Rapid solution of the wave equation in unbounded domains

Banjai, L; Sauter, S
RAPID SOLUTION OF THE WAVE EQUATION IN UNBOUNDED DOMAINS

L. BANJAI† AND S. SAUTER†

Abstract. In this paper we propose and analyze a new, fast method for the numerical solution of time domain boundary integral formulations of the wave equation. We employ Lubich’s convolution quadrature method for the time discretization and a Galerkin boundary element method for the spatial discretization. The coefficient matrix of the arising system of linear equations is a triangular block Toeplitz matrix. Possible choices for solving the linear system arising from the above discretization include the use of fast Fourier transform (FFT) techniques and the use of data-sparse approximations. By using FFT techniques, the computational complexity can be reduced substantially while the storage cost remains unchanged and is, typically, high. Using data-sparse approximations, the gain is reversed; i.e., the computational cost is (approximately) unchanged while the storage cost is substantially reduced. The method proposed in this paper combines the advantages of these two approaches. First, the discrete convolution (related to the block Toeplitz system) is transformed into the (discrete) Fourier image, thereby arriving at a decoupled system of discretized Helmholtz equations with complex wave numbers. A fast data-sparse (e.g., fast multipole or panel-clustering) method can then be applied to the transformed system. Additionally, significant savings can be achieved if the boundary data are smooth and time-limited. In this case the right-hand sides of many of the Helmholtz problems are almost zero, and hence can be disregarded. Finally, the proposed method is inherently parallel. We analyze the stability and convergence of these methods, thereby deriving the choice of parameters that preserves the convergence rates of the unperturbed convolution quadrature. We also present numerical results which illustrate the predicted convergence behavior.

Key words. wave equation, boundary element methods, convolution quadrature

AMS subject classifications. 65N38, 65R20, 35L05

DOI. 10.1137/070690754

1. Introduction. Boundary value problems governed by the wave equation

\[ \partial^2_t u - \Delta u = f \]

arise in many physical applications such as electromagnetic wave propagation or the computation of transient acoustic waves. Since such problems are typically formulated in unbounded domains, the method of integral equations is an elegant tool for transforming this partial differential equation (PDE) into an integral equation on the bounded surface of the scatterer.

Although this approach goes back to the early 1960s (cf. [19]), the development of fast numerical methods for integral equations in the field of hyperbolic problems is still in its infancy compared to the multitude of fast methods for elliptic boundary integral equations (cf. [38] and references therein). Existing numerical discretization methods include collocation methods with some stabilization techniques (cf. [7], [8], [14], [15], [16], [33], [37]), and Laplace-Fourier methods coupled with Galerkin boundary elements in space (see [3], [12], [17], [20]). Numerical experiments can be found, e.g., in [21].

*Received by the editors May 7, 2008; accepted for publication (in revised form) June 17, 2008; published electronically October 31, 2008. A short description of the results of this paper has been published in the Proceedings of the Waves 2007 Conference [6].

†Institut für Mathematik, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland (lehel.banjai@math.uzh.ch, stas@math.uzh.ch). The first author gratefully acknowledges the support given by SNF 200021-113481/1.
In [18] a fast version of the marching-on-in-time (MOT) method is presented which is based on a suitable plane wave expansion of the arising potential, which reduces the storage and computational costs.

We here employ the convolution quadrature method for the time discretization and a Galerkin boundary element method in space. The convolution quadrature method for the time discretization has been developed in [29], [30], [31], [32]. It provides a straightforward way to obtain a stable time stepping scheme using the Laplace transform of the kernel function. For applications to problems such as viscoelastic and poroelastic continua see [40], [41], [42].

The coefficient matrix in the arising system of linear equation is a block-triangular Toeplitz matrix consisting of $N$ blocks of dimension $M \times M$, where $N$ denotes the number of time steps and $M$ is the number of spatial degrees of freedom. Due to the nonlocalness of the arising boundary integral operators, the $M \times M$ matrix blocks are densely populated.

In the literature, there exist (at least) two alternatives for solving this system efficiently. In [24], fast Fourier transform (FFT) techniques are employed, which make use of the Toeplitz structure of the system matrix, and the computational complexity is reduced to $O(M^2 N \log^2 N)$, while the storage complexity stays at $O(N M^2)$. In [23], [22], [28], the $M \times M$ block matrices are approximated by data-sparse representations based on a cutoff and panel-clustering strategy. This leads to a significant reduction of the storage complexity. The computational complexity is reduced compared to the $O(N^2 M^2)$ cost of the naive approach but increased compared to the computational cost of the FFT approach.

Also the classical Galerkin discretization of the retarded boundary integral equation (see [3], [20]), leads to a block Toeplitz system matrix, where the matrix blocks are of size $M \times M$ and sparse. More precisely, the number of nonzero entries in the block Toeplitz matrix is, for piecewise constant boundary elements, of order $O(M^2)$ and, for piecewise linear boundary elements, of order $O(M^2 + \delta)$ for this approach. Here, the total cost for the computation of a full Galerkin approximation sums up to $O(M^2 N)$ for piecewise constant boundary elements and to $O(N^2 M^{3/2})$ for piecewise linear boundary elements. A drawback of this approach, however, is that the numerical quadrature for computing the coefficients of the system matrix has to be carried out on the intersections of the boundary element mesh with the discrete light cone. The stable handling of these intersections and the implementation is especially complicated for curved panels.

In this paper, we propose a new approach which combines the advantages of the FFT technique with the sparse approximation. We transfer the block Toeplitz system to the Fourier image by the discrete Fourier transform and then face the problem of computing approximate solutions of Helmholtz problems at different (complex) wave numbers. These Helmholtz problems are fully decoupled, and hence can be efficiently solved on parallel computers. Relatively standard, fast methods (e.g., fast multipole method, hierarchical matrices) for the solution of frequency domain scattering can effectively be applied to these problems; see [9], [35], and [5]. It may also be possible to further reduce the computational cost of assembling the matrices by using the techniques for multifrequency analysis described in [27], [44]. Further, we also show that if the boundary data are sufficiently smooth and compatible and of limited time duration, instead of $N$, only $O(N^\epsilon)$, for any fixed $\epsilon > 0$, Helmholtz systems need to be solved. Our method is similar and shares some properties (the need to solve a series of elliptic problems and the intrinsic parallelizability) of certain methods for
parabolic equations; see [26], [43]. A related, interesting variation of the convolution quadrature for convolution kernels whose Laplace transform is sectorial can be found in [39].

2. Integral formulation of the wave equation. Let \( \Omega \subset \mathbb{R}^3 \) be a Lipschitz domain with boundary \( \Gamma \); typically, e.g., in scattering problems, \( \Omega \) is an unbounded domain. In this paper, we present efficient methods for numerically solving the homogeneous wave equation

\[
\partial_t^2 u - \Delta u = 0 \quad \text{in } \Omega \times (0, T)
\]

with initial conditions

\[
u(\cdot, 0) = \partial_t u(\cdot, 0) = 0 \quad \text{in } \Omega
\]

and boundary conditions

\[
u = g \quad \text{on } \Gamma \times (0, T)
\]

on a time interval \( (0, T) \) for some \( T > 0 \). For its solution, we employ an ansatz as a single layer potential

\[
u(x, t) = \int_0^t \int_\Gamma k(x - y, t - \tau) \phi(y, \tau) d\Gamma_y d\tau, \quad (x, t) \in \Omega \times (0, T),
\]

where \( k(z, t) \) is the fundamental solution of the wave equation,

\[
k(z, t) = \frac{\delta(t - \|z\|)}{4\pi \|z\|},
\]

with \( \delta(t) \) being the Dirac delta distribution. The ansatz (2.2) satisfies the homogeneous equation (2.1a) and the initial conditions (2.1b). The extension \( x \to \Gamma \) is continuous, and hence the unknown density \( \phi \) in (2.2) is determined via the boundary conditions (2.1c), \( u(x, t) = g(x, t) \). This results in the boundary integral equation for \( \phi \),

\[
\int_0^t \int_\Gamma k(x - y, t - \tau) \phi(y, \tau) d\Gamma_y d\tau = g(x, t) \quad \forall (x, t) \in \Gamma \times (0, T).
\]

Existence and uniqueness results for the solution of the continuous problem are proved in [31] and [3, Prop. 3].


3.1. Time discretization via convolution quadrature. For the time discretization, we employ the convolution quadrature approach which has been developed by Lubich in [29], [30], [31] and Lubich and Schneider in [32]. We do not recall the theoretical framework here but directly apply the approach to the wave equation. We make use of the following notation for the time convolution:

\[
V(\partial_t)\phi := \int_0^t v(t - \tau)\phi(\tau)d\tau,
\]

where \( V \) denotes the Laplace transform of the operator \( v \); for the reasons behind using this notation see [29]. Note that, for the retarded single layer potential (2.2),
\( v \) is a parameter-dependent integral operator, i.e., \((v(t - \tau) \phi(\tau))(x) = \int_\Omega k(x - y, t - \tau) \phi(\tau, y) d\Gamma(y)\) (where we write \( \phi(\tau, y) \) for \((\phi(\tau))(y)\)) and \( V(s) \) is the Laplace transform of \( v \) given by (3.4).

To discretize the time convolution we split the time interval \([0, T]\) into \(N+1\) time steps of equal length \(\Delta t = T/N\) and compute an approximate solution at the discrete time levels \(t_n = n\Delta t\). The continuous convolution operator \( V(\partial_t) \) at the discrete times \(t_n\) is replaced by the discrete convolution operator, for \(n = 0, 1, \ldots, N\),

\[
(V(\partial_t^{\Delta t})\phi^{\Delta t})(t_n) := \sum_{j=0}^{n} \omega^{\Delta t}_{n-j}(V)\phi^{\Delta t}(t_j).
\]

The convolution weights \(\omega^{\Delta t}_n(V)\) are defined below (see (3.3)): whenever the underlying operator \(v\), respectively, \(V\), is clear from the context, we will write \(\omega^{\Delta t}_n\). The time-discrete problem is given as follows: Find \(\phi_j(\cdot) = \phi^{\Delta t}(\cdot, t_j)\), such that

\[
\sum_{j=0}^{n} (\omega^{\Delta t}_{n-j}\phi_j)(x) = g_n(x), \quad n = 1, \ldots, N, \quad x \in \Gamma,
\]

where \(g_n(x)\) is some approximation to \(g(x, t_n)\), or \(g(x, t_n)\) itself.

For the derivation, the general framework, and various applications, we refer the reader to [29], [30], [31], and for our concrete problem to [23]. If the time discretization is related to the unconditionally stable backward difference formula of second order (BDF2) scheme, the convolution weights \(\omega^{\Delta t}_n\) are implicitly defined by

\[
V \left( \frac{\gamma(\zeta)}{\Delta t} \right) = \sum_{n=0}^{\infty} \omega^{\Delta t}_n \zeta^n, \quad |\zeta| < 1.
\]

Here, \(V(s) : H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)\), \(\text{Re } s > 0\), is the single layer potential for the Helmholtz operator \(\Delta U - s^2 U = 0\),

\[
(V(s)\varphi)(x) = \int_{\Gamma} K(||x - y||, s)\varphi(y) d\Gamma_y, \quad \text{where } K(d, s) := \frac{e^{-sd}}{4\pi d}.
\]

Note that \(K\) is the Laplace transform of the original time domain kernel function (2.3). The function \(\gamma(\zeta)\) is the quotient of the generating polynomials of the BDF2 scheme and is given by

\[
\gamma(\zeta) = \frac{1}{2} (\zeta^2 - 4\zeta + 3).
\]

### 3.2. A decoupled system of Helmholtz problems

As recommended in [29, 31], the convolution weights \(\omega^{\Delta t}_n\) can be numerically computed by applying the trapezoidal rule to its representation as a contour integral,

\[
\omega^{\Delta t}_j(V) = \frac{1}{2\pi i} \int_C \frac{V(\gamma(\zeta)/\Delta t)}{\zeta^{j+1}} d\zeta,
\]

where \(C\) can be chosen as a circle centered at the origin of radius \(\lambda < 1\). The approximate convolution weights are then given by the scaled inverse discrete Fourier transform

\[
\omega^{\Delta t, \lambda}_j(V) := \frac{\lambda^{-j}}{N+1} \sum_{l=0}^{N} V(s_l) e^{i j \frac{s_l \zeta_{N+1}}{\Delta t}}, \quad \text{where } \zeta_{N+1} = e^{\frac{2\pi i}{N+1}}, \quad s_l = \frac{\gamma(\lambda \zeta_{N+1})}{\Delta t}.
\]
Let us extend the above two formulae to negative indices \( j < 0 \); note that this implies \( \omega_j^\Delta t = 0 \) for \( j < 0 \). As \( N \to \infty \) or \( \lambda \to 0 \), we have \( \omega_j^\Delta t - \omega_j^{\Delta t, \lambda} = O(\lambda^{N+1}) \), \( j = -N, \ldots, N \); see Proposition 5.4. By extending the sum in (3.1) to \( j = N \) and substituting the approximate weights in (3.2), we obtain the following new system of equations for the new unknown \( \phi^{\Delta t, \lambda} \):

\[
(3.6) \quad \left( V(\partial_t^{\Delta t, \lambda}) \phi^{\Delta t, \lambda} \right)(t_n) := \sum_{j=0}^{N} \omega_{n-j}^{\Delta t, \lambda}(V) \phi_j^\lambda = g_n, \quad n = 0, 1, \ldots, N.
\]

The effect of the approximation on the difference between \( \phi^{\Delta t, \lambda} \) and \( \phi^{\Delta t} \) is discussed later. Substituting the definition of \( \omega_{n-j}^{\Delta t, \lambda}(V) \) in (3.6), we obtain the system of equations

\[
(3.7) \quad \frac{\lambda^{-n}}{N+1} \sum_{l=0}^{N} \left( V(s_l) \hat{\phi}_l \right) (x) \zeta_N^{nl} = g_n(x), \quad n = 0, 1, \ldots, N,
\]

where

\[
\hat{\phi}_l := \sum_{j=0}^{N} \lambda^j \phi^\lambda_j \zeta_N^{-lj}.
\]

Note that the inverse transform is given by

\[
(3.8) \quad \phi^\lambda_n = \frac{\lambda^{-l}}{N+1} \sum_{j=0}^{N} \hat{\phi}_l \zeta_N^{lj}.
\]

Now, notice that, after multiplying by \( \lambda^n \), applying the discrete Fourier transform with respect to \( n \) to both sides gives \( N+1 \) decoupled problems as follows:

\[
(3.9) \quad \left( V(s_l) \hat{\phi}_l \right) (x) = \hat{g}_l(x) \quad \forall x \in \Gamma,
\]

where

\[
\hat{g}_l(x) = \sum_{n=0}^{N} \lambda^n g_n(x) \zeta_N^{-ln}.
\]

We have thereby reduced the problem of solving numerically the wave equation to a system of Helmholtz problems with complex wave numbers \( s_l, l = 0, 1, \ldots, N \). An example of the range of frequencies is given in Figure 1.

**Remark 3.1.** An important remark to make here is that \( V(\partial_t^{\Delta t, \lambda}) \phi^{\Delta t, \lambda} = g \) implies \( \phi^{\Delta t, \lambda} = V^{-1} \left( \partial_t^{\Delta t, \lambda} \right) g \).

This can be seen by applying the scaled discrete inverse Fourier transform (see (3.8)) to

\[
\hat{\phi}_l = V^{-1}(s_l) \hat{g}_l,
\]

thereby obtaining

\[
\phi^\lambda_n = \frac{\lambda^{-n}}{N+1} \sum_{l=0}^{N} \hat{\phi}_l \zeta_N^{nl} = \frac{\lambda^{-n}}{N+1} \sum_{l=0}^{N} V^{-1}(s_l) \hat{g}_l \zeta_N^{nl} = \sum_{j=0}^{N} \omega_{n-j}^{\Delta t, \lambda}(V^{-1}) g_j.
\]

The last step is obtained from the definition of \( \hat{g}_l \) and \( \omega_{n-j}^{\Delta t, \lambda}(V^{-1}) \); see also (3.6) and (3.7). This fact will help us in obtaining optimal error and stability estimates.
3.3. Spatial discretization. Galerkin boundary element methods. In the previous section we derived the following semidiscrete problem: For $n = 0, 1, \ldots, N$, find $\phi_n^\lambda \in H^{-1/2}(\Gamma)$ such that

\begin{equation}
\sum_{j=0}^{N} \omega_{n-j}^\lambda \phi_j^\lambda = \hat{g}_n, \quad n = 0, 1, \ldots, N.
\end{equation}

We have further shown that the above system is equivalent to a system of decoupled Helmholtz equations

\begin{equation}
(V(s_l)\hat{\phi}_l)(x) = \hat{g}_l(x) \quad \forall x \in \Gamma.
\end{equation}

In this paper we use a Galerkin boundary element method for the spatial discretization. Let $\mathcal{G}$ be a regular (in the sense of Ciarlet [11]) boundary element mesh on $\Gamma$ consisting of shape regular, possibly curved, triangles. For a triangle $\tau \in \mathcal{G}$, the (regular) pull-back to the reference triangle $\hat{\tau} := \text{conv}\{[0,0], [1,0], [0,1]\}$ is denoted by $\chi_\tau : \hat{\tau} \rightarrow \tau$. The space of piecewise constant, discontinuous functions is $S_{-1,0} := \{u \in L^\infty(\Gamma) : \forall \tau \in \mathcal{G} : u|_\tau \in P_0\}$, and, alternatively, we consider the space of continuous, piecewise linear functions $S_{0,1} := \{u \in C^0(\Gamma) : \forall \tau \in \mathcal{G} : (u \circ \chi_\tau)|_\tau \in P_1\}$

for the space discretization. As a basis for $S_{-1,0}$, we choose the characteristic functions for the panels $\tau \in \mathcal{G}$, while the basis for $S_{0,1}$ consists of the standard hat functions, lifted to the surface $\Gamma$. The general notation is $S$ for the boundary element space and $(b_m)_{m=1}^M$ for the basis. The mesh width is given by

\[ h := \max_{\tau \in \mathcal{G}} h_\tau, \quad \text{where} \quad h_\tau := \text{diam} (\tau). \]
RAPID SOLUTION OF THE WAVE EQUATION 233

For the space-time discrete solution at time $t_n$ we employ the ansatz

$$(3.12) \quad \phi_n^{h,\lambda}(y) = \sum_{m=1}^{M} \phi_{n,m} b_m(y),$$

where $(\phi_{n,m})_{m=1}^{M} \in \mathbb{R}^M$ are the nodal values of the discrete solution at time step $t_n$. Therefore, for the Helmholtz problems (3.11), the corresponding ansatz is

$$(3.13) \quad \hat{\phi}^h_l(y) = \sum_{m=1}^{M} \hat{\phi}_{l,m} b_m(y),$$

where the relationship between $\hat{\phi}_{l,m}$ and $\phi_{n,m}$ is given by $\hat{\phi}_{l,m} = \sum_{n=0}^{N} \lambda^n \phi_{n,m} s_{N+1}^l$.

To solve for the coefficients $\hat{\phi}_{l,m}$ we impose the integral equations (3.11) not pointwise but in a weak form as follows: Find $\hat{\phi}^h_l \in S$ of the form (3.13) such that

$$(3.14) \quad \sum_{m=1}^{M} \hat{\phi}_{l,m} \int_{\Gamma} \int_{\Gamma} K(||x-y||, st) b_m(y)b_k(x) d\Gamma_y d\Gamma_x = \int_{\Gamma} \hat{g}_l(x)b_k(x) d\Gamma_x,$$

for $l = 0, 1, \ldots, N, k = 1, 2, \ldots, M$. Note that this is equivalent to imposing (3.10) in a weak form in order to compute $\phi_n^{h,\lambda}$.

4. Algorithmic realization and sparse approximation. Applying the Galerkin boundary element method to the time-discrete equations (3.1) obtained by convolution quadrature results in a block-triangular, block Toeplitz system, where each block is a dense Galerkin boundary element matrix; see [31] and [22]. This block system can be solved by using FFT techniques (see [24]), with computational complexity of $O(M^2 N \log^2 N)$ and a storage complexity of $O(M^2 N)$. Alternatively (see [28]), one can approximate the block matrices $A_n$ by a cutoff strategy and panel-clustering and directly solve the system without the FFT. This reduces the storage cost significantly, while the computational complexity is $O(M^2 N^{1+s})$, where the small value of $s$ depends on the chosen discretization. By rewriting (3.1) as a system of decoupled Helmholtz problems, we are able to combine the advantages of both approaches.

We note that also the classical Galerkin discretization of the retarded boundary integral equation leads to a block Toeplitz system. Solving this system (see [3], [20]) nevertheless results in suboptimal, higher than linear, computational complexity.

4.1. Reduction of the number of Helmholtz problems to be solved. A closer look at the Helmholtz problems tells us that only half of the problems need to be solved. Since $\hat{\phi}_l$, $\hat{g}_l$, and $s_l$ are discrete Fourier transforms of real data, we know that they are, for $l = 1, 2, \ldots, \lfloor \frac{N}{2} + 1 \rfloor$, the complex conjugates of $\hat{\phi}_j$, $\hat{g}_j$, $s_j$, for $j = \lfloor \frac{N}{2} + 2 \rfloor, \ldots, N + 1$; for the case of $s_l$ see Figure 1. Most importantly, for us this means that

$$(4.1) \quad \hat{\phi}_{N+2-j} = \overline{\hat{\phi}_j}, \quad j = 1, 2, \ldots, \left\lfloor \frac{N}{2} + 1 \right\rfloor.$$

Depending on the properties of the right-hand side $g$, it is possible to avoid the solution of a much larger number of Helmholtz problems without destroying the accuracy of the overall approximation. A particularly favorable case arises if $g$ as a function of time can be extended to $\mathbb{R}$ as a smooth function with support contained in $[0,T]$. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
We plot \( \max_{\|x\|=1} |\hat{g}_n(x)| \) for \( N = 256, T = 2, \lambda^N = 10^{-8} \), and where \( g(x,t) \) is the Gaussian pulse given by (6.5). The solution to the \( n \)th Helmholtz problem, with \( n \) in the central plateau in the above plot, is accurately approximated by zero.

Let us assume that for some \( x \in \Gamma, g(x,\cdot) \in C^\infty([0,T]) \), and that
\[
\partial^n_t g(x,0) = \partial^n_t g(x,T) = 0 \quad \forall n \in \mathbb{N}_0.
\]
Further, define \( g_\lambda(x,t) := \lambda^{t/\Delta t} g(x,t) \). Then it is clear that \( g_\lambda(x,\cdot) \in C^\infty([0,T]) \) and that also all the partial derivatives with respect to time vanish at the end points of the time interval \([0,T]\). The reason for defining this function is that \( \hat{g}_n(x) \) is an approximation of a Fourier coefficient of \( g_\lambda(x,t) \), as we see next.

Let \( g_\lambda(x,\cdot) \) be extended to the domain \([0,T+\Delta t]\) by zero (i.e., in a smooth way) and further extended to \( \mathbb{R} \) in a periodic way with period \( T+\Delta t \). Let then
\[
g_\lambda(x,t) = \sum_{j=-\infty}^{\infty} a_j e^{\frac{2\pi i j t}{T+\Delta t}}, \quad a_j = \frac{1}{T+\Delta t} \int_0^{T+\Delta t} g_\lambda(x,\tau) e^{-\frac{2\pi i j \tau}{T+\Delta t}} \, d\tau
\]
be its Fourier expansion. Approximating the integral in the definition of the coefficients \( a_j \) by the trapezoidal rule, we obtain exactly the values \( \frac{1}{N+1} \hat{g}_j(x) \), where, assuming \( N \) is even,
\[
a_j \approx \frac{1}{N+1} \sum_{n=0}^{N} g_\lambda(x,t_n) e^{-\frac{2\pi i j n}{N+1}} = \frac{1}{N+1} \hat{g}_j(x) \quad \text{for } 0 \leq j \leq \frac{N}{2}.
\]
See Figure 2 for an example of a right-hand side with the above properties and the decay of its Fourier coefficients. The solutions of Helmholtz problems with right-hand sides that are close to zero (i.e., all the right-hand sides on the central plateau in Figure 2) can be set to zero with no adverse affect on the accuracy of the overall method.
Remark 4.1. A right-hand side \( g \) with the above properties can be thought of as a smooth signal of finite durability. If \( g \) does not have these properties, it may still be possible to split the signal into a number of smooth and time-limited signals.

4.2. Data-sparse approximation. To find a solution to (3.9) we need to solve a number of dense linear systems, each of size \( M \times M \). The cost of solving a single system by a direct method is \( O(M^3) \), and if a good preconditioner for an iterative method is available, this can be reduced to \( O(M^2) \). In both cases the storage costs are \( O(M^2) \). The cost of recovering the values \( \phi_{j,m} \) from \( \hat{\phi}_{l,m} \) is negligible since it can be done exactly (if we ignore errors due to finite precision arithmetic) and efficiently using the FFT in time \( O(MN \log N) \); see also Remark 5.11.

One possibility for reducing these costs is to use panel-clustering or fast multipole techniques. We explain the basic idea behind these methods.

Let \( A_n \) be the \( n \)th linear system to be solved in (3.9), i.e.,

\[
(A_n)_{kj} = \int_{\Gamma} \int_{\Gamma} K(\|x - y\|, s_n) b_j(y) b_k(x) d\Gamma_y d\Gamma_x.
\]

Further, we denote by \( I \) the index set \( I := \{1, 2, \ldots, M\} \), refer to subsets \( \tau \subset I \) as clusters, and define corresponding subsets of the boundary \( \Gamma \) by

\[
\Gamma_\tau := \bigcup_{j\in\tau} \text{supp} b_j.
\]

We call a pair of clusters \( \tau \times \sigma \) a block. The corresponding block of the matrix \( A_n \) is then given by

\[
(A_n|_{\tau \times \sigma})_{kj} = \begin{cases} (A_n)_{kj} & \text{if } k \in \tau \text{ and } j \in \sigma, \\ 0 & \text{otherwise}. \end{cases}
\]

In the following definition, \( B(c, r) \) denotes the ball centered at \( c \in \mathbb{R}^3 \) and radius \( r > 0 \).

**Definition 4.2.** A block \( b = \tau \times \sigma \) is said to be \( \eta \)-admissible, for some \( \eta < 1 \), if there exist \( r_\tau, r_\sigma > 0 \) and \( c_\tau, c_\sigma \in \mathbb{R}^3 \) such that

\[
r_\tau + r_\sigma \leq \eta \| c_\tau - c_\sigma \| \quad \text{and} \quad \Gamma_\tau \subset B(c_\tau, r_\tau), \quad \Gamma_\sigma \subset B(c_\sigma, r_\sigma).
\]

For an admissible block, our goal is to find a separable approximation of the following fundamental solution:

\[
K(\|x - y\|, s) \approx \sum_{l,k=1} L u^l_k(x) s^\tau_{ki} v^\sigma_l(y), \quad x \in \Gamma_\tau, \; y \in \Gamma_\sigma.
\]

As indicated by the notation, we require that the basis functions \( u^l_k(\cdot) \) (respectively, \( v^\sigma_l(\cdot) \)) depend only on the cluster \( \tau \) (respectively, \( \sigma \)), and that the coefficients \( s^\tau_{ki} \) depend only on the block cluster \( b = \tau \times \sigma \). Such an expansion allows us to approximate the block \( A_n|_{\tau \times \sigma} \) of the matrix by a low rank matrix as follows:

\[
A_n|_{\tau \times \sigma} \approx UV^T,
\]
where
\[(U)_{kl} := \begin{cases} \int_{\Gamma} u^\tau(x)b_k(x)d\Gamma_x & \text{if } k \in \tau, \ l = 1, \ldots, L, \\ 0 & \text{otherwise,} \end{cases} \]

and \[(V)_{jl} := \begin{cases} \int_{\Gamma} v^\sigma(y)b_j(y)d\Gamma_y & \text{if } j \in \sigma, \ l = 1, \ldots, L, \\ 0 & \text{otherwise.} \end{cases} \]

and \[(S)_{lm} := s^{\tau,\sigma}_{lm} . \]

Note that for \(A_n|\tau \times \sigma\) we need \(O(|\tau||\sigma|)\) amount of storage, whereas for \(USV^T\) we need \(O(|\tau|L + |\sigma|L)\). If \(L \ll \max\{|\tau|, |\sigma|\}\), it is significantly advantageous to use the low rank approximation of the block.

An extensive literature exists on the use of these methods to speed up the solution of the Helmholtz integral equations discretized by Galerkin boundary elements [2], [5], [13], [35], [36]. Most of this literature is, however, focused on the Helmholtz problem with a purely real wave number. For a purely real wave number the single layer potential representation is not always invertible; therefore certain stabilization methods need to be used. In our case the imaginary part of the wave number is strictly positive and we can use the single layer representation. The details of applying these “fast” methods to our case, together with algorithms and complexity estimates, will be given in a forthcoming paper. Here we investigate the effect of perturbations, due to the application of the fast methods, on the stability and accuracy. We assume that the kernel function \(K(\cdot, s_l)\) in (3.9) is replaced by a separable approximation \(K^{pc}(\cdot, s_l)\) such that
\[\left|K(d, s_l) - K^{pc}(d, s_l)\right| \leq \frac{\delta}{d} \text{ for some } \delta > 0.\]

The solution of the resulting perturbed system is denoted by \(\phi^{pc}_{lm}\). To obtain a uniform approximation (4.6), the length of expansion \(L\) needs to depend both on the block cluster \(b = \tau \times \sigma\) and on \(s_l\). Typically \(L\) is chosen so that
\[L \geq C \left(\text{Re } s_l \|c_\tau - c_\sigma\| + \log \frac{1}{\delta}\right)^{d-1},\]

where \(C\) depends on the admissibility parameter \(\eta\), and \(d = 2, 3\) is the space dimension. Explicit and sharp estimates on the optimal choice of \(L\) are difficult to obtain, especially for complex wave numbers. In practice, one would estimate the error by a product of a Bessel function and a Hankel function; see, e.g., [1], [9]. Nevertheless, an important observation that can be made is that once \(L\) is greater than some threshold, the threshold depending on \(s_l\), the convergence is exponential. This means that high accuracy can be obtained at little extra cost.

5. Error analysis. In the previous section we have introduced a method to reduce the numerical solution of the wave equation to a system of Helmholtz problems. We have also described two ways of reducing the cost of solving these systems by introducing further approximations. In this section we investigate the stability and convergence of both the basic method and the further approximations. This allows us to adjust the control parameters of these methods to the required accuracy in an optimal way.

Let the approximation to the unknown density \(\phi(x, t_n)\) obtained by the pure Lubich’s method, i.e., with exact convolution weights, be given by \(\phi^h_n \in S\). In [31] it is proved that if the data \(g\) are sufficiently smooth and compatible, then
\[\|\phi^h_n(\cdot) - \phi(\cdot, t_n)\|_{H^{-1/2}(\Gamma)} \leq C(\Delta t^2 + h^{m+3/2}),\]
where \( m = 0 \) for a piecewise constant basis and \( m = 1 \) for a piecewise linear basis. By “smooth and compatible” we mean that \( g \in H^m_0([0, T]; H^{1/2}(\Gamma)) \), where

\[
H^m_0([0, T]; H^{1/2}(\Gamma)) := \left\{ g : \Gamma \times [0, T] \to \mathbb{R} : \text{there exists } g^* \in H^m(\mathbb{R}; H^{1/2}(\Gamma)) \right\},
\]

with \( g = g^*|_{[0, T]} \) and \( g^* \equiv 0 \) on \((-\infty, 0)\),

\[
H^r(\mathbb{R}; H^{1/2}(\Gamma)) := \left\{ g : \Gamma \times \mathbb{R} \to \mathbb{R} : \int_{-\infty}^{\infty} (1 + |\omega|)^{2r} \| (\mathcal{F}g)(\cdot, \omega)\|^2_{H^{1/2}(\Gamma)} d\omega < \infty \right\},
\]

and \( \mathcal{F} \) denotes the integral Fourier transform with respect to the time variable \( t \in \mathbb{R} \).

Our goal is to prove that the parameters in our method can be chosen so that convergence rates in (5.1) are preserved.

5.1. Errors due to the perturbation of \( \omega_n^t \). Let \( V_h(s) : S \to S \) be defined by

\[
(V_h(s) \varphi, \psi)_{L^2(\Gamma)} := (V(s) \varphi, \psi)_{L^2(\Gamma)} \quad \forall \varphi, \psi \in S.
\]

Whenever necessary, we will identify the inner product \((\cdot, \cdot)_{L^2(\Gamma)}\) with its extension to the dual pairing \( H^{-1/2}(\Gamma) \times H^{1/2}(\Gamma) \). The solution by the convolution quadrature, i.e., with exact weights, is given by (see equation (5.5) in [31])

\[
g^h = V_h^{-1}(\partial_t^\Delta)g^h,
\]

whereas with the perturbed weights the solution is given by

\[
g^{h,\lambda} = \left( V_h^{-1}(\partial_t^{\Delta t,\lambda})g^h \right)(t_n)
\]

(see Remark 3.1), where \( g^h \in S \) is the \( L^2 \)-projection of \( g \) on \( S \) as follows:

\[
(g^h, \psi)_{L^2(\Gamma)} = (g, \psi)_{L^2(\Gamma)} \quad \forall \psi \in S.
\]

For the remainder of the paper we will make use of the following notation:

\[
\| \|_{+1} = \| \cdot \|_{H^{1/2}(\Gamma) - H^{-1/2}(\Gamma)} \quad \text{and} \quad \| \|_{-1} = \| \cdot \|_{H^{-1/2}(\Gamma) - H^{1/2}(\Gamma)}.
\]

**Lemma 5.1.** Let \( \Re s \geq \sigma_0 > 0 \). Then

\[
\| V_h^{-1}(s) \|_{-1} \leq \frac{C_{\text{stab}}}{\min(1, \sigma_0)} |s|^2.
\]

**Proof.** The result follows immediately from the definition of \( V_h(s) \) and the coercivity estimate for \( V(s) \) as follows (see [3]):

\[
\Re (sV(s)\psi, \psi)_{L^2(\Gamma)} \geq C_{\text{stab}} \frac{\min(1, \sigma_0)}{|s|} \| \psi \|^2_{H^{-1/2}(\Gamma)}.
\]

**Remark 5.2.** For \( \omega \in \mathbb{R} \), there holds

\[
\gamma (\lambda e^{i\omega}) = \frac{(\lambda + 3) (1 - \lambda) + 8 (1 - \lambda) \lambda \sin^2 \frac{\omega}{2} + 8 \lambda^2 \sin^4 \frac{\omega}{2}}{2} - i\lambda \sin \omega \left( 2 (1 - \lambda) + \lambda \left( 1 + 2 \sin^2 \frac{\omega}{2} \right) \right).
\]
For the real part, we obtain the estimate

$$\text{Re}(\gamma e^{i\omega}) / |\Delta t| \geq \left( \frac{1 - \lambda}{2} + 4\lambda^2 \sin^4 \frac{\omega}{2} \right) / |\Delta t|.$$  

For $0 < \lambda < 1$, we have the uniform bound with respect to $\omega$,

$$\text{Re} \left( \frac{\lambda e^{i\omega}}{\Delta t} \right) \geq \text{Re} \left( \frac{\lambda}{\Delta t} \right) = \frac{(3 + \lambda)(1 - \lambda)}{2\Delta t} \geq \frac{3(1 - \lambda)}{2\Delta t}.$$  

For the modulus, the (rough) upper estimate holds as follows:

$$\left| \frac{\gamma (e^{i\omega})}{\Delta t} \right| \leq \frac{C}{\Delta t} \quad \text{with} \quad C = 5^{3/2}.$$  

**Lemma 5.3.** Let $W_h(s) := V_h^{-1}(s)/s^2$. Then,

$$\|\omega_j^{\Delta t}(W_h)\|_{-1} \leq 2C_{\text{stab}}eT.$$  

Further, for sufficiently smooth and compatible $g$, the identities

$$(5.4) \quad V_h^{-1}(\partial_{\Delta t}^\lambda g) = W_h(\partial_{\Delta t}^\lambda ((\partial_{\Delta t}^\lambda)^2 g))$$  

and, for $N \geq 4$,

$$(5.5) \quad V_h^{-1}(\partial_{\Delta t}^{\lambda, \lambda} g) = W_h(\partial_{\Delta t}^{\lambda, \lambda} ((\partial_{\Delta t}^\lambda)^2 g)),$$  

hold, where $(\partial_{\Delta t}^\lambda)^2 g$ denotes the twofold application of the multistep approximation, which in our case is the BDF2 scheme.

**Proof.** The bound for $\|\omega_j^{\Delta t}(W_h)\|_{-1}$ follows from the Cauchy estimate by choosing the circle with radius $e^{-\Delta t/T}$ as the integration contour in (3.5), Remark 5.2, and Lemma 5.1 as follows:

$$\|\omega_j^{\Delta t}(W_h)\|_{-1} \leq e^{\Delta t/T} \max_{\|z\| = 1} \left| W_h \left( \gamma(e^{-\Delta t/T}z)/\Delta t \right) \right|_{-1} \leq \frac{C_{\text{stab}}}{\min(1, (1 - e^{-\Delta t/T})/(2\Delta t))} e^{\Delta t/T} \leq 2C_{\text{stab}}Te^{\Delta t/N}.$$  

Applying the (scaled) inverse discrete Fourier transform to the identity $V_h^{-1}(s_l) \tilde{g}_l = W_h(s_l) s_l^2 \tilde{g}_l$, we see that $V_h^{-1}(\partial_{\Delta t}^{\lambda, \lambda} g^h) = W_h(\partial_{\Delta t}^{\lambda, \lambda}) g^h$, where

$$\tilde{g}^h_n = \frac{\lambda^{-n}}{N + 1} \sum_{l=0}^{N+1} g_l^h s_l^2 \zeta_{N+1}^l,$$  

$$s_l = \frac{\gamma(\lambda N^{-1} - l)}{\Delta t}.$$  

The inverse discrete Fourier transform of $s_l^2$ is

$$(5.6) \quad \frac{1}{N + 1} \sum_{l=0}^{N+1} \left( \frac{\gamma(\lambda N^{-1} - l)}{\Delta t} \right)^2 \zeta_{N+1}^l \approx \frac{\lambda^2}{2\pi} \int_C \left( \frac{\gamma(\lambda \zeta)/\Delta t}{\zeta^{j+1}} \right)^2 d\zeta = \frac{\lambda^2}{\Delta t^2} \delta_j,$$  

where

$$(\gamma(\zeta))^2 = \sum_{k=-\infty}^{\infty} \delta_k \zeta^k = \left( \frac{3}{2} - 2\zeta + \frac{1}{2}\zeta^2 \right)^2.$$  

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Since \((\gamma(\zeta))^2\) is a polynomial of order 4 and \(N \geq 4\), the coefficients \(\frac{\lambda^j}{\Delta t^j}\delta_j\) are reproduced exactly, without any quadrature error in (5.6). Therefore

\[
\tilde{g}^h_n = \frac{1}{\Delta t^2} \sum_{j=0}^{n} \delta_{n-j} g_j,
\]

which is exactly the result of applying the BDF2 multistep method twice, where it is implicitly assumed that \(g(t) = 0\) for \(t \leq 0\). The result for \(V_h^{-1}(\partial_t^{\Delta t})g^h\) is proved similarly, but with no restriction on \(N\); see also [31].

**Proposition 5.4.** Let \(0 < \lambda < 1\). Then

\[
\|V_h^{-1}(\partial_t^{\Delta t})g^h - V_h^{-1}(\partial_t^{\Delta t,\lambda})g^h\|_{H^{-1/2}(\Gamma)} \leq 2C_{\text{stab}} e T^2 \frac{\lambda^{N+1}}{1 - \lambda^{N+1}} \Delta t^{-1}.
\]

**Proof.** Let \(a_j := \lambda^j \omega_j^{\Delta t}(W_h)\), and let \(\hat{a}_j := \lambda^j \omega_j^{\Delta t,\lambda}(W_h)\), \(W_h(s) = V_h^{-1}(s)/s^2\). Then \(\hat{a}_j\) is the discrete Fourier transform approximation to \(a_j\) for \(j = -N, \ldots, N\) and (see [25])

\[
\|a_j - \hat{a}_j\|_{-1} = \left\| \sum_{l=1}^{\infty} a_{j+l(N+1)} + a_{j-l(N+1)} \right\|_{-1} \leq \sum_{l=1}^{\infty} \|a_{j+l(N+1)}\|_{-1}
\]

\[
\leq \lambda^j \sum_{l=1}^{\infty} \lambda^{l(N+1)} \|\omega_j^{\Delta t}(W_h)\|_{-1} \leq 2C_{\text{stab}} e T \lambda^j \frac{\lambda^{N+1}}{1 - \lambda^{N+1}},
\]

where we have used the bound (5.3). Therefore

\[
\|\omega_j^{\Delta t}(W_h) - \omega_j^{\Delta t,\lambda}(W_h)\|_{-1} \leq 2C_{\text{stab}} T \frac{\lambda^{N+1}}{1 - \lambda^{N+1}},
\]

and the result follows from the definition of the discrete convolution and identities (5.4) and (5.5).

**Theorem 5.5.** Let the exact solution \(\phi(\cdot, t)\) be in \(H^{m+1}(\Gamma)\) for any \(t \in [0, T]\),

data \(g \in H_0^0([0, T]; H^{1/2}(\Gamma))\), \(0 < \lambda < 1\), and let the boundary element space be \(S = S_{m-1, m}\) for \(m \in \{0, 1\}\). Then the discrete solution

\[
\phi_h^{t,\lambda} = \left(V_h^{-1}(\partial_t^{\Delta t,\lambda})g^h\right)(t_n)
\]

satisfies the error estimate

\[
\|\phi_h^{t,\lambda} - \phi(\cdot, t_n)\|_{H^{-1/2}(\Gamma)} \leq C_g \left(\frac{\lambda^{N+1}}{1 - \lambda^{N+1}} T^2 \Delta t^{-1} + \Delta t^2 + h^{m+3/2}\right),
\]

where \(C_g\) depends on the right-hand side \(g\), \(C_{\text{stab}}\), and the time interval length \(T\).

**Proof.** The result is a direct consequence of Proposition 5.4 and (5.1); see [31, Theorem 5.4].

**5.2. Error due to the perturbation of \(V_h(s)\).** We investigate the effect of perturbing \(V_h(s)\), in particular the effect of approximate evaluation of the kernel \(K(d, s)\) by separable expansions. If these perturbations could be chosen to be analytic in \(s\), then a stability and error estimate from Lemma 5.5 in [31] could be used, in which there is no loss of powers of \(\Delta t\). Unfortunately due to numerical stability issues...
and investigate how this perturbation affects the final solution.

Lemma 5.6. Let \( \Re s > \sigma_0 > 0 \) and \( \varepsilon < \frac{1}{2} C_{\text{stab}}^{-1} \frac{\min(1, \sigma_0)}{|s|^2} \). Then \( (V^\varepsilon_h(s))^{-1} \) is bounded and

\[
\| (V^\varepsilon_h(s))^{-1} \|_{-1} \leq 2C_{\text{stab}} \frac{|s|^2}{\min(1, \sigma_0)}.
\]

Proof. Let us write

\[
V^\varepsilon_h(s) = V_h(s) \left[ I - V_h^{-1}(s) (V_h(s) - V^\varepsilon_h(s)) \right].
\]

From the estimate \( \| V_h^{-1}(s) \|_{-1} \leq C_{\text{stab}} |s|^2 / \min(1, \sigma_0) \) (see Lemma 5.1), we see that \( \varepsilon < \frac{1}{2} C_{\text{stab}}^{-1} \min(1, \sigma_0) / |s|^2 \) is sufficient for \( (V^\varepsilon_h(s))^{-1} \) to exist and to be bounded as above.

Lemma 5.7. Let \( \min_{l=0, \ldots, N} \Re s_l > \sigma_0 > 0 \) and \( \varepsilon < \frac{1}{2} C_{\text{stab}} \frac{\min(1, \sigma_0)}{\max_{l=0, \ldots, N} |s_l|^2} \).

Then

\[
\| \omega^\Delta t, \lambda(Q_h) - \omega^\Delta t, \lambda(Q^\varepsilon_h) \|_{-1} \leq CT \lambda^{-j} \varepsilon \Delta t^{-1},
\]

where

\[
C = \left( \frac{C_{\text{stab}}}{\min(1, \sigma_0)} \right)^2, \quad Q_h(s) := \frac{V_h^{-1}(s)}{s^4}, \quad \text{and} \quad Q^\varepsilon_h(s) := \frac{(V^\varepsilon_h(s))^{-1}}{s^4}.
\]

Proof. Using the fact that \( Q_h^{-1}(s) = s^4 V_h(s) \), we obtain the bound

\[
\| Q_h(s_l) - Q^\varepsilon_h(s_l) \|_{-1} = \| Q_h(s_l)(s_l^4 V^\varepsilon_h(s_l) - s_l^4 V_h(s_l))Q^\varepsilon_h(s_l) \|_{-1} \leq \frac{C_{\text{stab}}}{\min(1, \sigma_0)}^2 \varepsilon.
\]

From this and the definition of the perturbed convolution weights, the result follows.

Let us define the solution of the \( \varepsilon \)-perturbed convolution equation by

\[
\phi^\lambda, h, \varepsilon := (V^\varepsilon_h)^{-1}(\partial_t^\Delta t, \lambda) g = Q^\varepsilon_h(\partial_t^\Delta t, \lambda) \left( (\partial_t^\Delta t)^4 g \right)
\]

and, as before,

\[
\phi^\lambda, h := V_h^{-1}(\partial_t^\Delta t, \lambda) g = Q_h(\partial_t^\Delta t, \lambda) \left( (\partial_t^\Delta t)^4 g \right).
\]

In the next result we estimate the difference between the two.

Proposition 5.8. Let \( \min_{l=0, \ldots, N} \Re s_l > \sigma_0 > 0 \), let

\[
\varepsilon < \frac{1}{2} C_{\text{stab}} \frac{\min(1, \sigma_0)}{\max_{l=0, \ldots, N} |s_l|^2},
\]

and let the data \( g \) be sufficiently smooth and compatible. Then

\[
\| \phi^\lambda, h, \varepsilon - \phi^\lambda, h, n \|_{H^{-1/2}(\Gamma')} \leq C\varepsilon T^2 \lambda^{-n} \Delta t^{-2},
\]
with $C > 0$ as in Lemma 5.7.

Proof. The result is a direct consequence of the above lemma.

The above result, together with Remark 5.2, implies that to obtain optimal convergence it is sufficient to insure that $\varepsilon \leq C\lambda^N \Delta t^4$.

Let us now investigate the effect of perturbations on the kernel function $K(d, s)$. In order to do this, we assume

$$(5.8) \quad |K(||x-y||, s_i) - K^{pc}||x-y||, s_i)\leq \delta \frac{1}{||x-y||} \forall x, y \in \Gamma$$

for $l = 0, 1, \ldots, N$, and define the operator $V^{pc}_h(s) : S \rightarrow S$ by

$$(V^{pc}_h(s)\psi, \phi)_{L^2(\Gamma)} = \int_{\Gamma} \int_{\Gamma} K^{pc}(||x-y||, s)\psi(y)\phi(x)d\Gamma_y d\Gamma_x.$$ 

**Proposition 5.9.** Let (5.8) hold. Then, there exists $C_0 > 0$ such that

$$\|V^{pc}_h(s_i) - V_h(s_i)\|_{+1} \leq C_0 h^{-1}\delta.$$ 

Hence if $\delta \leq \frac{1}{\Delta} C_0 C_{\text{stab}} h_{\text{min}1,\sigma_0} \max |s|_2 \leq C h\Delta t^2$, the estimate

$$\|\phi_n^{\lambda, h} - \phi_n^{\lambda, h}\|_{H^{-1/2}(\Gamma)} \leq C\delta h^{-1} \lambda^{-N} \Delta t^{-2}$$

holds, where

$$\phi_n^{\lambda, h} = (V^{pc}_h)^{-1}(\partial t^{\lambda, h}) g.$$ 

Proof. Let $\phi \in S$. The well-known $L^2$-continuity of the single layer potential for the Laplacian and a scaling inequality for boundary element functions lead to

$$\|\phi_{V^{pc}_h(s_i) - V_h(s_i)}\|_{H^{1/2}(\Gamma)} \leq \delta \sup_{\phi \in S(\Gamma)} \int_{\Gamma \times \Gamma} |\phi(y)| |\psi(x)| \frac{1}{||x-y||} ds_x ds_y$$

$$\leq C\delta \sup_{\phi \in S(\Gamma)} \|\phi\|_{H^{-1/2}(\Gamma)} \|\psi\|_{L^2(\Gamma)} \leq C h^{-1} \delta \|\phi\|_{H^{-1/2}(\Gamma)}.$$ 

The estimate of the error in the solution is then a direct consequence of Proposition 5.8 and Remark 5.2.

In the following result, the binary relation $A \lesssim B$ is used to denote the existence of a constant $C$ independent of any discretization parameters such that $A \leq CB$. Further, $A \sim B$ implies $A \lesssim B$ and $B \lesssim A$.

**Corollary 5.10.** Let the conditions of Theorem 5.5 be satisfied, let (5.8) hold, and let

$$h^{m+3/2} \lesssim \Delta t^2, \quad \lambda^{N+1} \sim \Delta t^3, \quad \delta \lesssim \lambda^N h^{3} \Delta t^4 \lesssim h^{7m/2+25/4}.$$ 

Then the optimal rate of convergence is achieved

$$\|\phi_n^{\lambda, h} - \phi(s, t_n)\|_{H^{-1/2}(\Gamma)} \leq C \Delta t^2.$$
where $C$ depends on the data $g$.

Remark 5.11. According to the above result, $\lambda$ should be chosen as $\lambda \sim \Delta t^{3/(N+1)} = e^{\frac{1}{N+1} \log \frac{r}{N}}$. Since the rounding errors, in the same manner as the errors due to panel-clustering, are magnified by $\lambda^{-1}$, $\lambda$ should be chosen in the interval $\sqrt{\varepsilon_{ps}} < \lambda^{N} < 1$, where $\varepsilon_{ps}$ is the machine accuracy. In IEEE double precision this is approximately $10^{-16}$; therefore the accuracy of the method is limited by the choice $\lambda > 10^{-8/N}$. This accuracy limit can, however, be improved if an $n$-trapezoidal rule is used to compute the weights $\omega^{\Delta t, \lambda}$ with $n = jN$, $j > 1$.

Remark 5.12. The condition on the accuracy of the panel-clustering approximation is rather stringent. However, since the convergence of the separable expansion is exponential for a large enough length of expansion $L$ (see (4.7)), the computational costs of the panel-clustering method depend only logarithmically on the required accuracy. Therefore the overall computational cost is not significantly affected.

If we had assumed that $V_h^{\text{per}}(s) - V_h(s)$ is analytic in $s$ and could be bounded by $C|s|^2$, we could be obtained significantly better error estimates by using Lemma 5.5 in [31]. Unfortunately, due to the well-known numerical stability issues with the multipole expansions for the Helmholtz kernel [5], [9], [34], different types of expansions need to be used for different admissible blocks; the choice of the expansion depends on the wave number $s_l$ and the size of the block. This restricts us from using the more favorable results of Lemma 5.5 in [31].

5.3. Error due to the reduction of the number of linear systems.

Corollary 5.13. Let $0 \leq \lambda < 1$ and $\sigma_l = \text{Re } s_l$. Then

$$
\| \hat{\phi}_n^h \|_{H^{-1/2}(\Gamma)} \leq C_1 (\Delta t)^{-2} \| \hat{g}_n \|_{H^{1/2}(\Gamma)},
$$

where $C_1 = 5^3 \frac{C_{\text{min}}}{\min(1, \sigma_l)}$.

Proof. The result is a direct consequence of Lemma 5.1 and Remark 5.2.

Let $N_x \subset \{0, 1, \ldots, N\}$ determine the Helmholtz problems, the solution of which will be computed; the rest will be approximated by zero. Then we define the resulting approximation to $\phi_n^{h, \lambda}$ by

$$
\theta \phi_n^{h, \lambda}(x) := \frac{\lambda^{-n}}{N+1} \sum_{l \in N_x} \hat{\phi}_l^h(x) c_l^{h,n+1}.
$$

Corollary 5.14. Let $n \in \{0, 1, \ldots, N\}$. If

$$
\max_{l \in N_x} \| \hat{g}_l \|_{H^{1/2}(\Gamma)} \leq C_1^{-1} \lambda^n (\Delta t)^4,
$$

then we obtain optimal order convergence at time step $t_n$:

$$
\| \theta \phi_n^{h, \lambda} - \phi_n^{h, \lambda} \|_{H^{-1/2}(\Gamma)} \leq \Delta t^2.
$$

Proof. The proof follows directly from Corollary 5.13.

Next we show that if the right-hand side is smooth and of finite duration, it is sufficient to solve only a few Helmholtz systems. Let us introduce the space of functions that are zero at both $t = 0$ and $t = T$ as follows:

$$
H_0^0([0, T]; H^{1/2}(\Gamma)) := \left\{ g : \Gamma \times [0, T] \to \mathbb{R} : \text{there exists } g^* \in H^r(\mathbb{R}; H^{1/2}(\Gamma)) \text{ with } g = g^*|_{[0, T]} \text{ and } \text{supp } g^* \subset [0, T] \right\}.
$$
Theorem 5.15. Let \( g \in H^r_0([0,T];H^{1/2}(\Gamma)) \) for some \( r \geq 3.5 \), and let \( \epsilon > 0 \) be given. For any \( N \in \mathbb{N} \) let \( \lambda := c^\frac{1}{r} \). Then, \( N_z \) can be chosen so that \( \#N_z \leq C e^{-c\epsilon^{-1/2}N^{1-1/2}} \) and the optimal order convergence is retained. The constant \( C \) depends on \( r, (\log \epsilon)/T, \) and \( g \).

Proof. Let \( g_\lambda(x,t) := \lambda^{t/\Delta t} g(x,t) = e^{t\log \lambda} g(x,t) \) on \( t \in [0,T] \). Then we see that \( g_\lambda \) is independent of \( N \) and that \( g_\lambda \in H^r_0([0,T];H^{1/2}(\Gamma)) \). Then for \( \omega \in \mathbb{R} \), \( \| (\mathcal{F}g_\lambda)(\cdot,\omega) \|_{H^{1/2}(\Gamma)} = o(|\omega|^{-r-1/2}) \). Taking \( \omega_j = 2\pi j/(T+\Delta t) = 2\pi j/(T(N+1)) \), we define
\[
a_j := \| (\mathcal{F}g_\lambda)(\cdot,\omega_j) \|_{H^{1/2}(\Gamma)} = o(j^{-r-1/2}), \quad j \in \mathbb{Z}.
\]
Then using the aliasing formula (see [25]), we arrive at the following estimate for \( \hat{g}_n, \)
\[
\| \hat{g}_n \|_{H^{1/2}(\Gamma)} \leq a_n + \sum_{k>N/2} a_k = o(n^{-r-1/2} + N^{-r+1/2}) = o(n^{-r-1/2}).
\]
The constants in the \( o(\cdot) \) notation depend only on \( r, (\log \epsilon)/T, \epsilon, \) and \( g \). The result now follows from Corollary 5.14. \( \square \)

6. Numerical experiments. In this section we present the results of numerical experiments. Except for one simple example, the experiments will be done in two dimensions. All the steps in the method remain the same in two dimensions, except that the fundamental solution for the wave equation is given by
\[
k_{2D}(d,t) = \frac{H(t-d)}{2\pi \sqrt{t^2 - d^2}},
\]
where \( H \) is the Heaviside function,
\[
H(t) = \begin{cases} 
0 & \text{for } t < 0, \\
1 & \text{for } t > 0.
\end{cases}
\]
The Laplace transform \( K_{2D}(d,s) \) is again the fundamental solution of the Helmholtz equation \( \Delta U - s^2 U \) as follows:
\[
k_{2D}(d,s) = \frac{i}{4} H_0^{(1)}(isd),
\]
where \( H_0^{(1)}(\cdot) \) is the zero order Hankel function of the first kind.

Let us consider the case of \( \Gamma \) being the unit ball in \( \mathbb{R}^2 \) or \( \mathbb{R}^3 \) and a right-hand side that is separable in the time and the spatial variables: \( g(x,t) = g(t) e(x) \), where \( e(x) \) is an eigenfunction of the single layer potential \( V(s) \) with the eigenvalue \( \lambda_l(s) \). In two dimensions the eigenfunctions are the complex exponentials \( e^{i\lambda_l} \) and \( \lambda_l(s) = \frac{\pi}{2} \), whereas in three dimensions these are the spherical harmonics \( Y_l^m(\theta,\varphi) \) with \( \lambda_l(s) = -s j_l(h_l(s)); \) we have used the standard polar/spherical coordinates to describe the eigenfunctions. Here \( j_l(\cdot) \) (respectively, \( h_l(\cdot) \)) are cylindrical (respectively, spherical) Bessel functions of order \( l \), whereas \( H_l^{(1)}(\cdot) \) (respectively, \( h_l^{(1)}(\cdot) \)) are the cylindrical (respectively, spherical) Hankel functions of the first kind and order \( l \). The problem of finding the unknown density \( \phi(x,t) \) can then be reduced to the single, time, dimension. This can be seen by replacing the fundamental solution \( k \) in the
single layer representation formula by the inverse Laplace transform of its Laplace transform $K$ as follows:

$$g(t)e(x) = \int_0^t \int_\Gamma k(t - \tau, \|x - y\|)\phi(\tau, x)d\Gamma_y d\tau$$

$$= \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} e^{s\tau} \int_\Gamma K(s, \|x - y\|)\phi(t - \tau, y)d\Gamma_y ds$$

$$= \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \int_0^t e^{s\tau} \lambda_l(is)\phi(t - \tau, \cdot)(x)d\tau ds,$$

$x \in \Gamma$, for some $\sigma > 0$.

Therefore, we can use the ansatz $\phi(x, t) = \phi(t)e(x)$ to reduce the problem to finding $\phi(t)$ such that

$$g(t) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \int_0^t e^{s\tau} \lambda_l(is)\phi(t - \tau) d\tau ds.$$

Hence we need to solve a convolution integral equation in one dimension as follows:

(6.3) $$g(t) = \int_0^t \tilde{\lambda}_l(\tau)\phi(t - \tau) d\tau,$$

where $\tilde{\lambda}_l(\cdot)$ is the inverse Laplace transform of $\lambda_l(\cdot)$. The latter equation can then be solved by Lubich’s original method, which makes use of only $\lambda_l(\cdot)$ and not its inverse Laplace transform. The first few numerical examples will be of this type.

**6.1. Radial solution of scattering by unit sphere.** In this example we consider the three-dimensional case, $\Gamma = S^2$. Let $g(x, t) = g(t)$ be constant for a fixed time $t$, i.e., $e(x) = 2\sqrt{\pi}Y_0^0 = 1$. In this particularly simple case it can be shown that

$$\phi(t) = 2g'(t), \quad t \in [0, 2].$$

The restriction to the interval $[0, 2]$ is a consequence of the fact that the diameter of the sphere is 2. For time $t > 2$ the expression for $\phi(t)$ is more complicated.

The right-hand side of the $n$th Helmholtz problem is a constant,

$$\hat{g}_n = \sum_{j=0}^N \lambda^j g(t_j)\zeta^{-n_j},$$

and the solution of the Helmholtz problem is also a constant and is given by

$$\hat{\phi}_n = \frac{\hat{g}_n}{\lambda_0(s_n)}.$$

The approximation to the unknown density at time step $t_n$ is given by

$$\phi_n := \frac{\lambda - n}{N + 1} \sum_{j=0}^N \hat{\phi}_j s_{N+1}^n.$$

If $\lambda$ is chosen small enough, theoretical estimates predict the following behavior of the error:

$$\left(\sum_{n=0}^N \Delta t |\phi(t_n) - \phi_n|^2\right)^{1/2} \leq C\Delta t^2.$$
One more detail needs to be fixed before the experiments can be started, namely, the choice of $\lambda$. Recall that $\lambda$ needs to be chosen small enough to insure stability and accuracy (see Theorem 5.5) but also large enough to avoid numerical instability issues (see Remark 5.11). As suggested in Remark 5.11, we make the choice

$$
(6.4) \quad \lambda = \max(\Delta t^{3/N}, \text{eps}^{1/2}).
$$

Numerical results for the scattering by unit sphere are given in Table 1 and show that our theoretical estimates are sharp for this example.

### 6.2. A nonradial example

In this example we consider the two-dimensional case. We pick the right-hand side to be $g(x, t) = h(t) \cos(l\theta)$, where for the space variable we use the polar coordinate system $r \in \mathbb{R}_{\geq 0}$, $\theta \in [0, 2\pi)$. Since $\cos(l\theta)$ is an eigenfunction of the single layer potential $V(s_n)$, the Helmholtz problems can be solved exactly. However, to investigate the effect of spatial discretization we solve the problems using the Galerkin method, and hence obtain an approximation $\phi_{h,\lambda}(t_n, \theta)$ of the unknown density. To investigate the error, we use the fact that $\phi_{\theta,t} = \phi(t) \cos(l\theta)$ and solve with high accuracy for $\phi(t)$ by applying Lubich’s method to the one-dimensional problem (6.3). The error measure we use is the following:

$$
\|\phi - \phi_{h,\lambda}\|_{-1/2,2} := \left( \sum_{n=0}^{N} \Delta t \|\phi(t_n) \cos(l\cdot) - \phi_{h,\lambda}(t_n, \cdot)\|^2_{H^{-1/2}(\Gamma)} \right)^{1/2}.
$$

The theory predicts the above error to be proportional to $h^{m+3/2} + \Delta t^2$, where $m = 0$ for the Galerkin basis of piecewise constant functions and $m = 1$ for the basis of piecewise linear functions. In all the experiments, we choose $\lambda$ as in (6.4). To see if the spatial discretization has introduced significant errors, we compute the error obtained when the Helmholtz problems are solved exactly. The results are given in the following table:

<table>
<thead>
<tr>
<th>$N$</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|\phi - \phi_{h,\lambda}|_{-1/2,2}$</td>
<td>0.61</td>
<td>0.24</td>
<td>0.077</td>
<td>0.022</td>
<td>0.0057</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Comparing these results to Tables 2 and 3, we see that the error due to the discretization in space is not significant.

### 6.3. Reduction of the number of systems

Let us now consider a signal that is smooth and of nearly limited time duration as follows:

$$
(6.5) \quad g(r, t) = \cos(5t - r\.\alpha) \exp(-1.5(5t - r\.\alpha - 5)^2), \quad \alpha = (1, 0).
$$
The results for scattering by the unit disk with $g(x, t) = \sin^5(t) \cos(3x)$ and the piecewise constant Galerkin basis $S = S_{-1, 0}$. $M$ is chosen so that $h^{3/2} \propto \Delta t^2$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M$</th>
<th>$|\phi - \phi^{h, \lambda}|_{-1/2, 2}$</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>16</td>
<td>0.78</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>0.27</td>
<td>1.54</td>
</tr>
<tr>
<td>16</td>
<td>102</td>
<td>0.084</td>
<td>1.68</td>
</tr>
<tr>
<td>32</td>
<td>254</td>
<td>0.023</td>
<td>1.83</td>
</tr>
<tr>
<td>64</td>
<td>640</td>
<td>0.0062</td>
<td>1.93</td>
</tr>
<tr>
<td>128</td>
<td>1610</td>
<td>0.0016</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Table 3

The results for scattering by the unit disk with $g(x, t) = \sin^5(t) \cos(3x)$ and the piecewise linear Galerkin basis $S = S_{0, 1}$. $M$ is chosen so that $h^{5/2} \propto \Delta t^2$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M$</th>
<th>$|\phi - \phi^{h, \lambda}|_{-1/2, 2}$</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>22</td>
<td>0.66</td>
<td>-/–</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>0.26</td>
<td>1.34</td>
</tr>
<tr>
<td>16</td>
<td>68</td>
<td>0.082</td>
<td>1.67</td>
</tr>
<tr>
<td>32</td>
<td>116</td>
<td>0.023</td>
<td>1.84</td>
</tr>
<tr>
<td>64</td>
<td>204</td>
<td>0.0060</td>
<td>1.93</td>
</tr>
<tr>
<td>128</td>
<td>352</td>
<td>0.0015</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Table 4

The results for scattering by the unit disk where the incoming wave is a Gaussian pulse and the piecewise linear Galerkin basis $S = S_{1, 0}$ is used. The column $\#N_z$ shows the number of Helmholtz problems actually solved.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$#N_z$</th>
<th>$|\phi - \phi^{h, \lambda}|_{-1/2, 2}$</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>2.9</td>
<td>-/–</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>2.9</td>
<td>-0.03</td>
</tr>
<tr>
<td>16</td>
<td>9</td>
<td>1.4</td>
<td>1.09</td>
</tr>
<tr>
<td>32</td>
<td>17</td>
<td>0.42</td>
<td>1.70</td>
</tr>
<tr>
<td>64</td>
<td>24</td>
<td>0.11</td>
<td>1.92</td>
</tr>
<tr>
<td>128</td>
<td>24</td>
<td>0.028</td>
<td>1.98</td>
</tr>
<tr>
<td>256</td>
<td>24</td>
<td>0.0072</td>
<td>1.99</td>
</tr>
</tbody>
</table>

For such a Gaussian pulse our theory predicts that only $O(N^\epsilon)$, for any fixed $\epsilon > 0$, Helmholtz systems need to be solved to obtain optimal convergence; see also Figure 2. The results for scattering by the unit disk and for piecewise-linear basis functions $S = S_{1, 0}$ are given in Table 4. Since we approximate by zero only the solutions of those Helmholtz problems whose right-hand sides are zero almost to machine precision, the number of Helmholtz problems $\#N_z$ is constant for large enough $N$. For this more complicated problem, for each $N$ we have used as the reference solution the numerical solution using $2N$ steps in time and the corresponding number of nodes in the discretization in space.

7. Conclusion. We have described a method that requires the solution of a number of Helmholtz problems to obtain an approximate solution of the wave equation in an unbounded, homogeneous medium. We have proved stability and optimal convergence results for this approach. Further, we have indicated ways in which to efficiently solve the resulting system of Helmholtz problems. The stability and convergence results of the perturbations introduced by the efficient solvers have also been presented.

The fast methods we propose using are typically capable of computing a matrix-
vector product in almost linear time, i.e., $O(M \log^a M)$, of a single dense $M \times M$ system arising from the discretization of the Helmholtz single layer potential. In order to solve efficiently the linear system by an iterative method requiring only matrix-vector multiplication, a good preconditioner is needed. The investigation of such a preconditioner is beyond the scope of this paper. With a preconditioned iterative solver we expect to obtain computational costs which scale linearly, up to logarithmic terms, with respect to the number of unknowns $NM$. An important observation is that in some cases only a few Helmholtz systems need to be solved. Although this does not change the overall complexity (the discrete Fourier transformation still requires $O(MN \log N)$ operations), it can hugely reduce the absolute time for the computation. The storage costs will also scale linearly since at any one time only a single linear system representing the discretization of a Helmholtz problem needs to be stored. Since all the $NM$ coefficients $\phi_{j,n}$ are stored, the storage costs are not better than linear. Crucially, since the Helmholtz problems to be solved are entirely decoupled, the proposed method is easily parallelizable.

These asymptotic estimates significantly improve both the storage and computational costs compared to the previously proposed approaches for the solution of the wave equation using the convolution quadrature discretization in time; see [24] and [22], [23], [28]. The asymptotic costs of the MOT method presented in [10], [18] are also almost linear in the number of degrees of freedom. Advantages of our method include the intrinsic parallel nature of the method, proven convergence and stability properties, and the relatively simple implementation details. In a forthcoming paper, algorithmic details for the data sparse approximations, a more in-depth asymptotic complexity analysis, and large-scale computational results will be presented.

Acknowledgments. We gratefully acknowledge the fruitful discussions with Ch. Lubich which led to significant improvements in the perturbation analysis.

REFERENCES


