorem 3, we apply the interpolation method of Lemma 1 to the set of functions $\chi_1, \ldots, \chi_J$. This yields a set of points $\{y_q\}_{q=1}^J$ in Ω_S, and interpolants $\{\varphi_q\}_{q=1}^J$ satisfying (3.7). Combining (2.9) and (3.2), we find that

$$H_0(k|x - y|) = \sum_{j=1}^J \sum_{q=1}^J \psi_j(x) \chi_j(y_q) \varphi_q(ky) + E(x, y). \tag{3.8}$$

Another application of (3.2) yields

$$H_0(k|x - y|) = \sum_{q=1}^J (H_0(k|x - y_q|) - E(x, y_q)) \varphi_q(y) + E(x, y), \tag{3.9}$$

and the bound (3.6) is seen to be a direct consequence of (3.3) and (3.7). $\square$

The following theorem is the principal result of this section. It is simply a corollary of Theorem 3 obtained by integrating the expansion (3.5) against a charge distribution.

Theorem 4. *Let L, W, S, k, ε, Ω_S, Ω_T and J be as in Theorem 2, and let the functions $\varphi_1, \ldots, \varphi_J$ be as in Theorem 3. Furthermore, for any function $\sigma \in L^1(\Omega_S)$, we denote by u the potential on Ω_T induced by a charge distribution on Ω_S of density σ. In other words, for $x \in \Omega_T$,*

$$u(x) = \int_{\Omega_S} H_0(k|x - y|) \sigma(y) \mathrm{d}A(y). \tag{3.10}$$

Then, if we define for j = 1,...,J, the "equivalent" charges

$$\sigma_j = \int_{\Omega_S} \varphi_j(y)\sigma(y)\mathrm{d}A(y), \tag{3.11}$$

it is the case that, for $x \in \Omega_T$,

$$u(x) = \sum_{j=1}^{J} H_0(k|x - y_j|)\sigma_j + E(x), \tag{3.12}$$

where

$$|E(x)| \leqslant (J + 1)\varepsilon \int_{\Omega_S} |\sigma(y)|\mathrm{d}A(y). \tag{3.13}$$

Proof. Insert the expansion (3.5) into (3.10). □

Remark 3. In physical terms, Theorem 4 can be interpreted as follows: suppose that Ω_S and Ω_T are as shown in Fig. 1. Then there exist J points inside Ω_S with the property that any potential in Ω_T caused by a charge distribution in Ω_S, can to within precision ε be replicated by placing point charges at those J points only. Moreover, the number of points depends only logarithmically on the lengths of the boxes, and the precision required.

The functions $\{\varphi_j\}_{j=1}^{J}$ in Theorem 4 form a basis for the support of the integral operator that maps σ to u in (3.10). The following theorem provides an analogous basis $\{v_i\}_{i=1}^{J}$ for the range of the same integral operator.

Theorem 5. *Let L, W, S, k, ε, Ω_S, Ω_T and J be as in Theorem 4, and let for any $\sigma \in L^1(\Omega_S)$, the potential u be defined by (3.10). Then there exist J points $\{x_i\}_{i=1}^{J}$, and J interpolation functions $\{v_i\}_{i=1}^{J}$ with the property that, for $x \in \Omega_T$,*

$$u(x) = \sum_{i=1}^{J} v_i(x)u(x_i) + E(x), \tag{3.14}$$

where E satisfies (3.13), and, for i = 1,...,J and $x \in \Omega_T$,

$$|v_i(x)| \leqslant 1. \tag{3.15}$$

Proof. We let $\{\sigma_j\}_{j=1}^{J}$ and $\{y_j\}_{j=1}^{J}$ be defined as in Theorem 4, and set

$$u_\varepsilon(x) = \sum_{j=1}^{J} H_0(k|x - y_j|)\sigma_j. \tag{3.16}$$

Thus for any σ, the function u_ε belongs to the linear span of the functions $\{H_0(k|x - y_j|)\}_{j=1}^{J}$. Lemma 1 then directly yields the desired interpolation points and interpolation functions. □

Remark 4. Theorems 4 and 5 provide bases for the support and the range of the integral operator in (3.10), respectively. By combining the two theorems, we obtain a factorization of the integral operator into three factors where the middle one is simply the restriction of the original operator to the points $\{y_j\}_{j=1}^{J}$ and $\{x_i\}_{i=1}^{J}$. To be precise, by combining (3.12) and (3.14), we obtain the approximation

$$u(x) = \sum_{i,j=1}^{J} v_i(x)H_0(k|x_i - y_j|) \int_{\Omega_S} \varphi_j(y)\sigma(y)\mathrm{d}A(y) + E(x), \tag{3.17}$$

where E satisfies (3.13).

Remark 5. While Theorems 4 and 5 assert the existence of points $\{y_j\}_{j=1}^{J}$ and $\{x_i\}_{i=1}^{J}$ such that (3.12) and (3.14) hold, they do not say anything about how to actually find such points. Methods for this task are described in [6,13]. The points determined by these methods tend to lie close to the boundaries of Ω_S and Ω_T, as seen in Fig. 3b. In fact, it is possible to force the points to lie *exactly* on the boundary, although one must then use both monopole and dipole charges (as opposed to the representation (3.12) that uses monopoles only).

4. Fast direct solvers for the integral equations associated with elongated scatterers

In this section, we describe how to modify the scheme of [12] to obtain a fast direct solver for the integral equation

$$u(x) + \int_\Gamma H_0(k|x - y|)u(y)\mathrm{d}s(y) = f(x), \tag{4.1}$$

where f is a given function, Γ is an given contour, and k is the wave-number. Under quite weak conditions on Γ, the scheme described in [12] computes an approximate inverse of the Fredholm operator in (4.1) using $O(N + (kL)^3)$ arithmetic operations, where N is the number of discretization points, and L is the diameter of Γ. In this section, we describe a scheme for eliminating the superlinear dependence on kL for the special case where Γ is long and narrow. Similar schemes appear in [8,14]. For simplicity, we assume that k is real and positive, even though the scheme works for complex k as well, as long as the imaginary part of k is non-negative. The discussion is framed around Eq. (4.1) for simplicity, but all techniques described apply more generally to the integral equations discussed in Section 2.1.

We say that a contour Γ is "elongated," if it fits inside a rectangle of width W, and length L, where $W \ll L$, cf. Fig. 2. In discussing the asymptotic cost of solving (4.1) as the contour Γ grows, we assume that all contours considered fit inside rectangles of a fixed width W, but with different lengths L. Moreover, we assume that the number N of discretization nodes is proportional to kL, and that the nodes are more or less evenly distributed along the length of the contour. Under these conditions, the direct solver of [12] can be modified in such a way that its asymptotic cost is $O(N)$, or equivalently $O(kL)$. The scaling factor depends on kW. For practical purposes, W should be no larger than a few wave-lengths, cf. Remark 2, and Section 5.1.

The fast direct solver of [12] relies crucially on the fact that for Laplace problems, the rank of interaction between well-separated subsets of the contour Γ remains small regardless of the sizes of the subsets. For high-frequency scattering problems involving scatterers of general shapes, these interaction ranks typically grow too fast for the scheme of [12] to retain its $O(N)$ complexity. However, the results of Section 3 demonstrate that for the special case of elongated contours, the ranks grow only logarithmically with size, which in principle enables the scheme to achieve $O(N)$ complexity. To practically attain this, the scheme needs to be furnished with a technique for rapidly constructing certain low-rank approximations to the interaction matrices. The following observation provides such a technique:

Observation 6. *Let Γ denote an elongated contour of horizontal length L, split into two pieces Γ_1 and Γ_2, as shown in Fig. 3a. Furthermore, let m and n denote the number of points in the discretizations of Γ_1 and Γ_2, respectively, let B denote the $m \times n$ matrix discretizing the integral operator that maps a charge distribution on Γ_2 to a potential on Γ_1, and let $\{x_i\}_{i=1}^m$ denote the points in the discretization of Γ_1. Then for any given precision ε, there exists a set of points $\{y_j\}_{j=1}^p$ such that the vectors*

$$\begin{bmatrix} H_0(k|x_1 - y_1|) \\ \vdots \\ H_0(k|x_m - y_1|) \end{bmatrix}, \ldots, \begin{bmatrix} H_0(k|x_1 - y_p|) \\ \vdots \\ H_0(k|x_m - y_p|) \end{bmatrix}, \tag{4.2}$$

constitute a basis for the column space of B, accurate to within precision ε. Moreover, there exists a number C, depending on neither ε, nor L, such that
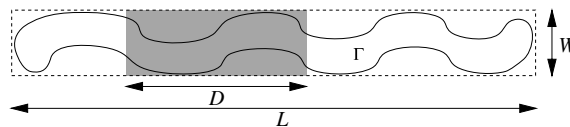

Fig. 2. An "elongated" contour Γ (drawn with a solid line). We assume that the number of discretization points m inside any box such as the one shaded above, is proportionate to its length along the horizontal direction, $m \sim kD$.
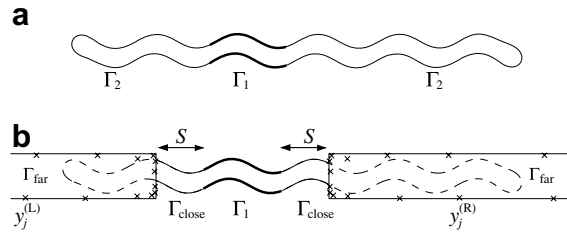
Fig. 3. (a) A contour $\Gamma = \Gamma_1 + \Gamma_2$. Γ_1 is drawn with a bold line and Γ_2 with a thin line. (b) A contour $\Gamma = \Gamma_1 + \Gamma_{\text{close}} + \Gamma_{\text{far}}$. Γ_1 is drawn with a bold line, Γ_{close} with a thin line, and Γ_{far} with a dashed line. The interaction between Γ_1 and Γ_{far} is replaced by the interaction between Γ_1 and the proxy points $y_j^{(L)}$, $y_j^{(R)}$, drawn with crosses.

$$p \leqslant C \log(kL) |\log \varepsilon|^2. \tag{4.3}$$

The cost of finding the points $\{y_j\}_{j=1}^p$ is $O(p)$.

Proof. We let $\{y_j^{(0)}\}_{j=1}^J$ denote the points provided by Theorem 4 in order to approximate a potential associated with a box of length L to precision ε, and let S be the corresponding prescribed separation between the target box and the source box. Given this separation S, we let Γ_{close} denote the part of Γ_2 that is within distance S of Γ_1 in the horizontal direction, and let Γ_{far} denote the rest, see Fig. 3b. Furthermore, we let B_{close} denote the matrix mapping a charge distribution on Γ_{close} onto a potential on Γ_1, and define B_{far} analogously. In other words, the matrices B and $[B_{\text{close}} \, B_{\text{far}}]$ are identical except for the ordering of the columns, and

$$\text{Col}(B) = \text{Col}(B_{\text{close}}) + \text{Col}(B_{\text{far}}), \tag{4.4}$$

where $\text{Col}(B)$ denotes the column space of B. We let $\{y_j^{(\text{close})}\}_{j=1}^{p_{\text{close}}}$ denote the points in the discretization of Γ_{close}. Then the vectors $\{[H_0(k|x_i - y_j^{(\text{close})}|)]_{i=1}^m\}_{j=1}^{p_{\text{close}}}$ form a basis for $\text{Col}(B_{\text{close}})$. It remains to find a basis of the required form for $\text{Col}(B_{\text{far}})$. But this is exactly what Theorem 4 does. Simply enclose the right part of Γ_{far} in a box of length L, and let $\{y_j^{(R)}\}_{j=1}^J$ denote the associated proxy source locations obtained from the set $\{y_j^{(0)}\}_{j=1}^J$ by shifting all points horizontally. Likewise, we enclose the left part in a similar box and obtain the source points $\{y_j^{(L)}\}_{j=1}^J$ by reflecting the points $\{y_j^{(0)}\}_{j=1}^J$ about a vertical axis and then shifting them. The union of the three sets $\{y_j^{(\text{close})}\}_{j=1}^{p_{\text{close}}}$, $\{y_j^{(L)}\}_{j=1}^J$ and $\{y_j^R\}_{j=1}^J$ now generates a basis of the required form. Finally, we note that $p = 2J + p_{\text{close}}$, that J satisfies the bound (3.4), that $p_{\text{close}} \sim kS$ since we assumed that the discretization points are evenly distributed along the horizontal axis, and that $kS \sim |\log \varepsilon|$ by (3.1). Together, these observations imply (4.3). $\square$

From Observation 6 it follows that only one modification is required to the direct solver of [12] to obtain an $O(N \log N)$ computational cost. Specifically, in Section 5 of [12] it is described how to replace the long-range interactions between a given patch $\Gamma_1 \subset \Gamma$ and the rest of the contour by the interaction between Γ_1 and a circular proxy contour $\widehat{\Gamma}_2$ that can be represented using only a handful of points. For the case of scattering problems involving an elongated scatterer, we can substitute the circular proxy contour by the set of points $\{y_j^{(L)}, y_j^{(R)}\}_{j=1}^J$. Thus, the interaction between Γ_1 and the well-separated part of the contour (called Γ_{far} in the proof of Observation 6) can be replaced by the interaction between Γ_1 and $2J$ proxy points. (We remark that Fig. 3b is analogous to Fig. 3b of [12].)

To further reduce the computational cost of the inversion from $O(N \log N)$ to $O(N)$, one must use sets of external proxy points of different sizes; when compressing the interactions with a small piece of the contour, only a small number of proxy points may be used. This is possible since the rank of interaction between the two pieces Γ_1 and Γ_{far} in Observation 6 is bounded from above by the size of the smaller of the two pieces.

5. Numerical examples

5.1. Numerical estimation of interaction ranks between elongated boxes

Theorem 2 states that to precision ε, the rank of interaction J between two elongated boxes of widths W and lengths L, separated by a distance S, as shown in Fig. 1, satisfies

$$J \leqslant C \log(kL)(\log(1/\varepsilon))^2, \tag{5.1}$$

for some number C, as $L \to \infty$ and $\varepsilon \to 0$ (the number k is the wave-number in the Helmholtz equations (2.1)). In this section, we present the results of some numerical experiments that indicate that in many cases of practical interest, J is quite small, typically between 20 and 40. Moreover, these examples indicate that even though Theorem 2 requires the separation distance S to be several times larger than the wavelength, a separation of only one half, or one wavelength, is sufficient for 5 or 10 digits of relative accuracy.

Table 1 gives upper bounds for the interaction ranks for a number of different values of L and W when the separation distance is one wave length (*i.e.* $kS = 2\pi$), while Table 2 gives the corresponding values when it equals one half wave length (*i.e.* $kS = \pi$). The ranks reported in the tables were computed using the methods described in [6] and [13].

5.2. Example: scattering off an elongated closed contour

We tested the performance of the inversion scheme of [12], with the modifications described in Section 4, by applying it to a set of closed contours having the undulating shape shown in Fig. 3a. The times required to invert the integral operator in (2.5) on a desktop PC with a 3 GHz Pentium IV processor and 3 Gb of RAM are reported in Table 3. In all experiments, the vertical size of the contour remained fixed at one wave-length. The horizontal size varied between 8 and 1024 wave-lengths and the contour was discretized using 50 points per wavelength. The "geometric" undulation of the contour has precisely twice the wavelength of the radiated field.

The integral equation was discretized using a Nyström method with the 6th order accurate quadrature rule described in [11]. With 50 points per wave-length, this yielded a discretization error of 10^{-10}.

Table 1
The table displays the number of interpolation points J actually required for the interpolation (3.12) for a separation S of one wave-length (*i.e.* $kS = 2\pi$)

$kL/2\pi$	$\varepsilon = 10^{-5}$						$\varepsilon = 10^{-10}$					
	10^1	10^2	10^3	10^4	10^5	10^6	10^1	10^2	10^3	10^4	10^5	10^6
$kW/2\pi = 1$	15	20	24	26	28	30	29	39	48	57	64	69
$kW/2\pi = 2$	20	25	28	30	32	34	36	48	57	66	73	79
$kW/2\pi = 3$	24	28	33	36	38	40	44	55	64	72	80	87

Table 2
The table displays the number of interpolation points J actually required for the interpolation (3.12) for a separation S of one half wave-length (*i.e.* $kS = \pi$)

$kL/2\pi$	$\varepsilon = 10^{-5}$						$\varepsilon = 10^{-10}$					
	10^1	10^2	10^3	10^4	10^5	10^6	10^1	10^2	10^3	10^4	10^5	10^6
$kW/2\pi = 1$	20	24	29	31	33	34	35	46	55	63	70	76
$kW/2\pi = 2$	27	33	37	40	42	44	45	55	65	73	82	87
$kW/2\pi = 3$	35	40	44	48	49	50	55	65	75	83	92	98

Table 3
Results of the numerical experiment described in Section 5.2.

n_{tot}	n_{wave}	t_{comp}	t_{apply}	E_{res}	M
800	8	2.38e0	6.60e−3	2.1e−6	7.5e0
1600	16	5.73e0	1.54e−2	3.3e−6	1.7e0
3200	32	1.27e1	3.26e−2	3.5e−6	3.6e1
6400	64	2.62e1	7.05e−2	5.2e−6	7.3e1
12,800	128	5.39e1	1.41e−1	7.5e−6	1.5e2
25,600	256	1.19e2	2.92e−1	9.9e−6	3.0e2
51,200	512	2.38e2	5.89e−1	1.2e−5	5.9e2
102,400	1024	4.79e2	1.19e−1	1.2e−5	1.2e3

n_{tot} is the total number of points in the discretization, n_{wave} the length of the scatterer in wavelengths (its width is 1 for all experiments), t_{comp} the time in seconds required to compute an approximate inverse, t_{apply} the time in seconds required to apply the compressed inverse to a vector, E_{res} the residual error in the computed inverse and M is the amount of memory (in megabytes) required to compute the inverse.

5.3. Example: scattering off an one-dimensional wavy surface

In this section, we consider the problem of determining the scattered wave from an infinite contour Γ of the form shown in Fig. 4. The contour Γ coincides with the x-axis everywhere, except on the finite piece Γ_+, drawn with a bold line in Fig. 4. To be precise, we are given an incoming field u_{in} of the form

$$u_{\text{in}}(x) = H_0(k|x - \hat{x}|),$$

where k is the wave-number, and $\hat{x}$ is a given point in the domain Ω above Γ, and we seek a field u_{out} such that

$$-\Delta u_{\text{out}} - k^2 u_{\text{out}} = 0, \quad \text{on } \Omega \tag{5.2}$$

and

$$u_{\text{out}} + u_{\text{in}} = 0, \quad \text{on } \Gamma, \tag{5.3}$$

while also satisfying the radiation condition (2.2). Letting $G(x, y)$ denote the kernel of the integral operator in (2.5), we make the Ansatz

$$u_{\text{out}}(x) = \int_{\Gamma_+} G(x, y)\sigma(y)\mathrm{d}s(y) - \int_{\Gamma_-} G(x, y_-)\sigma(y)\mathrm{d}s(y) - H_0(k|x - \hat{x}_-|), \tag{5.4}$$

where Γ_-, y_- and $\hat{x}_-$ are the reflections of Γ_+, y, and $\hat{x}$ in the x-axis, see Fig. 4. With this Ansats, the Helmholtz' equation (5.2) is automatically satisfied, and the boundary condition (5.3) is satisfied on $\Gamma \backslash \Gamma_+$. To ensure that (5.3) is satisfied on Γ_+ as well, it is necessary that σ satisfies the equation

$$-2\mathrm{i}\sigma(x) + \int_{\Gamma_+} G(x, y)\sigma(y)\mathrm{d}s(y) - \int_{\Gamma_-} G(x, y_-)\sigma(y)\mathrm{d}s(y) = H_0(k|x - \hat{x}_-|) - H_0(k|x - \hat{x}|), \quad \text{for } x \in \Gamma_+. \tag{5.5}$$
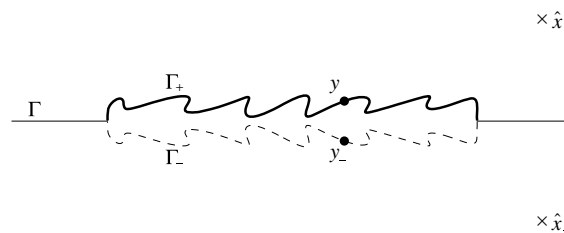
Fig. 4. Geometry of the scattering problem described in Section 5.3. The contour Γ is drawn with a solid line; the subset Γ_+ is highlighted using a bold line. The reflection Γ_- of Γ_+ is drawn with a dashed line. The point $\hat{x}$ is a source, and $\hat{x}_-$ its mirror image.

Table 4
Computational timings for the numerical experiment described in Section 5.3

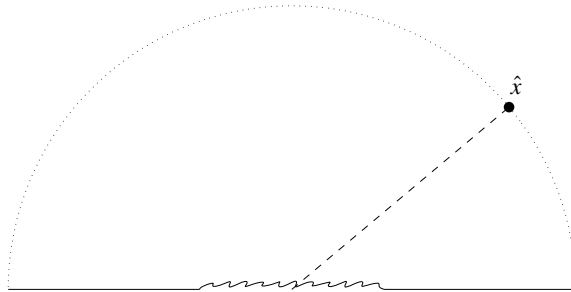| n_{tot}	n_{wave}	t_{comp}	t_{apply}	E_{res}	M
800	11	1.07e0	2.20e − 3	9.2e − 7	2.6e0
1600	23	2.32e0	4.70e − 3	1.8e − 7	5.5e0
3200	46	4.86e0	9.60e − 3	7.7e − 7	1.1e1
6400	91	1.12e1	1.94e − 2	6.2e − 7	2.3e1
12,800	183	2.30e1	3.90e − 2	8.8e − 7	4.7e1
25,600	366	4.58e1	7.92e − 2	4.7e − 6	9.4e1
51,200	731	1.01e2	1.59e − 1	4.9e − 6	1.9e2
102,400	1463	2.02e2	3.19e − 1	6.3e − 6	3.8e2
204,800	2926	4.06e2	6.36e − 1	1.7e − 5	7.6e2

The notation is the same as in Table 3.

Fig. 5. Computation of the reflection pattern of a wavy surface. A radiating source is placed at the point marked $\hat{x}$, the reflection problem is solved, and the intensity of the reflected wave at the point $\hat{x}$ is computed. A plot of this intensity, as a function of the angle between the dashed line and the positive x-axis, is given in Fig. 6. (Note that in the experiment reported, the wavy surface was 600 wave lengths long rather than the 11 shown here.)
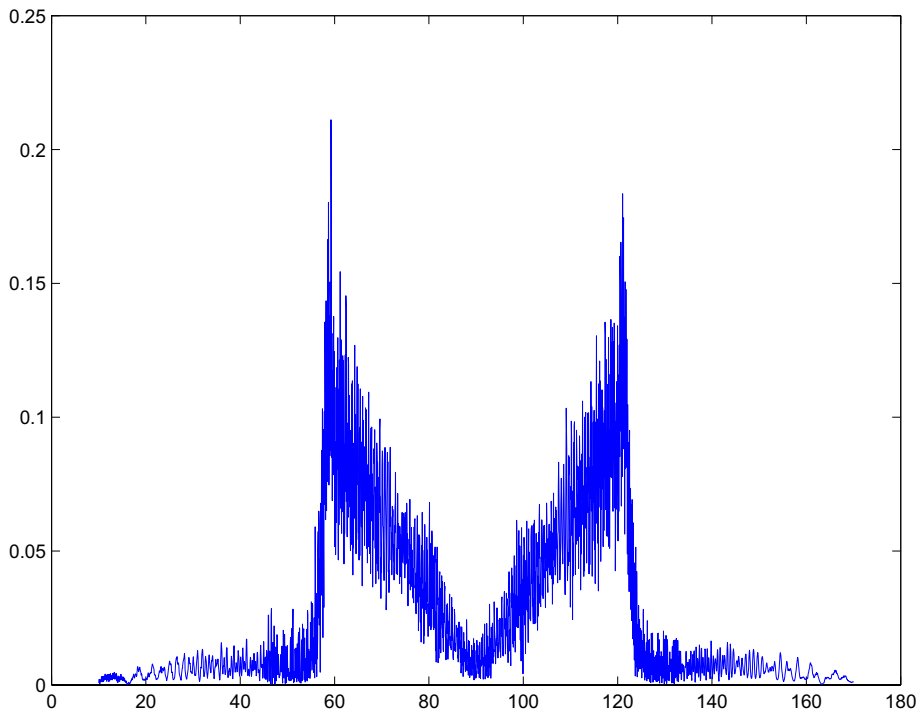

Fig. 6. This plot is described in the caption of Fig. 5.

The times required to invert the integral operator in (2.5) on a desktop PC with a 3 GHz Pentium IV processor and 3 Gb of RAM using the method described in Section 4 are reported in Table 4. In all experiments reported, the contour Γ_+ extends for half a wave-length above the x-axis, and for n_{wave} wave-lengths in the horizontal direction, where $10 \leqslant n_{\text{wave}} \leqslant 3000$. The "geometric" undulation of the wave pattern has about the same wavelength as the radiating field. The integral equation was discretized using the same techniques that were described in Section 5.2.

A benefit to using direct methods in the study of scattering problems, is that one frequently is interested in computing the scattered field for a large set of incoming wave patterns. In the direct method presented here, such calculations are very inexpensive once the approximate inverse has been calculated. As an example, we computed the inverse for the integral operator corresponding to a wavy surface such as the one depicted in Fig. 5, but 600 wave lengths across. We then computed the reflected field for a set of sources located at various positions on the semi-circle in Fig. 5 and computed the energy of the reflected field back at the source point. This resulted in a graph showing the amount of energy reflected back to the source as a function of location on the semi-circle, given in Fig. 6. In principle, the computation of the reflected energy for each source location requires the solution of a dense linear system involving 40,000 unknowns (in complex arithmetics). Since we had access to the inverse (which took about a minute to compute), each computation took less than one tenth of a second on a desktop PC.

6. Conclusions

In this paper, we have demonstrated that the integral operators associated with acoustic scattering from elongated scatterers in two dimensions can be inverted very rapidly. Numerical examples show that for a scatterer 3000 wavelengths long, the forward scattering problem can be solved in minutes on a desktop PC (at five digit accuracy). After the first solve, a compressed representation of the inverse of the integral operator is available, and additional forward scattering problems can be solved in less than one second each.

The techniques presented can with minor modifications be applied to electromagnetic scattering problems in both two and three dimensions, as long as the scatterer can be contained in a long, straight cylinder whose diameter as at most a few wave-lengths.

Acknowledgments

Appendix A. Proof of Theorem 2

The proof of Theorem 2 given in this appendix relies on the fact that any Hankel function H_n can be approximated by a sum of exponentials. The number of terms required to obtain an accuracy of ε on the interval $[n, n + L]$ scales as $\log L |\log \varepsilon|$. This result is stated in detail in Lemma 7. From Lemma 7, the special case of Theorem 2 in which the two boxes Ω_S and Ω_T have zero height follows immediately; we formulate it as Lemma 8. The proof that the results of Lemma 8 can be generalized to the environment of Theorem 2 is given at the end of this appendix.

Lemma 7 is a slight generalization of a result in [8].

Lemma 7. *Let L be any positive number larger than* 2, *let n be any integer, and let ε be a given precision such that* $0 < \varepsilon \leqslant 1/2$. *Then there exists an integer J, a number C, complex numbers* $\{\alpha_j, \beta_j\}_{j=1}^{J}$ *and non-negative real numbers* $\{t_j\}_{j=1}^{J}$ *with the following properties*: *the Hankel function H_n allows the expansion*

$$H_n(r) = \sum_{j=1}^{J} (\alpha_j e^{ir} + \beta_j) e^{-t_j r} + E(r), \tag{A.1}$$

where

$$|E(r)| \leqslant \varepsilon, \quad \text{for } n \leqslant r \leqslant n + L. \tag{A.2}$$

Moreover, the number C does not depend on either L, ε or n, and

$$J \leqslant C \log L |\log \varepsilon|. \tag{A.3}$$

Proof. Formula (9.1.22) of [1], says that

$$H_n(r) = \Psi_n(r) + \chi_n(r), \tag{A.4}$$

where

$$\Psi_n(r) = -\frac{1}{\pi} \int_0^\infty e^{-r \sinh t} (\sin(n\pi) e^{-nt} + i(e^{nt} + e^{-nt} \cos(n\pi))) dt, \tag{A.5}$$

and

$$\chi_n(r) = \frac{1}{\pi} \int_0^\pi e^{ir \sin \theta} e^{-in\theta} d\theta. \tag{A.6}$$

We convert the integral representation of Ψ_n to an expansion of the form (A.1) by simply estimating the integral by a quadrature rule supported on some points $t_1, \ldots, t_J$. To prove that the number of points J required in order to obtain an accuracy of ε is bounded as prescribed by (A.3), we note that (i) the integrand in the representation (A.5) decays exponentially (actually super-exponentially) and (ii) the derivative of the integrand at the origin is bounded by the condition that r satisfy $1 \leqslant r \leqslant 1 + L$, see [15] or [16].

It remains to derive an expansion for χ_n. To this end, we first effect the change of variables $t = \sin \theta$ in (A.6), whence

$$\chi_n(r) = \int_0^1 e^{irt} \chi_n(t) dt, \tag{A.7}$$

where

$$\chi_n(t) = \frac{1}{\pi \sqrt{1 - t^2}} \left(\left(\sqrt{1 - t^2} + it \right)^n + \left(-\sqrt{1 - t^2} + it \right)^n \right). \tag{A.8}$$

Since the integrand in (A.7) is holomorphic in the set $(0, 1) \times (0, \infty)$, we can change the path of integration from the horizontal line $\{t: 0 < t < 1\}$ to the vertical lines $\{it: 0 < t < \infty\}$, and $\{1 + it: 0 < t < \infty\}$. Taking into account the pole at $t = 1$, we then find that, for some number c,

$$\chi_n(r) = \int_0^\infty e^{-rt} \chi_n(it) dt - \int_0^\infty e^{ir-rt} \chi_n(1 + it) dt + c e^{ir}. \tag{A.9}$$

The integral representation (A.9) can now be converted to an expansion of the form (A.1) via the quadrature procedure already described for Ψ_n. $\square$

Lemma 8. *Let L be any positive number larger than 2, let n be any integer, let ε be a given precision such that $0 < \varepsilon \leqslant 1/2$, and consider the two sets $\Omega_S = \{(y_1, 0): -L \leqslant y_1 \leqslant 0\}$ and $\Omega_T = \{(x_1, 0): n \leqslant x_1 \leqslant n + L\}$. Then, there exists an integer J, a number C, and functions $\{\psi_j, \chi_j\}_{j=1}^J$ with the following properties: The Hankel function H_n allows the expansion*

$$H_n(|x - y|) = \sum_{j=1}^J \psi_j(x) \chi_j(y) + E(x, y), \quad \text{for } x \in \Omega_T, \ y \in \Omega_S, \tag{A.10}$$

where

$$J \leqslant C \log L |\log \varepsilon|, \tag{A.11}$$

and

$$|E(x,y)| \leqslant \varepsilon. \tag{A.12}$$

Moreover, the number C does not depend on either L, ε or n.

Proof. In the geometry given, $H_n(|x - y|) = H_n(x_1 - y_1)$, where $n \leqslant x_1 - y_1 \leqslant n + 2L$. Thus, the formula (A.1), which expresses H_n in terms of exponentials only, applies. The factorization (A.10) follows immediately. □

Proof of Theorem 2. Since k is simply a scaling parameter for the geometry of the problem, we can for the purposes of the proof set $k = 1$.

We first construct an expansion of the form (3.2) that is valid in the special case where x lies on the bottom boundary of Ω_T. Picking an arbitrary $y = (y_1, y_2) \in \Omega_S$ we set $\hat{y} = (y_1, -W/2)$, as shown in Fig. 1, and apply the Graf addition theorem (formula (9.1.79) of [1]) to obtain

$$H_0(|x - y|) = \sum_{n=-\infty}^{\infty} \mathrm{i}^n H_n(|x - \hat{y}|) J_n(|\hat{y} - y|). \tag{A.13}$$

Since $|\hat{y} - y| \leqslant W$, this series converges very fast; to be precise, we can for any ε find an integer N such that $N \sim |\log \varepsilon|$, and

$$\left| H_0(|x - y|) - \sum_{n=-N}^{N} \mathrm{i}^n H_n(|x - \hat{y}|) J_n(|\hat{y} - y|) \right| \leqslant \varepsilon \tag{A.14}$$

for any $y \in \Omega_S$, and any x on the bottom of Ω_T. Provided that $S \geqslant N$, Lemma 8 then applies to each term $H_n(|x - \hat{y}|)$ in (A.14), resulting in the desired expansion. The condition (3.1) follows since $N \sim |\log \varepsilon|$. (See Remark 1 for some notes on the truncation of the sum (A.13).)

Next we note that similar expansions can be constructed for x on all boundaries of Ω_T (the top boundary is of course entirely analogous to the bottom one, and the corresponding proof for the vertical boundaries is trivial). Similarly, it can be shown that there exist expansions of the desired form for the normal derivative of u on all boundaries of Ω_T.

Finally, we note that when we know the value of u and its normal derivative on the boundary of Ω_T, the value at any point $x \in \Omega_T$ can be constructed via Green's formula:

$$u(x) = \frac{1}{4\mathrm{i}} \int_{\partial \Omega_T} \left(\left(\frac{\partial}{\partial n(z)} H_0(|x - z|) \right) u(z) - H_0(|x - z|) \frac{\partial}{\partial n(z)} u(z) \right) \mathrm{d}s(z), \tag{A.15}$$

valid for any function u satisfying $-\Delta u - u = 0$ on Ω_T. □

References

[1] Milton Abramowitz, Irene A. Stegun (Eds.), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications Inc., New York, 1992. Reprint of the 1972 edition.

[2] K.E. Atkinson, The Numerical Solution of Integral Equations of the Second Kind, Cambridge University Press, 1997.

[3] O.M. Bucci, G. Franceschetti, On the degrees of freedom of scattered fields, IEEE Trans. AP-37 (1989) 918–926.

[4] S. Chandrasekaran, M. Gu, Fast and stable algorithms for banded plus semiseparable systems of linear equations, SIAM J. Matrix Anal. Appl. 25 (2) (2003) 373–384.

[5] Y. Chen, A fast direct algorithm for the Lippmann–Schwinger integral equation in two dimensions, Adv. Comp. Math. 16 (2) (2002) 175–190.

[6] H. Cheng, Z. Gimbutas, P.G. Martinsson, V. Rokhlin, On the compression of low rank matrices, SIAM J. Sci. Comp. 26 (4) (2005) 1389–1404.

[7] W.C. Chew, J.-M. Jin, E. Michielssen, J. Song, Fast and Efficient Algorithms in Computational Electromagnetics, Artech House, 2001.

[8] W.C. Chew, C.-C. Lu, A fast algorithm to compute the wave-scattering solution of a large strip, J. Comput. Phys. 107 (2) (1993) 378–387.

[9] G.T.D. Francia, Directivity, super-gain and information, IRE Trans. Anten. Propag. 4 (1956) 473–478.

[10] Lars Grasedyck, Wolfgang Hackbusch, Construction and arithmetics of $\mathscr{H}$-matrices, Computing 70 (4) (2003) 295–334.

[11] S. Kapur, V. Rokhlin, High-order corrected trapezoidal quadrature rules for singular functions, SIAM J. Numer. Anal. 34 (4) (1997) 1331–1356.

[12] P.G. Martinsson, V. Rokhlin, A fast direct solver for boundary integral equations in two dimensions, J. Comp. Phys. 205 (1) (2005) 1–23.
[13] P.G. Martinsson, V. Rokhlin, M. Tygert, On interpolation and integration in finite-dimensional spaces of bounded functions, Technical Report RR-1317, Department of Computer Science, Yale, 2005.
[14] E. Michielssen, A. Boag, W.C. Chew, Scattering from elongated objects: direct solution in $O(N\log^2 N)$ operations, IEE Proc. Microw. Anten. Propag. 143 (4) (1996) 277–283.
[15] V. Rokhlin, A fast algorithm for the discrete Laplace transformation, J. Complexity 4 (1) (1988) 12–32.
[16] N. Yarvin, V. Rokhlin, Generalized Gaussian quadratures and singular value decompositions of integral operators, SIAM J. Sci. Comput. 20 (2) (1998) 699–718.