SIAM J. MATRIX ANAL. APPL.
Vol. 41, No. 1, pp. 58–82

SOLVING THE THREE-DIMENSIONAL HIGH-FREQUENCY HELMHOLTZ EQUATION USING CONTOUR INTEGRATION AND POLYNOMIAL PRECONDITIONING[*]

XIAO LIU[†], YUANZHE XI[‡], YOUSEF SAAD[§], AND MAARTEN V. DE HOOP[†]

Abstract. We propose an iterative solution method for the three-dimensional high-frequency Helmholtz equation that exploits a contour integral formulation of spectral projectors. In this framework, the solution in certain invariant subspaces is approximated by solving complex-shifted linear systems, resulting in faster GMRES iterations due to the restricted spectrum. The shifted systems are solved by exploiting a polynomial fixed-point iteration, which is a robust scheme even if the magnitude of the shift is small. Numerical tests in three dimensions indicate that $O(n^{1/3})$ matrix-vector products are needed to solve a high-frequency problem with a matrix size n with high accuracy. The method has a small storage requirement, can be applied to both dense and sparse linear systems, and is highly parallelizable.

Key words. Helmholtz preconditioner, Cauchy integral, shifted Laplacian, polynomial iteration

AMS subject classifications. 15A06, 65F08, 65F10, 65N22, 65Y20

DOI. 10.1137/18M1228128

1. Introduction.

1.1. Problem of interest.

Helmholtz-type equations are second-order partial differential equations that model time-harmonic waves in materials with linear constitutive relations. For the scalar case, the Helmholtz operator can be written as

$$-\Delta - \omega^2/c^2(x), \tag{1.1}$$

where ω is the angular frequency, and $c(x)$ is the wavespeed. After one of several types of discretizations is applied, we end up with an $n \times n$ linear system of the following form to solve:

$$Au = f. \tag{1.2}$$

The main subject of this paper is the fast iterative solution of (1.2). The linear system (1.2) is challenging to solve because the coefficient matrix A is typically *highly indefinite* and *non-Hermitian*.

[*] Received by the editors November 21, 2018; accepted for publication (in revised form) by L. Giraud November 4, 2019; published electronically January 9, 2020.
https://doi.org/10.1137/18M1228128
Funding: The work of the second and third authors was supported by National Science Foundation grant DMS-1521573 and the Minnesota Supercomputing Institute. The work of the fourth author was supported by the Simons Foundation under the MATH + X program, National Science Foundation grant DMS-1559587, the corporate members of the Geo-Mathematical Group at Rice University, and Total.

[†] Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005 (xiao.liu@rice.edu, mdehoop@rice.edu).
[‡] Department of Mathematics, Emory University, Atlanta, GA 30322 (yxi26@emory.edu).
[§] Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455 (saad@umn.edu).

1.2. Existing work. Iterative methods can show fast convergence for non-Hermitian linear systems, when the spectrum of the coefficient matrix lies in some confined region in the complex plane that excludes the origin. Chebyshev iteration [44, 25, 26, 21] is among the first type of modern iterative methods developed for the non-Hermitian case. It is suitable for well-conditioned cases where the spectrum is enclosed in an ellipse some distance away from the origin. Later, alternatives using least-squares polynomials [33] were designed by minimizing some weighted L2 norm of the residual on a polygon that encloses the spectrum. At the same time Krylov subspace methods, such as GMRES [37], were found appealing as they did not require any prior spectral information. These methods converge well if the numerical range is not close to the origin [12, 6]. For (modified) Hermitian and skew-Hermitian splitting methods [4, 3], the spectrum is assumed to be in one of the four quadrants. However, the performance of these methods in solving (1.2) is usually unacceptably poor due to the unfavorable spectrum of the discretized Helmholtz operators.

In order to obtain iterative schemes that converge fast when solving (1.2), several efficient preconditioners such as the optimized Schwarz method [7, 18, 17], PML sweeping preconditioners [13, 41, 47], and shifted Laplacian preconditioners [5, 14, 22] have recently been proposed. These preconditioners are much more expensive to construct than those for solving standard elliptic PDEs. We highlight specifically the idea of shifted Laplacian preconditioners [14] which relies on the fact it is easier to solve linear systems with the shifted matrix $A-zI$, for some complex number z, than with the original matrix A. The idea of complex shifts is generalized in [46] based on contour integration formulations that are formerly used in eigenvalue computations [43, 45]. The method draws a circular contour in the complex plane to decompose the spectrum of A^{-1}, and the decomposed subproblems are then solved separately.

The application of shifted Laplacian–type preconditioners $A-zI$ is usually based on extending standard elliptic solvers to complex matrices. Two popular choices are incomplete LU (ILU) [32, 30] and multigrid methods [27]. One major issue associated with these preconditioners is that their performance deteriorates dramatically as the angular frequency ω increases. This is because on the one hand, a small magnitude of z is necessary for convergence when solving high-frequency problems [19], and on the other hand, a small $|z|$ will significantly increase the computational burden to approximate $(A-zI)^{-1}$: standard multigrid methods are no longer guaranteed to be effective [10], and ILU factors become dense and even unstable. The aim of this paper is to propose an efficient and robust preconditioning technique to overcome these difficulties.

1.3. Outline of the proposed method. In this paper, we solve the three-dimensional (3D) Helmholtz problems in a contour integration framework adapted from [46], with a new fixed-point iteration for the shifted systems. The fixed-point iteration is based on a polynomial approximation of the matrix exponential, which is suitable for the case when the spectrum is confined in a rectangle with a small separation away from the origin and the standard Chebyshev iteration diverges. Compared with existing methods for solving shifted problems, the new approach is robust and has a fixed storage requirement even if the imaginary part of the shift nears the origin. In the proposed contour integration framework, the fixed-point iteration is used to resolve components of the subproblem associated with large eigenvalues of A, and GMRES is used to resolve the remainder.

For the Helmholtz equation with the impedance boundary condition, we show for an idealized case that inside some contour the imaginary part of each eigenvalue is well separated from the real axis. Then, iA may have a positive definite Hermitian part for the problem inside the contour, in which case GMRES converges fast. We give several techniques to improve the effectiveness of the solver and demonstrate the performance of the proposed scheme for challenging high-frequency variable-coefficient problems in three dimensions. It is well known that spectral methods require fewer grid points per wavelength relative to finite difference and finite element methods [39]. Since the proposed method only accesses the matrix through matrix-vector products, it is ideally suited for solving dense linear systems resulting from spectral discretizations, for which matrix-vector products can be performed with almost linear complexity.

The remaining sections are organized as follows. In section 2, we review the contour integration framework for general indefinite linear systems. In section 3, we characterize the spectrum of the Helmholtz problem based on linear algebra assumptions. A fixed-point iteration method is developed in section 4 to solve shifted problems. In section 5, the distribution of eigenvalues of the interior impedance problem is studied. Some useful techniques are provided in section 6 to achieve an optimal performance of the proposed method. In section 7, numerical examples are presented using Fourier spectral and finite difference methods. Conclusions are drawn in section 8.

The following notation will be used throughout the remaining sections:

- Range(G) and Null(G) denote the range and null space, respectively, of a matrix G;
- $\rho(G)$ represents the spectral radius of a matrix G;
- $G \succ (\succeq)\ 0$ means G is Hermitian positive (semi-)definite.

2. Review of the contour integration framework.

The inverse of a matrix A can be approximated by a linear combination of the resolvent $(A - zI)^{-1}$ with several complex shifts [35, 46]. In this section, we provide theoretical justifications of the key ideas in [46] and also suggest some new improvements.

In [46], the authors only consider circular contours. Here, we first generalize their results to arbitrary contours. Let γ be a closed piecewise smooth Jordan curve in the complex plane $\mathbb{C}$ that encloses the origin and such that no eigenvalue of A lies on γ. Then the eigenprojector P associated with the eigenvalues *outside* γ can be expressed as

$$P = \frac{1}{2\pi \mathrm{i}} \int_\gamma \left(I - zA^{-1}\right)^{-1} \frac{\mathrm{d}z}{z}, \tag{2.1}$$

where the integral is taken counterclockwise on γ. This is because

$$P = \frac{-1}{2\pi \mathrm{i}} \int_\gamma \left(\frac{1}{z} - A^{-1}\right)^{-1} \mathrm{d}\frac{1}{z} = \frac{1}{2\pi \mathrm{i}} \int_{\gamma^{-1}} \left(z' I - A^{-1}\right)^{-1} \mathrm{d}z', \tag{2.2}$$

where $\gamma^{-1} = \{z^{-1} : z \in \gamma\}$ and the last integral is taken counterclockwise on γ^{-1}. The right-hand side of (2.2) takes the standard form of an eigenprojector of A^{-1} associated with eigenvalues enclosed by γ^{-1}; see, for example, [35, Theorem 3.3]. Assuming that the $1/\lambda_i$'s are those eigenvalues of A^{-1} enclosed by γ^{-1}, we then have

$$\mathrm{Range}(P) = \bigoplus_i \mathrm{Null}\left(\frac{1}{\lambda_i} I - A^{-1}\right)^{l_i} = \bigoplus_i \mathrm{Null}\left(\lambda_i I - A\right)^{l_i},$$

where l_i is the *index* of λ_i [35, section 1.8.2]. The above equality implies that P in (2.1) is equal to the spectral projector of A associated with eigenvalues outside γ.

The method proposed in [46] is based on the Cauchy integral representation of PA^{-1}:

$$PA^{-1} = \frac{1}{2\pi \mathrm{i}} \int_{\gamma} (A - zI)^{-1} \frac{\mathrm{d}z}{z}. \tag{2.3}$$

Again, the integral is taken counterclockwise on γ. PA^{-1} ignores the eigenvalues of A that are inside γ, and attempts to solve the restricted problem in the subspace spanned by the eigenvalues outside γ. This restricted problem is referred to as the *outer problem*.

If a numerical quadrature rule is applied to discretize the right-hand side of (2.3), PA^{-1} can be approximated as

$$PA^{-1} \approx \sum_i \frac{\sigma_i}{z_i} (A - z_i I)^{-1}, \tag{2.4}$$

where $\{z_i\}$ are the quadrature nodes along γ and $\{\sigma_i\}$ are the corresponding weights.

For a given right-hand side f, the method proposed in [46] first approximates $PA^{-1}f \approx w := \sum_i \frac{\sigma_i}{z_i}(A - z_i I)^{-1} f$ and then tries to minimize the residual of the solution in the range of $I - P$ with an iterative method,

$$\min_v \|Av - (f - Aw)\|_2. \tag{2.5}$$

This problem focuses on eigenvalues inside γ and is referred to as the *inner problem*.

Afterward, $v + w$ serves as an approximate solution. The problem (2.5) is easier to solve than the original problem because the spectrum is restricted inside γ. This framework has some flexibilities regarding the selection of contours and quadrature points and is summarized in Algorithm 2.1. Since the linear system is not solved to high accuracy in a single run of FCI, it is necessary to use flexible iterative methods, such as FGMRES [36], or iterative refinement to improve the accuracy.

Regarding line 4 of Algorithm 2.1, we introduce a scalar multiplier d to further reduce the two-norm residual error. If $Aw \neq 0$, the objective function has a unique minimizer at $d = \frac{(Aw)^H f}{(Aw)^H Aw}$. It takes one matrix-vector product and two inner products

Algorithm 2.1. Fast contour integration approximation of $A^{-1}f$

```
1: procedure FCI ($f \in \mathbb{C}^n, A \in \mathbb{C}^{n \times n}, \{z_i \in \gamma\}, \{\sigma_i \in \mathbb{C}\}$)
                    ▷ $z_i$ and $\sigma_i$ are quadrature points and weights on a contour $\gamma$
2:     Solve $(A - z_i I) y_i = f$ for each quadrature point $z_i$
3:     Approximate $PA^{-1}f$ with a quadrature
           $w = \sum_i \frac{\sigma_i}{z_i} y_i$
4:     Compute a scalar multiplier $d$ as follows to compensate quadrature error
           $d = \mathrm{argmin}_{d \in \mathbb{C}} \|f - dAw\|_2 = \frac{(Aw)^H f}{(Aw)^H Aw}$
5:     Solve $v = \mathrm{argmin}_v \|Av - (f - dAw)\|_2$ with a few steps of GMRES
6:     return the approximate solution $v + dw$
7: end procedure
```

to compute d. Theoretically speaking, when P is an orthogonal projection and w exactly equals $PA^{-1}f$, this step has no effect because

$$\frac{(Aw)^H f}{(Aw)^H Aw} = \frac{f^H P^H f}{f^H P^H P f} = 1.$$

In practice, this step becomes meaningful because it makes the method more robust to quadrature error.

In the following sections, we will discuss how to maximize the efficiency of Algorithm 2.1 to solve (1.2) by exploiting the spectral properties of the discretized Helmholtz operators.

3. Eigenvalue distribution of the discretized Helmholtz operators. In order to apply the FCI preconditioner (Algorithm 2.1) to solve the linear system (1.2), the spectrum information of the coefficient matrix is crucial for the selection of the contour as well as the resulting preconditioning effect. In this section, we will systematically study the eigenvalue distribution of the discretized Helmholtz operators as well as some of its variants. More specifically, we will investigate the spectrum of two types of matrices: (1) the coefficient matrix A in (1.2) and (2) a related double-size matrix.

Case One. For this simplest case, Algorithm 2.1 is applied to solve (1.2) directly. The skew-Hermitian part of the coefficient matrix A comes from absorbing boundary conditions or various types of damping. Here we assume the skew-Hermitian part of A is $-\mathrm{i}$ multiplied by a positive semidefinite matrix. That is,

$$A = A_1 - \mathrm{i}A_2,$$

where both A_1 and A_2 are Hermitian, and A_2 is positive semidefinite. This assumption has previously appeared in [22, equation (12)] and also in [19, equation (1.7)]. Under this assumption, it is easy to characterize the spectrum of A as follows.

LEMMA 3.1. *If the Hermitian matrices A_1, A_2 satisfy $A_1 + I \succeq 0$ and $A_2 \succeq 0$, then the spectrum of $A = A_1 - \mathrm{i}A_2$ is contained in the closed rectangle*

$$\{\lambda \in \mathbb{C} : \mathrm{Re}(\lambda) \in [-1, \rho_1 - 1], \quad \mathrm{Im}(\lambda) \in [-\rho_2, 0]\},$$

where $\rho_1 = \rho(A_1 + I)$ and $\rho_2 = \rho(A_2)$.

Proof. Let v be a unit right eigenvector of A. Then the corresponding eigenvalue λ satisfies

$$\lambda = v^H A v = v^H A_1 v - \mathrm{i} v^H A_2 v.$$

This implies that

$$\mathrm{Re}(\lambda) = v^H A_1 v \in [-1, \rho_1 - 1], \quad \mathrm{Im}(\lambda) = -v^H A_2 v \in [-\rho_2, 0]. \qquad \square$$

Lemma 3.1 can also be derived from Theorem 1 in [8].

Remark 3.1. The assumption $A_1 + I \succeq 0$ is not essential. One can always normalize the matrix A first to make the assumption hold. This normalization does not affect the conditioning of the matrix or the relative (residual) accuracy of the solution, but either the absolute error or the absolute residual error is changed. In Lemma 3.1, ρ_1 and ρ_2 represent the horizontal and the vertical stretch of the spectrum, respectively, and ρ_1/ρ_2 measures the aspect ratio of this rectangle.

Case Two. We now consider a double-size linear system:

$$(3.1)\qquad (\mathrm{i}C - I)\begin{pmatrix}\mathrm{i}u\\ u\end{pmatrix} = \begin{pmatrix}0\\ f\end{pmatrix},\quad C = \begin{pmatrix} & I\\ -(A_1+I) & -A_2\end{pmatrix}.$$

One can check that u in (3.1) is exactly the solution of (1.2). We can apply Algorithm 2.1 to solve the system (3.1) instead. Although the size of the coefficient matrix $\mathrm{i}C-I$ in (3.1) is twice as large as that of (1.2), it could be less costly to solve (3.1) than (1.2). This is because the spectrum of $\mathrm{i}C - I$ can be more compact than that of A under some discretization schemes, which is analyzed in the following theorem.

Theorem 3.2. *Following the same assumption as in Lemma* 3.1*, the spectrum of the matrix* $\mathrm{i}C - I$ *defined in* (3.1) *is contained in the rectangle*

$$\left\{\mu\in\mathbb{C} : |\operatorname{Re}(\mu)+1| \le \frac{\rho_2}{2} + \sqrt{\left(\frac{\rho_2}{2}\right)^2 + \rho_1},\quad \operatorname{Im}(\mu)\in[-\rho_2, 0]\right\},$$

where ρ_1 *and* ρ_2 *are defined in Lemma* 3.1*. Furthermore, if* $\rho_2 = 0$*, then the set of eigenvalues of* $\mathrm{i}C - I$ *is*

$$\{-1 \pm \sqrt{\lambda_i} : \lambda_i \textit{ is an eigenvalue of } A_1 + I\}.$$

Proof. It suffices to prove that the spectrum of the matrix C defined in (3.1) is contained in

$$\left\{\mu\in\mathbb{C} : |\mu| \le \frac{\rho_2}{2} + \sqrt{\left(\frac{\rho_2}{2}\right)^2 + \rho_1},\quad \operatorname{Re}(\mu)\in[-\rho_2, 0]\right\},$$

and $\{\pm\mathrm{i}\sqrt{\lambda_i} : \lambda_i$ is an eigenvalue of $A_1 + I\}$ is the set of eigenvalues of C for $\rho_2 = 0$. If μ is a nonzero eigenvalue of C, then the Schur complement S of $\mu I - C$

$$S = \mu I + A_2 + \mu^{-1}(A_1 + I)$$

has to be singular.

Denote μS by E. For a nonzero vector v in the null space of E, we have

$$0 = \left|\frac{v^H E v}{v^H v}\right| \ge |\mu^2| - |\mu|\rho_2 - \rho_1.$$

Thus, $|\mu| \le \frac{\rho_2}{2} + \sqrt{\left(\frac{\rho_2}{2}\right)^2 + \rho_1}$.

If μ is real, then for any vector w, we have

$$\left|\frac{w^H E w}{w^H w}\right| \ge |\mu^2| - |\mu|\rho_2.$$

This implies that $\mu \in [-\rho_2, 0]$, because otherwise $E \succ 0$ is a contradiction. If μ has a nonzero imaginary part, then the Hermitian part of $E/(\operatorname{Im}(\mu)\mathrm{i})$ is

$$\frac{\operatorname{Im}(\mu^2)}{\operatorname{Im}(\mu)} I + A_2 = 2\operatorname{Re}(\mu) I + A_2,$$

which is positive definite for $\operatorname{Re}(\mu) > 0$ and is negative definite for $\operatorname{Re}(\mu) < -\rho_2/2$. Therefore, $\operatorname{Re}(\mu) \in [-\rho_2, 0]$.

Finally, we consider the special case when $\rho_2 = 0$. Let $V^{-1}\Lambda V$ be the diagonalization of $A_1 + I$. If $\rho_2 = 0$, then $A_2 = 0$ and we have

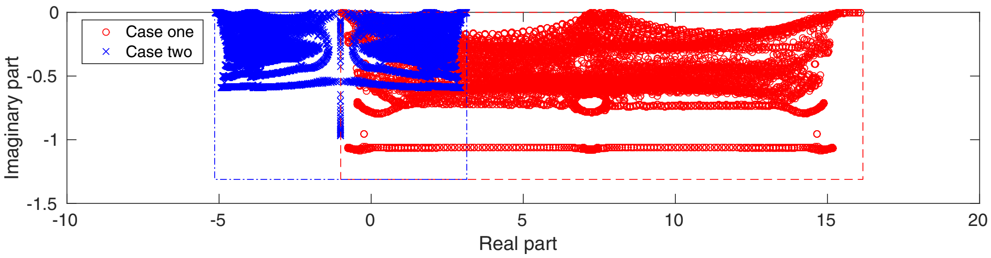


FIG. 3.1. *Comparison of the spectrum between Case One* (1.2) *and Case Two* (3.1). *The test matrix is based on the finite difference method on a* 100^2 *grid, with absorbing layers near the boundary. The eigenvalues are computed by* **eig** *in MATLAB. The rectangles are the results of Lemma* 3.1 *and Theorem* 3.2, *where* $\rho_1 \approx 17.2$, $\rho_2 \approx 1.3$. *Case Two shows the spectrum of the matrix* $\mathrm{i}C - I$.

$$\begin{pmatrix} V & \\ & V \end{pmatrix} C \begin{pmatrix} V^{-1} & \\ & V^{-1} \end{pmatrix} = \begin{pmatrix} & I \\ -\Lambda & \end{pmatrix}.$$

For the characteristic polynomial, we have

$$\det(\mu I - C) = \det \begin{pmatrix} \mu I & -I \\ \Lambda & \mu I \end{pmatrix} = \prod_i (\mu^2 + \lambda_i).$$

This shows $\{\pm \mathrm{i}\sqrt{\lambda_i}\}$ are the eigenvalues of C when $\rho_2 = 0$. □

Remark 3.2. We can compare $\mathrm{i}C - I$ with the matrix A in terms of the spreading of spectrum. The spectrum of $\mathrm{i}C - I$ is contained in a $\rho_2 + \sqrt{\rho_2^2 + 4\rho_1}$ by ρ_2 rectangle, but the spectrum of A is contained in a ρ_1 by ρ_2 rectangle. A can have a more elongated spectrum when ρ_1/ρ_2 is large. Therefore, although $\mathrm{i}C - I$ is double in size, it may be more suitable for iterative solvers because the spectrum is less spread out. See Figure 3.1 for a 2D example.

4. Polynomial preconditioners for solving shifted problems.

The application of Algorithm 2.1 to solve discretized Helmholtz equations involves several linear system solutions with shifted problems:

$$(A - zI)y = f. \tag{4.1}$$

If $|\operatorname{Im}(z)|$ is large enough, then according to the previous section $\mathrm{i}(A - zI)$ has a sign definite Hermitian part, and many existing elliptic solvers or preconditioners can be used. However, they become expensive to compute or to store as $|\operatorname{Im}(z)|$ reduces. In this section, we will propose several efficient polynomial preconditioning techniques to solve (4.1) even when $|\operatorname{Im}(z)|$ is relatively small.

We can write the general form of a *polynomial fixed-point iteration* of (4.1) as

$$y^{(m+1)} = y^{(m)} + p(A - zI)r^{(m)}, \tag{4.2}$$

where p is a polynomial, and $r^{(m)} = f - (A - zI)y^{(m)}$ is the residual at the mth step. If all the roots of the polynomial p are known explicitly, then (4.2) can be rewritten as a cyclic Richardson iteration. Motivated from Lemma 3.1 in section 3, we assume in this section that the spectrum of A is contained in a closed rectangle $\mathcal{B}$. For fixed z, we define

$$R(\lambda) = 1 - (\lambda - z)p(\lambda - z). \tag{4.3}$$

R is usually called the *residual polynomial.* A desirable polynomial p should solve the following minimax problem in some special set of polynomials denoted by $\mathcal{P}$:

$$\min_{p\in\mathcal{P}} \max_{\lambda\in\mathcal{B}} |R(\lambda)| = \min_{p\in\mathcal{P}} \max_{\lambda\in\partial\mathcal{B}} |R(\lambda)|. \tag{4.4}$$

The polynomial p can either be used directly in the fixed-point iteration (4.2) or be used as a preconditioner in Krylov methods.

4.1. Stationary Richardson iteration. The scheme (4.2) is a stationary Richardson iteration if p is a complex constant. The minimax problem (4.4) is essentially solved by the method in [29]. Here, we present a slight generalization to the case of rectangular spectrum. The following theorem shows how to choose the complex constant for optimal convergence rate.

THEOREM 4.1. *Let $\mathcal{B} \subset \mathbb{C}$ be the rectangle with vertices $\{b_1, b_2, \beta_1, \beta_2\}$ such that*

$$\mathrm{Im}(b_j) = 0, \quad \mathrm{Im}(\beta_j) < 0, \quad j \in \{1, 2\}.$$

Assuming that $p \in \mathbb{C}$, the minimax value of $R(\lambda)$ in (4.3)–(4.4) *is taken on the vertices*

$$\min_{p\in\mathbb{C}} \max_{\lambda\in\mathcal{B}} |R(\lambda)| = \min_{p\in\mathbb{C}} \max_{\lambda\in\{b_1,b_2,\beta_1,\beta_2\}} |R(\lambda)|.$$

Furthermore, if $\mathrm{Im}(z) \notin [\mathrm{Im}(\beta_1), 0]$ *and z is enclosed by the circumcircle of $\mathcal{B}$, then the minimax value of $R(\lambda)$ equals $\frac{|\alpha_1-\alpha_2|}{|\alpha_1|+|\alpha_2|}$ at $p^* = \frac{|\alpha_1|/\alpha_1+|\alpha_2|/\alpha_2}{|\alpha_1|+|\alpha_2|}$, where*

$$\alpha := (\alpha_1, \alpha_2) = \begin{cases} (b_1 - z, b_2 - z) & \text{if } \mathrm{Im}(z) > 0, \\ (\beta_1 - z, \beta_2 - z) & \textit{otherwise.} \end{cases}$$

Proof. Let $o = 1/p + z$ be the root of the residual polynomial $R(\lambda) = 1-(\lambda-z)p$. The absolute value of $R(\lambda)$ is related to the distance from o

$$|R(\lambda)| = |1 - (\lambda - z)p| = \left|1 - \frac{\lambda - z}{o - z}\right| = \frac{|\lambda - o|}{|z - o|}.$$

For fixed z and o, $|R(\lambda)|$ is convex, so the maximum value on each line segment is taken on a vertex. $\partial\mathcal{B}$ has four sides, and the maximum value is taken on one of the four vertices.

For the remaining part of the theorem, it suffices to prove for the case $\mathrm{Im}(z) > 0$ because the other case follows from symmetry. The minimax problem restricted to a line segment is solved in [29, Example 5.1], which suggests

$$\min_{p\in\mathbb{C}} \max_{\lambda\in[b_1,b_2]} |1 - (\lambda - z)p| = \frac{|\alpha_1 - \alpha_2|}{|\alpha_1| + |\alpha_2|} < 1,$$

and the optimal value of p is p^*. For circles containing b_1, b_2, the one centered at $o^* = 1/p^* + z$ does not contain z because $|b_1 - o^*| = |b_2 - o^*| < |z - o^*|$, but the circumcircle of $\mathcal{B}$ contains z. So, o^* is closer to β_1, β_2 than to b_1, b_2. That is,

$$|\beta_1 - o^*| = |\beta_2 - o^*| < |b_1 - o^*| = |b_2 - o^*|.$$

Therefore,

$$\frac{|b_1 - o^*|}{|z - o^*|} = \min_{p\in\mathbb{C}} \max_{\lambda\in\{b_1,b_2,\beta_1,\beta_2\}} |R(\lambda)| = \min_{p\in\mathbb{C}} \max_{\lambda\in\mathcal{B}} |R(\lambda)|.$$

p^* solves the minimax problem (4.4) for $\mathcal{B}$. □

Remark 4.1. If $b_1 b_2 < 0$, which means the matrix A in (4.1) is indefinite, for an imaginary shift z (or $|\operatorname{Re}(z)|$ is much smaller than $|b_1|$ or $|b_2|$), the convergence rate is close to

$$\frac{|\alpha_1 - \alpha_2|}{|\alpha_1| + |\alpha_2|} = \frac{|b_1| + |b_2|}{\sqrt{b_1^2 + \operatorname{Im}^2(z)} + \sqrt{b_2^2 + \operatorname{Im}^2(z)}}.$$

Using the Taylor expansion of $\sqrt{1+x^2}$, one can check that an $O(|\operatorname{Im}(z)|^{-2})$ number of iterations is needed to reach certain relative accuracy. This result can be improved by considering high-order polynomials.

4.2. High-order polynomials. High-order polynomials $p(\lambda - z)$ have the capability to improve the convergence rate. For well-conditioned problems, existing work such as Chebyshev iterations [44, 25, 26, 21] and Leja points [31] can select the roots of the polynomial near the spectrum for asymptotic optimal convergence. Some more advanced polynomial preconditioners for solving a sequence of shifted linear systems can be found in [1, 16]. Since the shifted system (4.1) may not be sufficiently well conditioned, we will design such a polynomial from the approximation of the exponential function.

For given z, the residual polynomial $R(\lambda)$ defined in (4.3) can be reformulated as

$$R(\lambda) = \frac{\tilde{p}(\lambda)}{\tilde{p}(z)} \tag{4.5}$$

for some polynomial $\tilde{p}$ because this form also takes the value of 1 at $\lambda = z$. The ideal polynomial should yield small value of $|R(\lambda)|$ at every eigenvalue λ.

The choice of $\tilde{p}$ is motivated from the exponential function. If $\operatorname{Im}(\lambda - z)$ has a fixed sign for $\lambda \in \mathcal{B}$, then with a suitable choice of $\delta \in \mathbb{R}$, the following quantity can be arbitrarily small:

$$\frac{|e^{-\mathrm{i}\delta\lambda}|}{|e^{-\mathrm{i}\delta z}|} = e^{\delta \operatorname{Im}(\lambda - z)} \ll 1.$$

Choosing $\tilde{p}(\lambda) \approx e^{-\mathrm{i}\delta\lambda}$ can possibly reduce $|R(\lambda)|$. The simplest choice of $\tilde{p}$ is based on the Taylor expansion of the exponential function

$$\tilde{p}^{(q)}(\lambda) = \sum_{j=0}^{q} \frac{(-\mathrm{i}\delta(\lambda - z_0))^j}{j!}, \tag{4.6}$$

where z_0 is the center of the Taylor expansion. Then the polynomial p of degree $q-1$ in (4.2) has the form

$$p(\lambda - z) = \left(1 - \frac{\tilde{p}^{(q)}(\lambda)}{\tilde{p}^{(q)}(z)}\right) \Big/ (\lambda - z). \tag{4.7}$$

The explicit form of $p(\lambda - z)$ is quite complex, but there is a recursive expression that makes it easier to compute. Notice that

$$\tilde{p}^{(q)}(z) p(\lambda - z) = \frac{\tilde{p}^{(q)}(z) - \tilde{p}^{(q)}(\lambda)}{\lambda - z} = \sum_{j=1}^{q} \kappa_j,$$

where κ_j is defined as $-\frac{(-\mathrm{i}\delta)^j}{j!}\sum_{l=0}^{j-1}(\lambda - z_0)^l(z-z_0)^{j-1-l}$. It is easy to see that

$$\begin{aligned}\kappa_1 &= \mathrm{i}\delta,\\ \kappa_j &= \frac{-\mathrm{i}\delta}{j}\left[(\lambda - z_0)\kappa_{j-1} + \frac{(-\mathrm{i}\delta(z-z_0))^{j-1}}{(j-1)!}\right], \quad j \in \{2,3,\ldots,q\}.\end{aligned}$$

Therefore, for the specific choice (4.6)–(4.7), the fixed-point iteration (4.2) can be rewritten as

$$\begin{aligned}k_1 &= \mathrm{i}\delta r^{(m)},\\ k_j &= \frac{-\mathrm{i}\delta}{j}\left((A - z_0 I)k_{j-1} + \frac{(-\mathrm{i}\delta(z-z_0))^{j-1}}{(j-1)!}r^{(m)}\right), \quad j \in \{2,3,\ldots,q\},\\ y^{(m+1)} &= y^{(m)} + \frac{1}{\tilde{p}^{(q)}(z)}\sum_{j=1}^{q} k_j.\end{aligned} \tag{4.8}$$

For solving (4.1), the error at the mth step satisfies

$$y^{(m)} - y = \left(\frac{\tilde{p}^{(q)}(A)}{\tilde{p}^{(q)}(z)}\right)^m \left(y^{(0)} - y\right).$$

The optimal parameters z_0 and δ in the scheme (4.8) can be computed by solving an optimization problem for each fixed polynomial degree q. Here we propose a heuristic to simplify this procedure. We choose z_0 to guarantee robustness and δ for fast convergence. By robustness we mean for sufficiently small $|\delta|$, the spectral radius $\rho(\tilde{p}^{(q)}(A)/\tilde{p}^{(q)}(z))$ is less than 1. This is done by considering the following equation:

$$\left|\frac{\tilde{p}^{(q)}(\lambda)}{\tilde{p}^{(q)}(z)}\right|^2 = \left|\frac{1-\mathrm{i}\delta(\lambda - z_0)}{1-\mathrm{i}\delta(z-z_0)}\right|^2 + o(|\delta|) = \frac{1+2\delta\,\mathrm{Im}(\lambda - z_0)}{1+2\delta\,\mathrm{Im}(z-z_0)} + o(|\delta|).$$

If we fix $\mathrm{Im}(z_0) = \mathrm{Im}(z)$ and assume $\mathrm{Im}(\lambda - z_0)$ has a fixed sign, then we can always find some δ with a small absolute value such that $\delta\,\mathrm{Im}(\lambda - z_0) < 0$, which controls the spectral radius. Since $z_0 = (b_1 + b_2)/2 + \mathrm{i}\,\mathrm{Im}(z)$ is the optimal choice for $q = 1$, we will always follow this choice for high-order polynomials. After that, δ is determined by numerically minimizing the convergence rate

$$\nu = \min_{\delta\in\mathbb{R}}\max_{\lambda\in\partial\mathcal{B}}\left|\frac{\tilde{p}(\lambda)}{\tilde{p}(z)}\right|. \tag{4.9}$$

This is a 2D optimization problem which is easy to solve. Table 4.1 compares the convergence rate ν for different order q. We prefer choosing q with a minimum $\nu^{1/q}$ value since this quantity gives the fastest converging method for a given number of matrix-vector products.

Table 4.1

Convergence rate of (4.8) *for different order q. The spectrum of A in* (4.1) *is within a rectangle $\mathcal{B}$ with the real part in* $[-1, 2.8]$ *and the imaginary part in* $[-0.65, 0]$*, and the shift is* $z = \mathrm{i}$*. The test matrix is based on Fourier spectral method on a* 80^3 *grid, with absorbing layers near the boundary. δ^* is the optimal choice of δ. ν is the estimated convergence rate, and $\nu^{1/q}$ quantifies the convergence rate per matrix-vector product.*

q	δ^*	ν	$\nu^{1/q}$
1	0.250	0.866	0.866
2	0.688	0.537	0.733
3	0.750	0.530	0.810
4	0.750	0.547	0.860
5	0.938	0.416	0.839

Table 4.2

Estimated number of matrix-vector products that reduces the residual by 10^2 *times. The shifts are purely imaginary. The test matrix A is the same as Table* 4.1.

	$z = \mathrm{i}$	$z = \mathrm{i}/2$	$z = \mathrm{i}/4$	$z = \mathrm{i}/8$
$q = 1$	33	110	420	1656
$q = 2$	16	38	110	304
$q = 3$	24	30	60	135
$q = 4$	32	44	72	128
$q = 5$	30	60	120	180

We are also concerned with how the cost depends on the shift z. The proof of Theorem 4.1 mentions a symmetry argument that by choosing the horizontal symmetry axis of $\mathcal{B}$ as the axis of reflection, any point z above the axis is equivalent with another point below the axis. The solution method (4.8) is also symmetric when z and z_0 are simultaneously reflected. Therefore, it suffices to study points above the horizontal symmetry axis of $\mathcal{B}$, which are points with positive imaginary parts.

As z approaches an example $\mathcal{B}$ along the positive imaginary axis, Table 4.2 gives the estimated number of matrix-vector products for reducing the residual by 10^2 times. The estimated number is $q\lceil \log 10^{-2}/\log\nu\rceil$, where ν is defined in (4.9) and $\lceil\cdot\rceil$ is the ceiling function. The cost of the stationary Richardson iteration ($q = 1$) quadruples as the distance reduces by $1/2$. For high-order methods such as $q = 3$, the results are much better. The cost roughly doubles when the imaginary shift reduces by $1/2$. As the imaginary shift decreases, one might want to increase q slightly to approach a desirable performance. In practice, we choose q to have the best convergence by solving

$$\min_{q\in\mathbb{N}^+}\min_{\delta\in\mathbb{R}}\max_{\lambda\in\partial\mathcal{B}}\left|\frac{\tilde{p}^{(q)}(\lambda)}{\tilde{p}^{(q)}(z)}\right|.$$

Table 4.3 shows the costs of solving several complex shifted problems. The estimated costs in Table 4.3(b) match well with the actual costs in Table 4.3(c) in a sample run. The estimate is reliable and insensitive to the right-hand sides because it directly comes from solving the minimax problem (4.4). One can see that the real part of the shift plays a minor role in determining the cost.

Finally, we compare the cost of solving a pair of shifted problems corresponding to the two cases (1.2) and (3.1). In order to draw a fair comparison, we force the shifted problems to be equivalent. Let A be a Hermitian indefinite matrix and complex numbers z and s satisfy $z + 1 = (s + 1)^2$. Then the pair of shifted problems are

$$(A - zI)y = f \tag{4.10}$$

Table 4.3

*Costs of solving shifted problems for general complex shifts. The number of matrix-vector products (*mvs*) that reduces the residual by* 10^2 *times is both estimated and tested. The optimal order q is selected. The test matrix A is the same as Table* 4.1*. The right-hand side has a single nonzero value at grid location* $(40, 40, 40)$.

(a) Complex shift z

$-1/2+\mathrm{i}$	$-0.25+\mathrm{i}$	i	$0.25+\mathrm{i}$	$0.5+\mathrm{i}$
$-1/2+0.75\mathrm{i}$	$-0.25+0.75\mathrm{i}$	$0.75\mathrm{i}$	$0.25+0.75\mathrm{i}$	$0.5+0.75\mathrm{i}$
$-0.5+0.5\mathrm{i}$	$-0.25+0.5\mathrm{i}$	$0.5\mathrm{i}$	$0.25+0.5\mathrm{i}$	$0.5+0.5\mathrm{i}$
$-0.5+0.25\mathrm{i}$	$-0.25+0.25\mathrm{i}$	$0.25\mathrm{i}$	$0.25+0.25\mathrm{i}$	$0.5+0.25\mathrm{i}$

(b) Estimated mvs for shifts from (a)

14	14	16	16	16
18	20	22	22	22
33	33	30	30	30
66	66	60	54	54

(c) Actual mvs for shifts from (a)

12	12	14	14	14
18	18	20	20	20
33	30	30	27	27
63	60	54	51	51

Table 4.4

Number of matrix-vector products for reducing the residual by 10^2 *times. Boldface numbers indicate a superior performance over the alternative case.*

(a) Case One: solving (4.10)

Spectrum	$z=\mathrm{i}$	$z=\mathrm{i}/2$	$z=\mathrm{i}/4$	$z=\mathrm{i}/8$
$[-1,8]$	**28**	**67**	**156**	**276**
$[-1,16]$	**52**	**145**	381	721
$[-1,32]$	**97**	436	861	1691
$[-1,64]$	331	916	1831	3736

(b) Case Two: solving (4.11)

Spectrum	$z=\mathrm{i}$	$z=\mathrm{i}/2$	$z=\mathrm{i}/4$	$z=\mathrm{i}/8$
$[-1,8]$	67	121	209	441
$[-1,16]$	85	157	**229**	**609**
$[-1,32]$	113	**193**	**397**	**681**
$[-1,64]$	**149**	**221**	**449**	**909**

and

$$\begin{pmatrix} -(s+1)I & \mathrm{i}I \\ -\mathrm{i}(A+I) & -(s+1)I \end{pmatrix} \begin{pmatrix} \mathrm{i}y \\ (s+1)y \end{pmatrix} = \begin{pmatrix} 0 \\ f \end{pmatrix}. \tag{4.11}$$

The results are tabulated in Table 4.4. Note that the cost of each matrix-vector product is roughly the same. (4.10) is more suitable for the case that the spectrum of A is compact and the imaginary shift is large; otherwise (4.11) is better suited.

5. Characterization of small eigenvalues.

In this section, we analyze the distribution of eigenvalues with small magnitude for the Helmholtz equation. It is important to study near-zero eigenvalues because they govern the conditioning of the problem. A stability estimate is proved in [28, 11, 19] for the constant-coefficient Helmholtz equation in a star-shaped domain Ω with the impedance boundary condition.

The part of the statement we want to highlight is that

$$\omega\|u\|_{L^2(\Omega)} \leq \alpha\|f\|_{L^2(\Omega)},$$

where u is the solution of the right-hand-side f, and α is a constant independent from the angular frequency ω. This result implies that *all the eigenvalues of A are at least $O(\omega)$ distance away from the origin.*

For the constant-coefficient case in a hypercube, we can further describe the distribution of eigenvalues with the impedance boundary condition. In fact, we can show that *the imaginary part is at least $O(\omega)$ for eigenvalues with magnitude smaller than ω^2.* Let the wavespeed be normalized as one in the d-dimensional unit hypercube $[0,1]^d$. Consider the eigenvalue problem in multiple dimensions ($d \geq 2$), and an eigenpair (λ, v) satisfies

$$\begin{cases} -\Delta v - \omega^2 v = \lambda v & \text{in } [0,1]^d, \\ \partial_n v - \mathrm{i}\omega v = 0 & \text{on } \partial[0,1]^d, \end{cases} \tag{5.1}$$

where ∂_n means taking the directional derivative along the outward unit normal.

THEOREM 5.1. *For any fixed $\rho \in (0,1)$, there exists a positive constant s such that every eigenvalue λ of* (5.1) *with $|\lambda| \leq \rho\omega^2$ satisfies $|\operatorname{Im}(\lambda)| \geq s\omega$ for sufficiently large ω.*

The theorem gives a more detailed characterization of small eigenvalues. Figure 5.1 illustrates how this compares with existing stability results. The proof is based on a separation of variables in the following form:

$$v(x) = \prod_{j=1}^{d} \varphi_j(x_j), \quad \varphi_j''(x_j) + \xi_j^2 \varphi_j(x_j) = 0, \tag{5.2}$$

where $\xi_j \in \mathbb{C}$ and $\operatorname{Re}(\xi_j) \geq 0$. An eigenvalue can be written as $\lambda = \xi \cdot \xi - \omega^2$, where $\xi = \begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_d \end{pmatrix}$. The general solution of (5.2) is $\varphi_j(x_j) = a_j^+ e^{\mathrm{i}\xi_j x_j} + a_j^- e^{-\mathrm{i}\xi_j x_j}$. The boundary condition in (5.1) suggests

$$-\varphi_j'(0) = \mathrm{i}\omega\varphi_j(0), \quad \varphi_j'(1) = \mathrm{i}\omega\varphi_j(1). \tag{5.3}$$

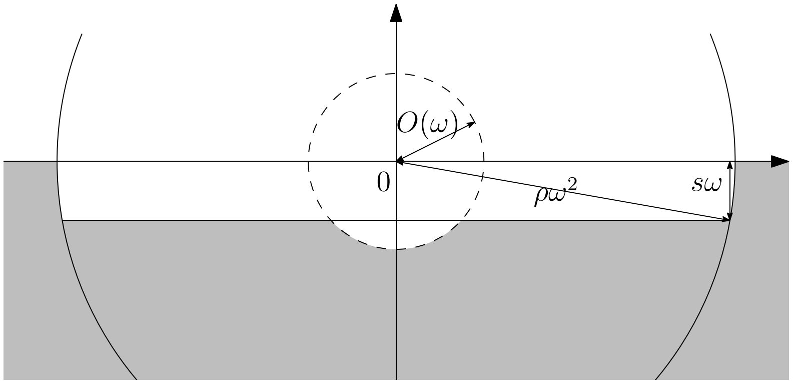


FIG. 5.1. *Distribution of eigenvalues in the complex plane for the Helmholtz equation with impedance boundary condition. All the eigenvalues are in the shaded area of the lower half plane. The dashed circle around the origin with radius $O(\omega)$ is the result of* [28, 11, 19]. *In the circle with radius $\rho\omega^2$, the minimum $s\omega$ distance from the real axis is the result of Theorem* 5.1.

Substituting the general solution into (5.3), we have

$$a_j^+(\omega+\xi_j) = -a_j^-(\omega-\xi_j), \tag{5.4}$$

$$a_j^+(\omega-\xi_j)e^{\mathrm{i}\xi_j} = -a_j^-(\omega+\xi_j)e^{-\mathrm{i}\xi_j}. \tag{5.5}$$

Eliminating $a_j^\pm$, we have that every ξ_j solves the same equation of z,

$$\frac{\omega-z}{\omega+z} = \pm e^{-\mathrm{i}z}. \tag{5.6}$$

There is no root in the first quadrant $\{z : \mathrm{Re}(z)\geq 0, \mathrm{Im}(z) > 0\}$ because

$$\frac{|\omega-z|}{|\omega+z|} \leq 1, \quad |e^{-\mathrm{i}z}| > 1.$$

So we are only interested in the fourth quadrant $\{z : \mathrm{Re}(z)\geq 0, \mathrm{Im}(z)\leq 0\}$. The following lemma will be used in the proof of Theorem 5.1.

Lemma 5.2. *Given a sequence of angular frequencies* $\{\omega_m\}$ *that goes to infinity, and for a sequence of complex numbers in the fourth quadrant* $\{z_m = a_m - \mathrm{i}b_m : a_m, b_m \geq 0\}$, *if there exists* $c_1, c_2 > 0$ *such that* $b_m \geq c_1\omega_m, b_m \geq c_2 a_m$, *then* z_m *does not solve* (5.6) *for sufficiently large* m.

Proof. The lemma can be easily proved by taking the absolute value on both sides of (5.6). For the left-hand side of (5.6), we have

$$\left|\frac{\omega_m - a_m + \mathrm{i}b_m}{\omega_m + a_m - \mathrm{i}b_m}\right| \geq \frac{b_m}{|\omega_m + a_m - \mathrm{i}b_m|} \geq \frac{1}{\sqrt{(1/c_1+1/c_2)^2+1}}.$$

The right-hand side of (5.6) satisfies

$$\lim_{m\to\infty} |\exp(-b_m - \mathrm{i}a_m)| = \lim_{m\to\infty}\exp(-b_m) = 0. \qquad \square$$

So the equality does not hold for sufficiently large m.

Proof of Theorem 5.1. If the statement is false, then there exists a $\rho > 0$, a sequence of angular frequencies $\{\omega_m\}$ that goes to infinity, and a sequence of complex phase vectors $\{\xi^{(m)} \in \mathbb{C}^d\}$ satisfying (5.2) and (5.3), but for $\lambda_m = \xi^{(m)}\cdot\xi^{(m)} - \omega_m^2$, we have that

$$|\lambda_m| \leq \rho\omega_m^2, \quad \lim_{m\to\infty}\frac{\mathrm{Im}(\lambda_m)}{\omega_m} = 0. \tag{5.7}$$

We can assume that there exists a sequence of indices $\{j_m \in \{1,2,\dots,d\}\}$ such that

$$|\xi_{j_m}^{(m)}| \geq \omega_m\sqrt{\frac{1-\rho}{d}}, \tag{5.8}$$

because otherwise $|\xi^{(m)}|^2 < \omega_m^2(1-\rho)$, and we have

$$|\lambda_m| \geq \omega_m^2 - |\xi^{(m)}|^2 > \rho\omega_m^2,$$

which contradicts with (5.7).

Let $\xi_{j_m}^{(m)} = a_m - \mathrm{i}b_m$. Because of (5.6)–(5.8), $a_m, b_m \geq 0$ satisfy

$$a_m^2 + b_m^2 \geq \frac{1-\rho}{d}\omega_m^2, \tag{5.9}$$

$$\lim_{m\to\infty} \frac{a_m b_m}{\omega_m} = 0, \tag{5.10}$$

$$\frac{\omega_m - a_m + \mathrm{i}b_m}{\omega_m + a_m - \mathrm{i}b_m} = \pm\exp(-b_m - \mathrm{i}a_m). \tag{5.11}$$

From (5.9) and (5.10), we get

$$\frac{\omega_m a_m b_m}{a_m^2 + b_m^2} = \frac{a_m b_m}{\omega_m}\frac{\omega_m^2}{a_m^2 + b_m^2} \to 0.$$

Then we can find subsequences (still denoted by $\{a_m\}, \{b_m\}$) such that either $\omega_m a_m/b_m$ or $\omega_m b_m/a_m$ goes to zero.

If $\omega_m a_m/b_m \to 0$, then $a_m/b_m \to 0$ for large ω_m. From (5.9), we also have

$$\frac{\omega_m}{b_m} \leq \sqrt{\frac{d}{1-\rho}\left(1 + \frac{a_m^2}{b_m^2}\right)}.$$

$\{a_m - \mathrm{i}b_m\}$ satisfies the assumptions in Lemma 5.2. Hence they are not roots of (5.6) for large m, a contradiction.

If $\omega_m b_m/a_m \to 0$, then from (5.10)

$$b_m^2 = \frac{\omega_m b_m}{a_m}\frac{a_m b_m}{\omega_m} \to 0.$$

From (5.9), we have

$$\frac{a_m^2}{\omega_m^2} \geq \frac{1-\rho}{d} - \frac{b_m^2}{\omega_m^2}.$$

So a_m/ω_m is bounded above zero. From (5.11),

$$1 = \exp\left(-\lim_{m\to\infty} b_m\right) = \lim_{m\to\infty}\frac{|1 - a_m/\omega_m|}{1 + a_m/\omega_m}.$$

So $a_m/\omega_m \to \infty$. For the complex phase vectors $\{\xi^{(m)}\}$,

$$\lambda_m = (a_m^2 - b_m^2 - \omega_m^2) - \mathrm{i}2a_m b_m + \sum_{l\neq j_m}\xi_l^{(m)}\xi_l^{(m)}.$$

For large m, we have $a_m^2 - b_m^2 - \omega_m^2 > 2\rho\omega_m^2$. We can find another sequence $\{\xi_{l_m}^{(m)} : l_m \neq j_m\}$ such that

$$\mathrm{Re}\left(\xi_{l_m}^{(m)}\xi_{l_m}^{(m)}\right) < -\frac{\rho}{d-1}\omega_m^2.$$

Because otherwise

$$\mathrm{Re}(\lambda_m) > 2\rho\omega_m^2 + \sum_{l\neq j_m}\mathrm{Re}\left(\xi_l^{(m)}\xi_l^{(m)}\right) \geq 2\rho\omega_m^2 - (d-1)\frac{\rho}{d-1}\omega_m^2 = \rho\omega_m^2,$$

which contradicts with (5.7). Let $\xi_{l_m}^{(m)} = \tilde{a}_m - \mathrm{i}\tilde{b}_m$ with nonnegative $\tilde{a}_m$ and $\tilde{b}_m$. We have

$$\tilde{a}_m^2 - \tilde{b}_m^2 < -\frac{\rho}{d-1}\omega_m^2.$$

So $\tilde{b}_m > \tilde{a}_m$ and $\tilde{b}_m > \sqrt{\rho/(d-1)}\omega_m$. Because of Lemma 5.2, this sequence does not solve (5.6) for large m, which is a contradiction. ∎

Theorem 5.1 depicts the fine structure of spectrum of discretized Helmholtz operators near the origin. Since the smallest eigenvalues of A are some distance away from the real axis, GMRES is expected to converge fast for the inner problem (2.5).

6. Techniques for improved performance.

In this section, we will discuss several special techniques to improve the speed and robustness of Algorithm 2.1.

6.1. Quadrature on an ellipse.

The contour γ is usually selected as a circle in existing methods. By shrinking, say, the real axis, the circle is transformed into an ellipse in the form of

$$\{tr\cos\theta + \mathrm{i}r\sin\theta :\ t, r > 0,\ \theta\in[0,2\pi]\}\,. \tag{6.1}$$

For fixed r, the advantage of using a small ratio t is that there are fewer eigenvalues enclosed by the contour. Hence the contour integration (2.3) is closer to the true inverse. The disadvantage is that the shape of γ becomes irregular, so the number of quadrature points may need to increase.

One way to derive quadrature rules on an ellipse is by mapping it to the unit circle parametrized by the angle θ. Note that

$$\frac{1}{2\pi\mathrm{i}}\mathrm{d}z = \frac{1}{2\pi\mathrm{i}}\mathrm{d}(tr\cos\theta + \mathrm{i}r\sin\theta) = (r\cos\theta + \mathrm{i}tr\sin\theta)\frac{\mathrm{d}\theta}{2\pi}.$$

For an equispaced set of angles $\{\theta_j\}$, the quadrature weights can be chosen as

$$\sigma_j = (r\cos\theta_j + \mathrm{i}tr\sin\theta_j)/J,$$

where J is the number of quadrature points. Choosing J as an even number gives symmetric quadrature points with respect to the major and minor axes.

6.2. Shifting the center.

In order to avoid an ill-conditioned shifted system, we shift the center of the ellipse according to the spectrum and choose a small number of points such as 6 so that the quadrature points are not close to any eigenvalue. This makes the contour integration method robust.

After determining t and an even J from the previous subsection, we simply choose the center of the contour as $-tr\cos\frac{\pi}{J} - \mathrm{i}\frac{\rho_2}{2}$. Recall from Lemma 3.1 that the imaginary part of each eigenvalue belongs to the interval $[-\rho_2, 0]$. Then, each quadrature point z_j satisfies

$$z_j = tr\left(\cos\frac{(2j-1)\pi}{J} - \cos\frac{\pi}{J}\right) + \mathrm{i}\left(r\sin\frac{(2j-1)\pi}{J} - \frac{\rho_2}{2}\right), \quad j\in[1,\ldots,J]. \tag{6.2}$$

$\mathrm{Re}(z_j) \le 0$ so that the shifted matrices cannot be more indefinite than the original one.

Let ϵ be a positive parameter. The following choice of r ensures that each z_j is at least ϵ-distance away from the spectrum:

$$r = \frac{\frac{\rho_2}{2} + \epsilon}{\sin\frac{\pi}{J}} \tag{6.3}$$

so that the imaginary part is

$$\mathrm{Im}(z_j) \notin (-\epsilon - \rho_2, \epsilon).$$

The new parameter ϵ further pushes the quadrature points away from the spectrum.

In summary, the quadrature points satisfy (6.2)–(6.3). The imaginary part of the spectrum uses ρ_2 from Lemma 3.1 explicitly. One can reduce t, ϵ and increase J for more accurate approximation of A^{-1}, and vice versa for faster approximation. The quadrature points are visualized in Figure 6.1.

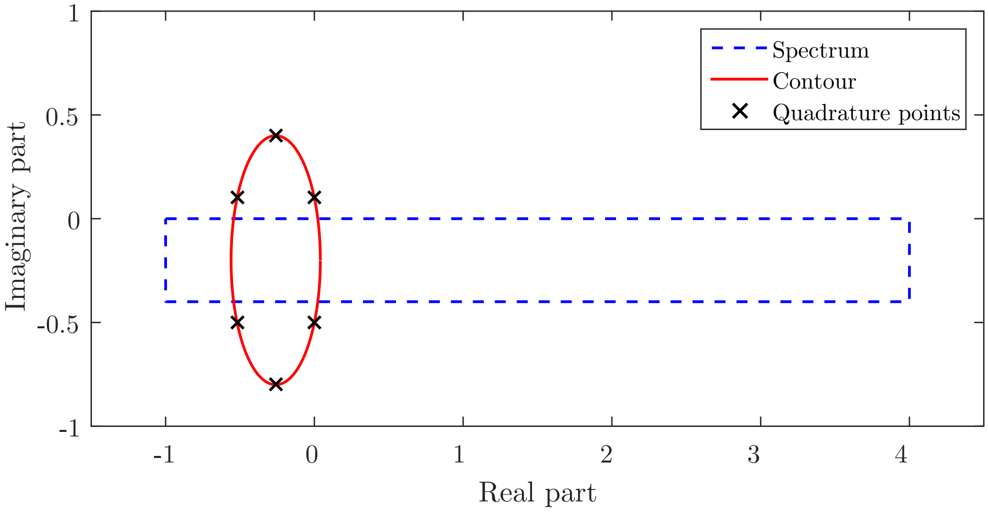


FIG. 6.1. *Illustration of the quadrature points* (6.2)*. Here, $t = 0.5$, $r = 0.6$, and $J = 6$.*

6.3. Preconditioning the inner problem. The inner problem (2.5) discussed in section 5 is an idealized case where the outer problem is exactly solved. Error in the solution and quadrature may affect the convergence of the inner problem. One can apply some standard preconditioners to improve the convergence. We find that an (approximate) discrete Laplacian usually gives a satisfactory performance by deflating the large eigenvalues of the Helmholtz problem. For the sparse case, this can be achieved by computing an ILU factorization without fill-in.

7. Numerical examples.

To illustrate the performance of the proposed method, we present the test results for solving a challenging 3D high-frequency Helmholtz equation. Since the method has little restrictions on the type of discretization, we attempt to solve both dense linear systems from a Fourier spectral method as well as sparse ones based on finite difference methods. The solution algorithms are implemented by MATLAB. The test machine is a Linux workstation having 3.5GHz CPU and 64GB RAM. In this section, we use the following notation:

- its: number of outer iterations;
- mvs: total number of matrix-vector products;
- i-t: iteration time in seconds.

7.1. Description of test problems. Every test matrix can be written formally as

$$A = S - M + \mathrm{i}D \tag{7.1}$$

for some Hermitian positive semidefinite matrices S, M, and D. S is the negative discrete Laplacian, M is the mass matrix that generates the indefiniteness of A, and D gives the non-Hermitian part. For solving free-space problems, D is used to reduce artificial reflections near the boundaries of the computational domain. One example is a diagonal matrix which has positive diagonal entries for points near the boundary and zero elsewhere; see, for example, [40].

If $S, M, D \succeq 0$, and $\rho(M) \leq 1$, then $S - M + I \succeq 0$ and A satisfies the assumptions of Lemma 3.1. It is helpful to know $\rho_1 = \rho(S - M + I)$ and $\rho_2 = \rho(D)$, because they characterize the spectrum of A. Since ρ_2 only depends on D, we study the combined effect of S and M on ρ_1 for our test problem.

Consider the 3D problem in a cuboid domain with an equispaced regular grid. The Fourier spectral method can be applied because the proposed method only requires matrix-vector products of the discrete Laplacian. S can be diagonalized by a 3D fast Fourier transform (FFT):

$$S = F^H \Lambda F,$$

where F is the transformation matrix of a forward FFT, and Λ is the diagonal matrix consisting of the eigenvalues of S. Each eigenvalue can be written as

$$\lambda_i = \left(\frac{l_{\min}}{N}\right)^2 \sum_{j=1}^{3} \min(i_j^2, (N - i_j)^2), \tag{7.2}$$

where i is a zero-based multi-index in an N^3 grid, and $l_{\min}$ is the minimum sampling rate (minimum number of points per wavelength). Clearly,

$$\rho(S) = \max_i |\lambda_i| \le \left(\frac{l_{\min}}{N}\right)^2 \sum_{d=1}^{3} \left(\frac{N}{2}\right)^2 = \frac{3}{4} l_{\min}^2.$$

The standard seven-point stencil can also be used to generate a sparse matrix $\tilde{S}$; then the eigenvalues are replaced by

$$\tilde{\lambda}_i = \left(\frac{l_{\min}}{2\pi}\right)^2 \sum_{j=1}^{3} 2\left(1 - \cos\frac{i_j}{N}\pi\right).$$

The spectral radius is instead $\rho(\tilde{S}) = \max_i |\tilde{\lambda}_i| \le \frac{3}{\pi^2} l_{\min}^2$. Low-order finite difference methods may need a large sampling rate $l_{\min}$, which increases the size and spectral radius of A.

The mass matrix M contains the variations of the wavespeed. For the simplest diagonal case, the ith nonzero diagonal entry is simply

$$M_{ii} = \frac{l_{\min}^2}{l_i^2},$$

where l_i is the local sampling rate on the ith grid point. For this case, ρ_1 defined in Lemma 3.1 satisfies

$$\rho_1 = \rho(S - M + I) \le \|S\|_2 + \|I - M\|_2 = O(l_{\min}^2) + \frac{l_{\max}^2 - l_{\min}^2}{l_{\max}^2}.$$

Regarding the spreading of the spectrum, the minimum sampling rate ($l_{\min}$) has a primary influence, and the variations of the wavespeed ($l_{\max}/l_{\min}$) play a secondary role.

Finally we discuss the generation of the diagonal matrix D. For a point that is j grid points away from the boundary, the corresponding diagonal value is $\rho_2(1 + \cos(\min(j, m)\pi/m))$. m is the thickness of the absorbing layers and is chosen as 10 here.

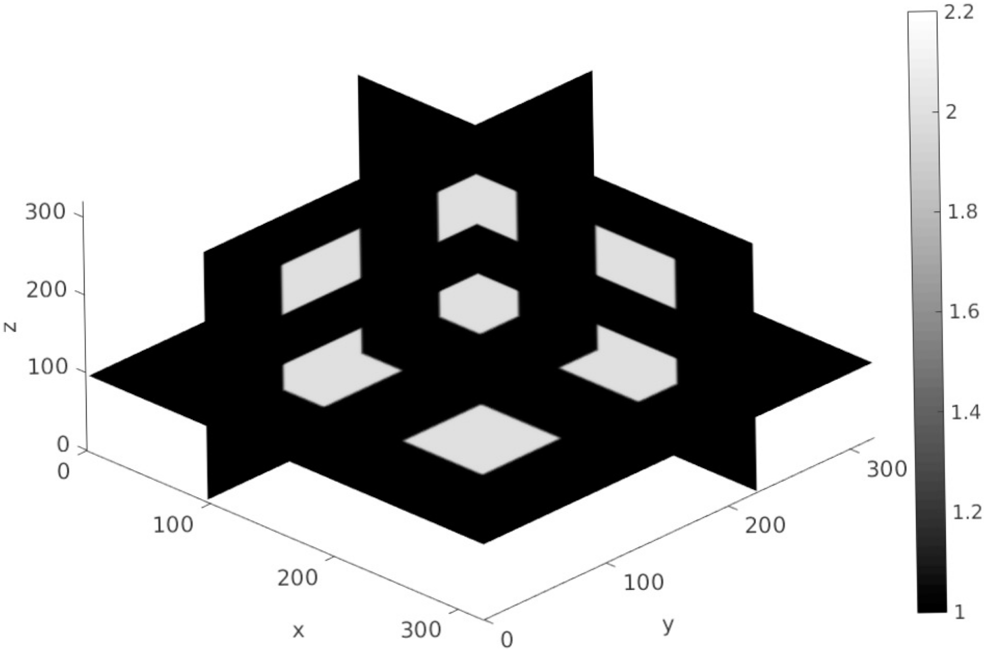


FIG. 7.1. *Wavespeed function $c(x)$ in* (1.1).

7.2. Scaling test. This scaling test checks the cost of solving (1.2) as the frequency and the problem size increase. We choose a wavespeed function that has eight high-wavespeed anomalies. Figure 7.1 visualizes the wavespeed function. The parameters needed by the matrix (7.1) are given as follows:

- For the spectral method, the sampling rate is $l_{\min} = 2.25$ in the background, and is $l_{\max} = 4.5$ inside the anomalies. For the constants in Lemma 3.1, $\rho_1 \approx 4.55, \rho_2 \approx 0.65$.
- For the finite difference approach, the frequency is reduced by four times to obtain a minimum sampling rate of $l_{\min} = 9$ and a maximum sampling rate of $l_{\max} = 18$. $\rho_1 \approx 25.37, \rho_2 \approx 0.65$.

The right-hand side has a single nonzero at the center of the grid. Figure 7.2 shows the solution of the largest problem size using spectral methods.

We set up the solver based on the techniques described in section 6. The solution is computed by the FCI preconditioned flexible GMRES. Six quadrature points are used in Algorithm 2.1, and their locations change with respect to the angular frequency ω. For example, for the four problems in Table 7.2(a), we fix $t = 0.1$ in (6.2) and select $\epsilon \propto 1/\omega$ in (6.3); then the sets of quadrature points are

$$\begin{pmatrix} 0.00 + 0.80\mathrm{i} \\ -0.22 + 2.05\mathrm{i} \\ -0.43 + 0.80\mathrm{i} \\ -0.43 - 1.70\mathrm{i} \\ -0.22 - 2.96\mathrm{i} \\ 0.00 - 1.70\mathrm{i} \end{pmatrix}, \quad \begin{pmatrix} 0.00 + 0.40\mathrm{i} \\ -0.15 + 1.25\mathrm{i} \\ -0.30 + 0.40\mathrm{i} \\ -0.30 - 1.30\mathrm{i} \\ -0.15 - 2.16\mathrm{i} \\ 0.00 - 1.30\mathrm{i} \end{pmatrix}, \quad \begin{pmatrix} 0.00 + 0.27\mathrm{i} \\ -0.12 + 0.99\mathrm{i} \\ -0.25 + 0.27\mathrm{i} \\ -0.25 - 1.17\mathrm{i} \\ -0.12 - 1.89\mathrm{i} \\ 0.00 - 1.17\mathrm{i} \end{pmatrix}, \quad \begin{pmatrix} 0.00 + 0.20\mathrm{i} \\ -0.11 + 0.85\mathrm{i} \\ -0.23 + 0.20\mathrm{i} \\ -0.23 - 1.10\mathrm{i} \\ -0.11 - 1.76\mathrm{i} \\ 0.00 - 1.10\mathrm{i} \end{pmatrix}.$$

The first quadrature point in each set is closest to the origin. The sets of quadrature weights are

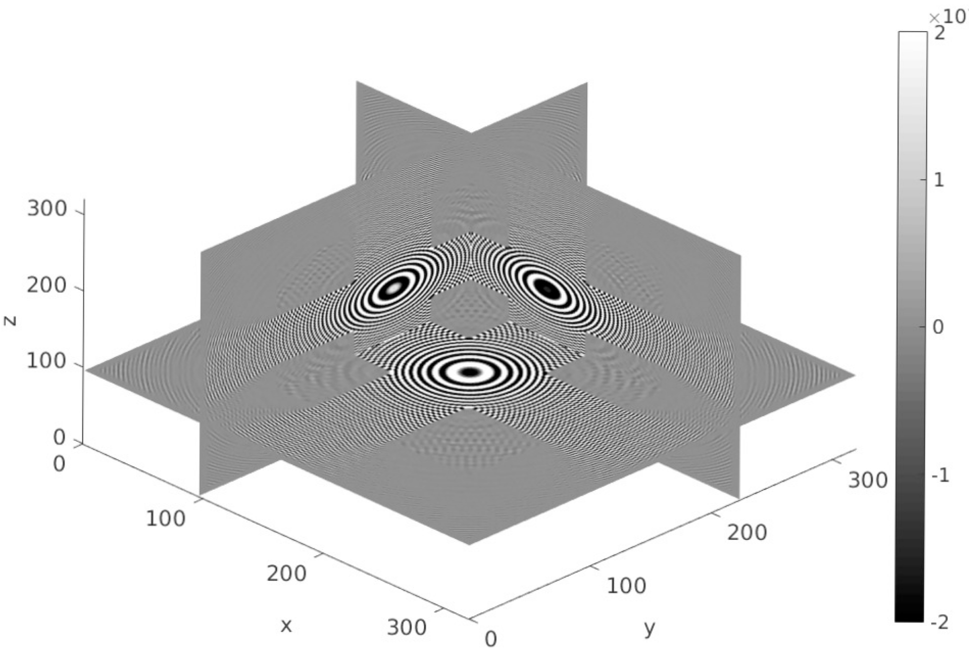


FIG. 7.2. *Solution of the problem in Figure* 7.1 *using the spectral method.*

$$\frac{1}{10}\begin{pmatrix}0.26-4.52\mathrm{i}\\0.20-0.02\mathrm{i}\\2.10+3.38\mathrm{i}\\0.62-1.96\mathrm{i}\\0.14+0.01\mathrm{i}\\0.12+2.12\mathrm{i}\end{pmatrix},\ \frac{1}{10}\begin{pmatrix}0.36-6.15\mathrm{i}\\0.22-0.03\mathrm{i}\\3.17+3.81\mathrm{i}\\0.51-1.77\mathrm{i}\\0.13+0.00\mathrm{i}\\0.11+1.89\mathrm{i}\end{pmatrix},\ \frac{1}{10}\begin{pmatrix}0.45-7.78\mathrm{i}\\0.24-0.03\mathrm{i}\\4.12+3.93\mathrm{i}\\0.46-1.67\mathrm{i}\\0.13+0.01\mathrm{i}\\0.10+1.77\mathrm{i}\end{pmatrix},\ \frac{1}{10}\begin{pmatrix}0.54-9.42\mathrm{i}\\0.25-0.03\mathrm{i}\\4.91+3.87\mathrm{i}\\0.43-1.62\mathrm{i}\\0.12+0.01\mathrm{i}\\0.10+1.71\mathrm{i}\end{pmatrix}.$$

Note that each contour is far from being circular and the number of points is rather small. The quadrature weights described in section 6 do not sum to one. We rely on the scalar multiplier d in line 4 of Algorithm 2.1 to compensate for this error. For the poles with smaller quadrature weights, the shifted problems can be solved to a lower accuracy. Our heuristic of choosing the relative tolerance τ_i at the ith pole is

$$\tau_i = \tau_1\sqrt{|\sigma_1|/|\sigma_i|},$$

where σ_i is the ith quadrature weight, and $\tau_1 = 0.20$ is the tolerance at the first pole on the positive imaginary axis. The method in section 4.2 can be used to determine the precise scheme for each shifted problem. Take the example of the first pole in each set; the parameters for applying (4.8) are listed in Table 7.1.

Regarding different cases in section 3, by estimating the spectrum we find that for the spectral method one can apply Algorithm 2.1 to the original matrix A because the spectrum is more compact, and the modified matrix i$C - I$ in (3.1) is suitable for the finite difference method. The spectral method case includes a regularized inverse Laplacian preconditioner and is diagonalized by FFT with eigenvalues $\{1/\max(\lambda_i, 1)\}$, where λ_i is defined in (7.2); the finite difference case includes an ILU(0) preconditioner based on the seven-point stencil discrete Laplacian. Since the cost of applying the Laplacian preconditioner is similar to one matrix-vector product, the counts are

Table 7.1

Parameters of the polynomial iteration (4.8) *for one pole. The parameters are computed by solving* (4.9)*. z is the complex shift. The method stops when the residual is reduced by five times.*

n	z	q	δ	mvs
80^3	0.8i	2	0.51	10
160^3	0.4i	3	0.91	18
240^3	0.27i	3	0.88	27
320^3	0.2i	3	0.85	39

Table 7.2

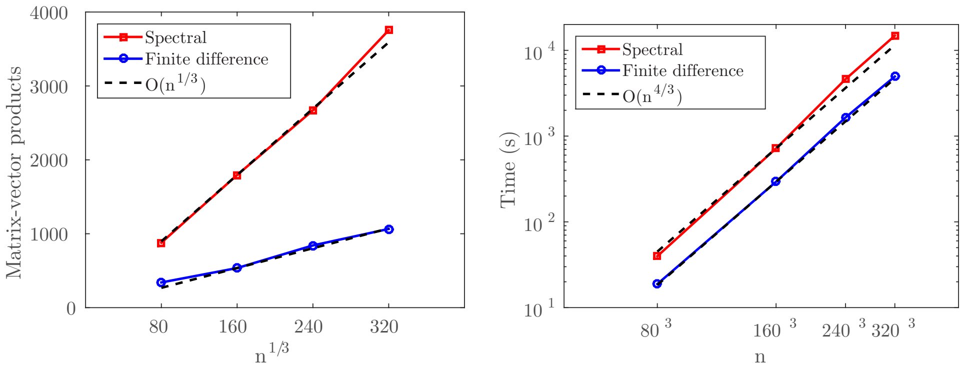
Scaling test for fixed sampling rate and increasing problem sizes.

(a) Fourier spectral method

n	$\omega/(2\pi)$	its	mvs	i-t
80^3	35.56	6	879	39.8
160^3	71.11	8	1795	719.6
240^3	106.67	9	2670	4610.2
320^3	142.22	11	3754	14841.6

(b) Finite difference method

n	$\omega/(2\pi)$	its	mvs	i-t
80^3	8.89	9	341	18.9
160^3	17.78	8	536	293.7
240^3	26.67	11	842	1649.2
320^3	35.56	10	1065	4954.5



(a) *Matrix-vector products* (b) *Iteration time*

Fig. 7.3. *3D scaling test.*

combined in mvs. The inner problem is more challenging, so we distribute 80% of the cost there. For each call of Algorithm 2.1, this is done by first counting the number of matrix-vector products for solving shifted problems, and then setting the number of Laplacian-preconditioned GMRES iterations to be double that number.

Table 7.2 and Figure 7.3 are the test results for reducing the residual by 10^3. For both cases, the number of matrix-vector products is proportional to the angular frequency $O(\omega)$. Because of the lack of sparsity, spectral methods are rarely considered for 3D Helmholtz problems. As can be seen from Table 7.2(a), the proposed solution method is suitable for solving this type of problems. Spherical patterns can be observed in Figure 7.2. This shows that the solution is qualitatively speaking meaningful. Note that we have chosen a straightforward Fourier spectral method here as the first attempt. In order to have accurate solution of variable-coefficient problems, the discretization method and the sampling rate may be improved in the future. In [15], a pseudospectral method is proposed for the variable-coefficient wave equation, and it is likely that similar techniques can be used for the Helmholtz equation.

7.3. SEG/EAGE salt-dome model. The SEG/EAGE salt-dome model [2] is a 3D wavespeed model commonly used in exploration geophysics. The physical size is 12km×12km×4.5km. The wavespeed ranges between 1500m/s and 4500m/s; see Figure 7.4 for sections of the wavespeed. At high frequency, 33.33Hz, we apply Fourier spectral discretization on a 201×676×676 grid. Figure 7.5 visualizes the

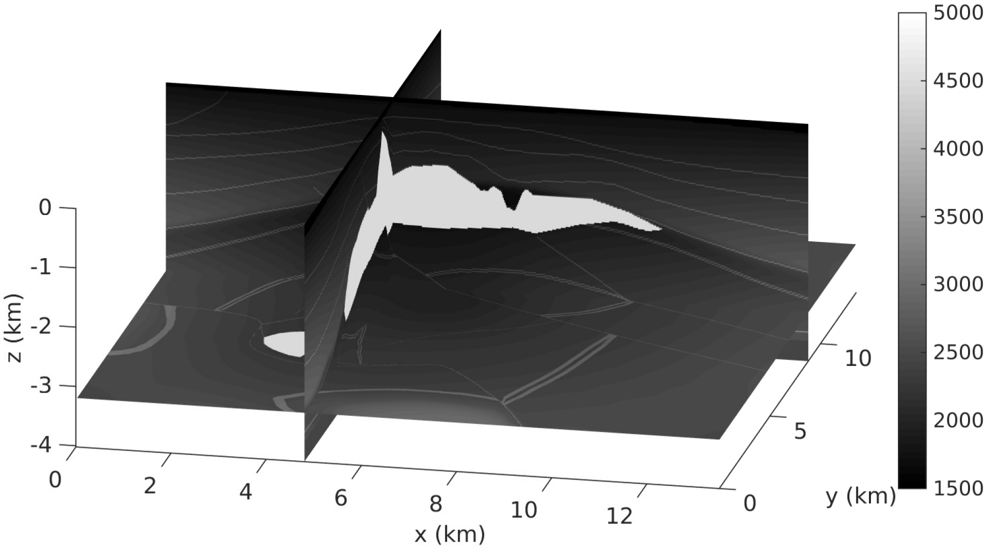


FIG. 7.4. *Wavespeed function of the SEG/EAGE salt-dome model.*

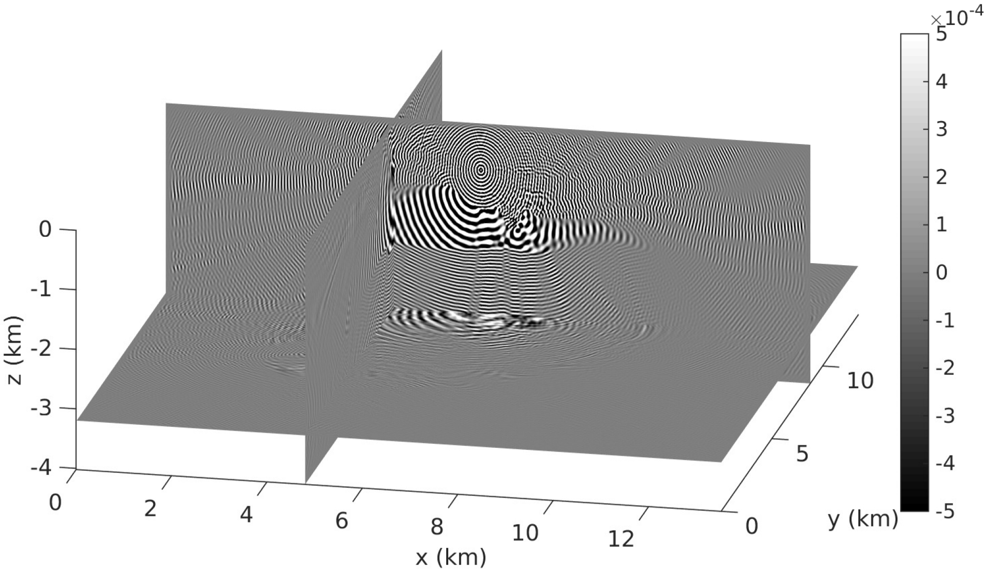


FIG. 7.5. 33.33*Hz solution wavefield corresponding to Figure* 7.4.

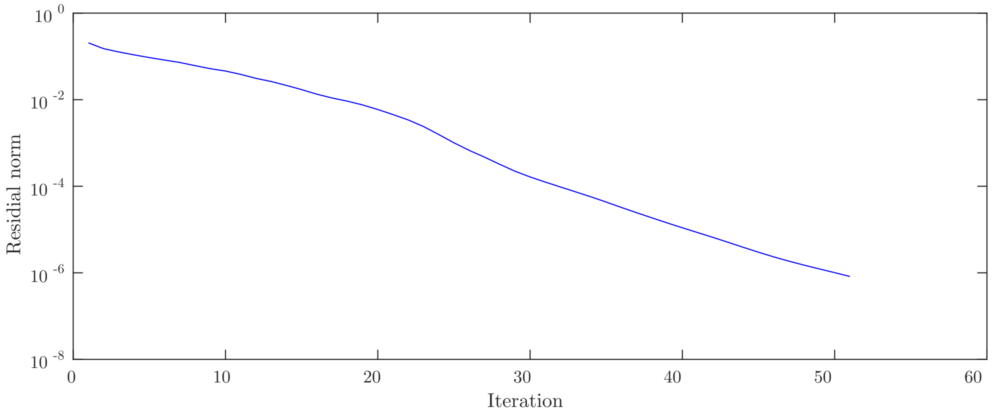


FIG. 7.6. *Residual history of FCI iterations of SEG/EAGE salt-dome model.*

solution wavefield. For 10^{-2}, 10^{-4}, and 10^{-6} relative residual, the proposed method takes 17 iterations (984 min), 32 iterations (1850 min), and 51 iterations (2943 min), respectively. Figure 7.6 shows that linear convergence of the residual still holds even when the wavespeed is rather complex. This test shows the capability of solving a realistic 3D high-frequency problem with limited memory consumption. This is possible because spectral methods can reduce the matrix size, and the solution method here does not generate or factorize any block of the matrix directly.

8. Conclusions.

An iterative method was proposed to solve the discretized 3D high-frequency Helmholtz equation. In the framework of the contour integration method which implicitly decomposes the original problem into an inner and an outer problem, a fixed-point iteration was introduced to solve the outer problem. GMRES is suitable for solving the inner problem because of our theoretical estimates on the distribution of eigenvalues. 3D numerical examples show that the computational cost of this method scales as $O(\omega^4)$ or $O(n^{4/3})$. The method is especially suitable for solving high-frequency problems when combined with spectral methods.

Acknowledgment. We would like to thank both referees for their suggestions.

REFERENCES

[1] M. AHMAD, D. SZYLD, AND M. VAN GIJZEN, *Preconditioned multishift BiCG for $\mathcal{H}$_2-optimal model reduction*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 401–424.

[2] F. AMINZADEH, J. BRAC, AND T. KUNZ, *3-D Salt and Overthrust Models*, SEG/EAGE 3-D Modeling Series 1, Society of Exploration Geophysicists, 1997.

[3] Z. Z. BAI, M. BENZI, AND F. CHEN, *Modified HSS iteration methods for a class of complex symmetric linear systems*, Computing, 87 (2010), pp. 93–111.

[4] Z. Z. BAI, G. H. GOLUB, AND M. K. NG, *Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 603–626.

[5] A. BAYLISS, C. I. GOLDSTEIN, AND E. TURKEL, *An iterative method for the Helmholtz equation*, J. Comput. Phys., 49 (1983), pp. 443–457.

[6] B. BECKERMANN, S. A. GOREINOV, AND E. E. TYRTYSHNIKOV, *Some remarks on the Elman estimates for GMRES*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 772–778.

[7] J. BENAMOU AND B. DESPRÈS, *A domain decomposition method for the Helmholtz equation and related optimal control problems*, J. Comput. Phys., 136 (1997), pp. 68–82.

[8] I. BENDIXSON, *Sur les racines d'une équation fondamentale*, Acta Math., 25 (1902), pp. 359–365.

[9] J. P. Berenger, *A perfectly matched layer for the absorption of electromagnetic waves*, J. Comput. Phys., 114 (1994), pp. 185–200.

[10] P. H. Cocquet and M. J. Gander, *How large a shift is needed in the shifted Helmholtz preconditioner for its effective inversion by multigrad?*, SIAM J. Sci. Comput., 39 (2017), pp. A438–A478.

[11] P. Cummings and X. Feng, *Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equation*, Math. Models Methods Appl. Sci., 16 (2006), pp. 139–160.

[12] H. C. Elman, *Iterative Methods for Sparse Nonsymmetric Systems of Linear Equations*, Ph.D. thesis, Yale University, 1982.

[13] B. Engquist and L. Ying, *Sweeping preconditioner for the Helmholtz equation: Moving perfectly matched layers*, Multiscale Model. Simul., 9 (2011), pp. 686–710.

[14] Y. A. Erlangga, C. Vuik, and C. W. Oosterlee, *On a class of preconditioners for the Helmholtz equation*, Appl. Numer. Math., 50 (2005), pp. 409–425.

[15] J. T. Etgen and S. Brandsberg-Dahl, *The pseudo-analytical method: Application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation*, in SEG Technical Program Expanded Abstracts, Society of Exploration Geophysicists, 2009, pp. 2552–2556.

[16] R. Freund, *On conjugate gradient type methods and polynomial preconditioners for a class of complex non-hermitian matrices*, Numer. Math., 57 (1990), pp. 285–312.

[17] M. J. Gander, *Optimized Schwarz methods*, SIAM J. Numer. Anal., 44 (2006), pp. 699–731.

[18] M. J. Gander, F. Magoules, and F. Nataf, *Optimized Schwarz methods without overlap for the Helmholtz equation*, SIAM J. Sci. Comput., 24 (2002), pp. 38–60.

[19] M. J. Gander, I. G. Graham, and E. A. Spence, *Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: What is the largest shift for which wavenumber-independent convergence is guaranteed?*, Numer. Math., 131 (2015), pp. 567–614.

[20] A. George, *Nested dissection of a regular finite element mesh*, SIAM J. Numer. Anal., 10 (1973), pp. 345–363.

[21] M. H. Gutknecht and S. Röllin, *The Chebyshev iteration revisited*, Parallel Comput., 28 (2002), pp. 263–283.

[22] M. B. Van Gijzen, Y. A. Erlangga, and C. Vuik, *Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted Laplacian*, SIAM J. Sci. Comput., 29 (2007), pp. 1942–1958.

[23] C. T. Kelley and D. E. Keyes, *Convergence analysis of pseudo-transient continuation*, SIAM J. Numer. Anal., 35 (1998), pp. 508–523.

[24] X. Liu, J. Xia, and M. V. de Hoop, *Parallel randomized and matrix-free direct solvers for large structured dense linear systems*, SIAM J. Sci. Comput., 38 (2016), pp. S508–S538.

[25] T. A. Manteuffel, *The Tchebyshev iteration for nonsymmetric linear systems*, Numer. Math., 28 (1977), pp. 307–327.

[26] T. A. Manteuffel, *Adaptive procedure for estimation of parameter for the nonsymmetric Tchebyshev iteration*, Numer. Math., 28 (1978), pp. 187–208.

[27] S. P. MacLachlan and C. W. Oosterlee, *Algebraic multigrid solvers for complex-valued matrices*, SIAM J. Sci. Comput., 30 (2008), pp. 1548–1571.

[28] J. M. Melenk, *On Generalized Finite Element Methods*, Ph.D. thesis, University of Maryland, 1995.

[29] G. Opfer and G. Schober, *Richardson's iteration for nonsymmetric matrices*, Linear Algebra Appl., 58 (1984), pp. 343–361.

[30] D. Osei-Kuffuor and Y. Saad, *Preconditioning Helmholtz linear systems*, Appl. Numer. Math., 60 (2010), pp. 420–431.

[31] L. Reichel, *The application of Leja points to Richardson iteration and polynomial preconditioning*, Linear Algebra Appl., 154 (1991), pp. 389–414.

[32] Y. Saad, *ILUT: A dual threshold incomplete LU factorization*, Numer. Linear Algebra Appl., 1 (1994), pp. 387–402.

[33] Y. Saad, *Least squares polynomials in the complex plane and their use for solving nonsymmetric linear systems*, SIAM J. Numer. Anal., 24 (1987), pp. 155–169.

[34] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003.

[35] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Classics in Appl. Math. 66, SIAM, Philadelphia, 2011.

[36] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.

[37] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.

[38] E. B. Saff and R. S. Varga, *Zero-free parabolic regions for sequences of polynomials*, SIAM J. Math. Anal., 7 (1976), pp. 344–357.

[39] J. Shen and L. Wang, *Spectral approximation of the Helmholtz equation with high wave numbers*, SIAM J. Numer. Anal., 43 (2005), pp. 623–644.

[40] C. Shin, *Sponge boundary condition for frequency-domain modeling*, Geophysics, 60 (1995), pp. 1870–1874.

[41] C. C. Stolk, *A rapidly converging domain decomposition method for the Helmholtz equation*, J. Comput. Phys., 241 (2013), pp. 240–252.

[42] G. Szegö, *Über eine eigenschaft der exponentialreihe*, Berlin Math. Ges. Sitzungsber., 23 (1924), pp. 50–64.

[43] P. Tang and E. Polizzi, *FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 354–390.

[44] H. E. Wrigley, *Accelerating the Jacobi method for solving simultaneous equations by Chebyshev extrapolation when the eigenvalues of the iteration matrix are complex*, Comput. J., 6 (1963), pp. 169–176.

[45] Y. Xi and Y. Saad, *Computing partial spectra with least-squares rational filters*, SIAM J. Sci. Comput., 38 (2016), pp. A3020–A3045.

[46] Y. Xi and Y. Saad, *A rational function preconditioner for indefinite sparse linear systems*, SIAM J. Sci. Comput., 39 (2017), pp. A1145–A1167.

[47] L. Zepeda-Núñez and L. Demanet, *The method of polarized traces for the 2D Helmholtz equation*, J. Comput. Phys., 308 (2016), pp. 347–388.