Superlinear convergence using block preconditioners for the real system formulation of complex Helmholtz equations

Owe Axelsson, János Karátson, Frédéric Magoulès

Superlinear convergence using block preconditioners for the real system formulation of complex Helmholtz equations

Owe Axelsson* János Karátson† Frédéric Magoulès‡

Abstract

Complex-valued Helmholtz equations arise in various applications, and a lot of research has been devoted to finding efficient preconditioners for the iterative solution of their discretizations. In this paper we consider the Helmholtz equation rewritten in real-valued block form, and use a preconditioner in a special two-by-two block form. We show that the corresponding preconditoned Krylov iteration converges at a mesh-independent superlinear rate.

Keywords: Helmholtz equations; Iterative solution; Preconditioning; Mesh-independent superlinear convergence

1 Introduction

Complex-valued Helmholtz equations arise in the modelling of various applied problems, for instance, when air is periodically compressed into some closed compartment, e.g., in a car. For the iterative solution of their discretization, standard preconditioning methods such as incomplete factorization or (algebraic) multigrid methods are not efficient, mainly due to the effect of high indefiniteness and large wave-numbers [8, 14], therefore more efficient iterative solvers are still of great interest. A lot of recent research has been devoted to preconditioners arising as the discretization of the so-called "complex shifted Laplace" problems, see, e.g., [7, 8, 9, 10, 13]. These preconditioners require, however, use of complex arithmetics and solution of a still somewhat involved preconditioner. In this note we modify the preconditoner so that it can be solved directly in real arithmetics and still preserve the favourable properties of convergence. Each application of the action of the preconditioner involves essentially the solution of only two standard elliptic problems.

We consider the Helmholtz equation in a bounded domain Ω, with impedance boundary conditions, i.e.,

$$\begin{cases} -\Delta \underline{u} - \kappa^2 \underline{u} &= \underline{f} \quad \text{in } \Omega \\ \dfrac{\partial \underline{u}}{\partial n} - i\kappa \underline{u} &= 0 \quad \text{on } \partial\Omega. \end{cases} \tag{1}$$

For simplicity, we assume that Ω is a polygonal domain. Here $\underline{u} = u + iv$, where u, v are real valued, the wave number $\kappa > 0$ and $\underline{f} = f + ig$, i being the imaginary unit. Due to the imaginary coefficient in the boundary condition, the positive real number κ^2 cannot attain an eigenvalue, i.e. the homogeneous problem with $g \equiv 0$ has only the trivial solution $u \equiv 0$, see, e.g., [12]. The Fredholm alternative then ensures that problem (1) has a unique weak

*Institute of Geonics AS CR, Ostrava, Czech Republic.

†Department of Applied Analysis & MTA-ELTE Numerical Analysis and Large Networks Research Group, ELTE University; Department of Applied Analysis, Technical University; Budapest, Hungary.

‡CentraleSupélec, Université Paris-Saclay, France (correspondence, frederic.magoules@hotmail.com).

solution in $H^1(\Omega)$. Moreover, Fredholm's well-posedness result involves the invertibility of the corresponding operator on the left-hand side.

As proposed in the previously cited papers, one can form a preconditioner by use of a complex shifted Laplace operator with perturbation terms of lower order,

$$\begin{cases} -\Delta \underline{u} - (\kappa^2 + i\varepsilon)\underline{u} &= \underline{f} \quad \text{in } \Omega \\ \dfrac{\partial \underline{u}}{\partial n} - i\mu\, \underline{u} &= 0 \quad \text{on } \partial\Omega, \end{cases} \tag{2}$$

where $\varepsilon > 0$, $\mu > 0$ are perturbation coefficients. This corresponds to a compact perturbation of the given Helmholtz operator equation. In this situation the corresponding preconditoned Krylov iteration method typically converges at a superlinear rate, which is a main interest in this paper, see also [3, 4, 17, 18, 19] on superlinear results for coercive problems and [5] in the case of the complex formulation (2). However, discretizing (2) with standard finite element methods and solving systems with this preconditioner entails complex valued systems and complex valued arithmetics and, even though they have a more favourable distribution of eigenvalues, the systems are still somewhat complicated to solve. Hence often the real-valued block form of the Helmholtz equation is preferred, see, e.g., [11].

In this paper we rewrite the Helmholtz equation in real-valued block form and use a somewhat modified version of the shifted Laplacian. We thus get a preconditioner in a special two-by-two block form, the solution of which only involves real arithmetics in the form of some vector operators and two solution steps for real-valued elliptic problems. For the latter, standard solution methods can be used. It will be shown with a proper analysis of the method that the corresponding preconditoned Krylov iteration still converges at a superlinear rate.

2 Superlinear convergence of Krylov type methods

Here we very briefly summarize the required basic facts on the solution of linear systems

$$Au = b \tag{3}$$

with a given nonsingular matrix $A \in \mathbf{R}^{n\times n}$, with focus on superlinear convergence rates for the iterative solution of (3).

When A is an s.p.d. (i.e. symmetric positive definite) matrix, then the widespread way of iterative solution is the standard CG method, for which a well-known superlinear convergence estimate is expressed in terms of the decomposition $A = I + E$, where I is the identity matrix [1, 18]. For non-s.p.d. matrices A, several Krylov algorithms exist (see e.g., [1, 16]), in particular, GMRES and its variants are widely used. There exist similar superlinear convergence estimates for the GMRES as for the standard CG, using singular values and the residual error vectors $r_k := Au_k - b$. In fact, the sharpest one, proved in [15] on the Hilbert space level for an invertible operator $A \in B(H)$, uses the product of singular values, and one can readily derive a simplified form as follows [5]:

$$\left(\frac{\|r_k\|}{\|r_0\|}\right)^{1/k} \leq \frac{\|A^{-1}\|}{k} \sum_{j=1}^{k} s_j(E) \qquad (k = 1, 2, ...). \tag{4}$$

Here the right-hand side is decreasing (and on the operator level it tends to zero as k tends to infinity), which means that the convergence is superlinear.

3 The block preconditioner

3.1 Construction of the preconditioner

The method to be used is based on first rewriting the given Helmholtz equation in real-valued system form,

$$\begin{cases} -\Delta u - \kappa^2 u = f \text{ in } \Omega, & \dfrac{\partial u}{\partial n} + \kappa v = 0 \text{ on } \partial\Omega, \\ -\Delta v - \kappa^2 v = g \text{ in } \Omega, & \dfrac{\partial v}{\partial n} - \kappa u = 0 \text{ on } \partial\Omega. \end{cases} \tag{5}$$

For the weak form we involve the real Hilbert space $H^1(\Omega)^2 := H^1(\Omega) \times H^1(\Omega)$, endowed with the inner product

$$\left\langle \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} z \\ w \end{pmatrix} \right\rangle_{(H^1)^2} := \int_\Omega (\nabla u \cdot \nabla z + uz) + \int_\Omega (\nabla v \cdot \nabla w + vw). \tag{6}$$

Then the weak formulation of (5) reads as follows: find $(u, v) \in H^1(\Omega)^2$ such that

$$\int_\Omega (\nabla u \cdot \nabla z - \kappa^2 uz) + \kappa \int_{\partial\Omega} vz - \kappa \int_{\partial\Omega} uw + \int_\Omega (\nabla v \cdot \nabla w - \kappa^2 vw) \;=\; \int_\Omega (fz + gw)$$

for all $(z, w) \in H^1(\Omega)^2$. As mentioned in the introduction, we have a well-posedness result via Fredholm theory.

The construction of the preconditioner can be indicated on the operator level. Namely, the above equations can be perturbed by the addition of zero-th order modified shifted terms to form the PDE system

$$\begin{cases} -\Delta u - \kappa^2 u + \varepsilon v = f \text{ in } \Omega, & \dfrac{\partial u}{\partial n} + \mu v = 0 \text{ on } \partial\Omega, \\ -\Delta v - \kappa^2 v + 2\varepsilon v - \varepsilon u = g \text{ in } \Omega, & \dfrac{\partial v}{\partial n} + 2\mu v - \mu u = 0 \text{ on } \partial\Omega, \end{cases} \tag{7}$$

where $\varepsilon > 0$, $\mu > 0$ are perturbation parameters. The weak formulation of the perturbed problem reads as follows: find $(u, v) \in H^1(\Omega)^2$ such that

$$\int_\Omega (\nabla u \cdot \nabla z - \kappa^2 uz) + \varepsilon \int_\Omega vz + \mu \int_{\partial\Omega} vz$$

$$-\varepsilon \int_\Omega uw - \mu \int_{\partial\Omega} uw + \int_\Omega (\nabla v \cdot \nabla w - \kappa^2 vw) + 2\varepsilon \int_\Omega vw + 2\mu \int_{\partial\Omega} vw \;=\; \int_\Omega (fz + gw)$$

for all $(z, w) \in H^1(\Omega)^2$. The reasons for this choice of the preconditioning operator will be discussed in the next subsection.

Now let us consider the discretization of the above problems using the finite element method (FEM). Let $V_h \subset H^1(\Omega)$ be a finite element function space corresponding to a finite element partitioning of Ω in a triangular/polygonal mesh, to be used for both solution components u and v. Let $\{\varphi_i\}_1^n$ be the set of basis functions in V_h. Let A_h correspond to the finite element discretization of the operator $-\Delta - \kappa^2 I$, that is, $A_h = [a_{ij}]$ where

$$a_{ij} = a(\varphi_j, \varphi_i)$$

with the bilinear form

$$a(u, z) = \int_\Omega (\nabla u \cdot \nabla z - \kappa^2 uz) \quad \forall z \in V_h.$$

Further, let M_h and B_h be the domain mass matrix and boundary mass matrix, respectively, that is,

$$M_h = [m_{ij}], \quad m_{ij} = p(\varphi_j, \varphi_i), \quad p(u,z) = \int_\Omega uz \quad \forall z \in V_h.$$

$$B_h = [b_{ij}], \quad b_{ij} = q(\varphi_j, \varphi_i), \quad q(u,z) = \int_{\partial\Omega} uz \quad \forall z \in V_h.$$

Then the finite element matrix, corresponding to the real Helmholtz system (5), takes the saddle-point form

$$\mathcal{A}_h = \begin{bmatrix} A_h & \kappa B_h \\ -\kappa B_h & A_h \end{bmatrix}.$$

We note that the finite element mesh should be sufficiently fine to enable capturing of the waves, so some six node points within each wave is a proper choice.

The discretization of the preconditioning operator in system (7) leads to the matrix

$$\tilde{\mathcal{A}}_h = \begin{bmatrix} A_h & \varepsilon M_h + \mu B_h \\ -(\varepsilon M_h + \mu B_h) & A + 2(\varepsilon M_h + \mu B_h) \end{bmatrix} \equiv \begin{bmatrix} A_h & C_h \\ -C_h & A_h + 2C_h \end{bmatrix}, \tag{8}$$

where

$$C_h := \varepsilon M_h + \mu B_h.$$

The constants ε, μ shall be chosen such that $A_h + C_h$ becomes regular. Then the matrix $\tilde{\mathcal{A}}_h$ is also regular, as seen in the next subsection.

3.2 Advantages of the preconditioner

A major advantage of the proposed preconditioner $\tilde{\mathcal{A}}_h$ is due to its very special structure. Namely, it admits a factorization that leads to efficient solution of the arising systems in the preconditioning steps, since these are reduced to two standard symmetric positive definite subproblems, as will be described below. Such a factorization property could not be achieved by using either simple block diagonal preconditioning or a preconditioner based on the real block form of (2). We note that even if an iterative solution method has a superlinear rate of convergence, in practical applications this superlinear rate may only be seen after many initial iterations unless an efficient preconditioner, leading to a small condition number, is used. It has been shown in [2, 6] that preconditioners with such factorization lead to a small condition number, further, they have been efficiently tested in detail therein. The present paper is not aimed at numerical testing, instead, robust superlinear estimates will be derived, showing that it holds independently of the mesh size.

The systems corresponding to (8) can be reduced to systems with matrix $A_h + C_h$. Namely, as observed, e.g., in [2], an elementary computation shows that such an $\tilde{\mathcal{A}}_h$ can be directly factorized as

$$\tilde{\mathcal{A}}_h = \begin{bmatrix} I_h & I_h \\ 0 & I_h \end{bmatrix} \begin{bmatrix} A_h + C_h & 0 \\ -C_h & A_h + C_h \end{bmatrix} \begin{bmatrix} I_h & -I_h \\ 0 & I_h \end{bmatrix}.$$

Hence, the solution of a system $\tilde{\mathcal{A}}_h \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$ leads to the solution of

$$\begin{bmatrix} A_h + C_h & 0 \\ -C_h & A_h + C_h \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} I_h & -I_h \\ 0 & I_h \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} h \\ g \end{bmatrix},$$

where $z = x - y$ and $h = f - g$. Therefore, it involves the following computations:

(i) $h = f - g$

(ii) solve $(A_h + C_h)z = h$

(iii) compute $k = g + C_h z$

(iv) solve $(A_h + C_h)y = k$

(v) $x = z + y$.

Hence, besides three vector additions and a matrix-vector product, it involves only the solution of two systems with the real-valued matrix $A_h + C_h$. Here C_h involves the domain and boundary mass matrices, and thus the matrix $A_h + C_h$ corresponds to the operator

$$-\Delta u + (\varepsilon - \kappa^2) \quad \text{in } \Omega$$

with boundary condition

$$\frac{\partial u}{\partial n} + \mu u = 0 \quad \text{on } \partial\Omega.$$

Choosing $\varepsilon \geq \kappa^2$ and $\mu > 0$ implies that the operator is strictly positive and hence one-to-one. Consequently, with such choices the symmetric matrix $A_h + C_h$ will be positive definite (and hence regular as required), and the solution of the subproblems can be achieved with standard methods, such as (algebraic) multilevel inner preconditioners.

Remark 3.1. The preconditioning matrix is also related to some favourable approximation properties. Firstly, if we define the intermediate matrix

$$\overline{\mathcal{A}}_h = \begin{bmatrix} A_h & C_h \\ -C_h & A_h \end{bmatrix}, \tag{9}$$

then $\overline{\mathcal{A}}_h$ differs from $\mathcal{A}_h$ only in a term arising from a compact perturbation of (5), hence $\overline{\mathcal{A}}_h^{-1}\mathcal{A}_h$ has a condition number bounded uniformly in h and expected to be small. In turn, $\tilde{\mathcal{A}}_h$ differs from $\overline{\mathcal{A}}_h$ in a further compact perturbation term for which it has been shown, e.g., in [2, 6] that the condition number of such a preconditioned matrix $\tilde{\mathcal{A}}_h^{-1}\mathcal{A}_h$ is bounded by two. The above also shows that the overall preconditioner also arises from a compact perturbation of (5). Our goal is to prove that this leads to a mesh independent superlinear rate of convergence.

Altogether, we wish to solve the preconditioned system

$$\tilde{\mathcal{A}}_h^{-1}\mathcal{A}_h c_h = \tilde{\mathcal{A}}_h^{-1} b_h \tag{10}$$

using a GMRES iteration.

4 Mesh independent superlinear convergence for the preconditioned system

4.1 Decomposition of the matrices

Let us introduce the bounded linear operators $L_S : H^1(\Omega)^2 \to H^1(\Omega)^2$ and $N_S : H^1(\Omega)^2 \to H^1(\Omega)^2$, defined via the following equalities:

$$\left\langle L_S \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} z \\ w \end{pmatrix} \right\rangle_{(H^1)^2} := \int_\Omega (\nabla u \cdot \nabla z - \kappa^2 uz) + \kappa \int_{\partial\Omega} vz - \kappa \int_{\partial\Omega} uw + \int_\Omega (\nabla v \cdot \nabla w - \kappa^2 vw) \tag{11}$$

$$\begin{aligned} \left\langle N_S \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} z \\ w \end{pmatrix} \right\rangle_{(H^1)^2} &:= \int_\Omega (\nabla u \cdot \nabla z - \kappa^2 uz) + \varepsilon \int_\Omega vz + \mu \int_{\partial\Omega} vz \\ -\varepsilon \int_\Omega uw - \mu \int_{\partial\Omega} uw + \int_\Omega (\nabla v \cdot \nabla w - \kappa^2 vw) + 2\varepsilon \int_\Omega vw + 2\mu \int_{\partial\Omega} vw \end{aligned} \tag{12}$$

The operators L_S and N_S correspond to the left-hand sides of the weak forms of problems (2) and (3), respectively.

Remark 4.1. The existence of L_S and N_S is ensured by the Riesz theorem. The subscript S refers to the following notation: using the setting of [4], the space $H^1(\Omega)^2$ is in fact the energy space of the operator

$$S\begin{pmatrix} u \\ v \end{pmatrix} := \begin{pmatrix} -\Delta u + u \\ -\Delta v + v \end{pmatrix}$$

defined for $\frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = 0$, further, the Helmholtz and preconditioning operators are S-bounded, giving rise to their weak forms L_S and N_S in $H^1(\Omega)^2$. This setting is used in order to keep the operators within the space $H^1(\Omega)^2$ instead of mapping into $H^{-1}(\Omega)^2$.

The operator S induces the weight stiffness matrix

$$\mathcal{S}_h = \begin{bmatrix} S_h & 0 \\ 0 & S_h \end{bmatrix}$$

where, corresponding to (6),

$$\big(S_h\big)_{i,j} := \langle \varphi_i, \varphi_j \rangle_{H^1} = \int_\Omega \Big(\nabla\varphi_i \cdot \nabla\overline{\varphi}_j + \beta\varphi_i\overline{\varphi}_j \Big) \qquad (i,j = 1,...,n). \tag{13}$$

The operators L_S and N_S satisfy

$$L_S = N_S + Q_S, \tag{14}$$

where

$$\left\langle Q_S \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} z \\ w \end{pmatrix} \right\rangle_{(H^1)^2} :=$$

$$-\varepsilon \int_\Omega vz + (k-\mu) \int_{\partial\Omega} vz + \varepsilon \int_\Omega uw - (k-\mu) \int_{\partial\Omega} uw - 2\varepsilon \int_\Omega vw - 2\mu \int_{\partial\Omega} vw\,. \tag{15}$$

For finite element matrices, the decomposition analogous to (14) is

$$\mathcal{A}_h = \tilde{\mathcal{A}}_h + \mathcal{Q}_h$$

where

$$\mathcal{Q}_h = \begin{bmatrix} 0 & -\varepsilon M_h + (\kappa-\mu) B_h \\ \varepsilon M_h - (\kappa-\mu) B_h & -2(\varepsilon M_h + \mu B_h) \end{bmatrix} \tag{16}$$

or, using the previously used notation $C_h := \varepsilon M_h + \mu B_h$,

$$\mathcal{Q}_h = \begin{bmatrix} 0 & \kappa B_h - C_h \\ -\kappa B_h + C_h & -2C_h \end{bmatrix}.$$

In preconditioned form we have

$$\tilde{\mathcal{A}}_h^{-1}\mathcal{A}_h = \mathcal{I}_h + \tilde{\mathcal{A}}_h^{-1}\mathcal{Q}_h\,.$$

We apply the GMRES algorithm for the matrix $\tilde{\mathcal{A}}_h^{-1}\mathcal{A}_h$. Since the desired mesh independence property relies on the underlying operator level in $H^1(\Omega)^2$, we use the discrete Sobolev inner product induced by the weight matrix $\mathcal{S}_h$. In particular, the adjoint of a matrix M w.r.t. this inner product (the "$\mathcal{S}_h$-adjoint") is $M^*_{\mathcal{S}_h} = \mathcal{S}_h^{-1} M^T \mathcal{S}_h$, and the corresponding singular values are $s_j(M) = \lambda_j(\mathcal{S}_h^{-1} M^T \mathcal{S}_h M)^{1/2}$.

This leads to the following counterpart of estimate (4): the matrix $E := \tilde{\mathcal{A}}_h^{-1}\mathcal{Q}_h$ and its $\mathcal{S}_h$-adjoint $E^* = \mathcal{S}_h^{-1}\mathcal{Q}_h^T\tilde{\mathcal{A}}_h^{-T}\mathcal{S}_h$ provide the singular values $s_j(E) = \lambda_j(E^*E) = \lambda_j(\mathcal{S}_h^{-1}\mathcal{Q}_h^T\,\tilde{\mathcal{A}}_h^{-T}\mathcal{S}_h\tilde{\mathcal{A}}_h^{-1}\mathcal{Q}_h)^{1/2}$, hence (4) implies

$$\left(\frac{\|r_k\|_{\mathcal{S}_h}}{\|r_0\|_{\mathcal{S}_h}}\right)^{1/k} \le \frac{\|(\tilde{\mathcal{A}}_h^{-1}\mathcal{A}_h)^{-1}\|_{\mathcal{S}_h}}{k} \sum_{i=1}^{k} \lambda_i(\mathcal{S}_h^{-1}\mathcal{Q}_h^T\,\tilde{\mathcal{A}}_h^{-T}\mathcal{S}_h\tilde{\mathcal{A}}_h^{-1}\mathcal{Q}_h)^{1/2} \tag{17}$$

$(k = 1, 2, ..., n)$.

Remark 4.2. Regarding the sum in (17), the sensitivity of the eigenvalues to the parameter κ is determined by the underlying decomposition of $\mathcal{Q}_h$:

$$\mathcal{Q}_h = \begin{bmatrix} 0 & -C_h \\ C_h & -2C_h \end{bmatrix} + \kappa \begin{bmatrix} 0 & B_h \\ -B_h & 0 \end{bmatrix}. \tag{18}$$

This shows on the one hand that the arising eigenvalues in (17), and hence the overall estimate, grows at most linearly with κ. On the other hand, since κ multiplies an antisymmetric matrix in (18), it only contributes to the imaginary parts of the eigenvalues of $\mathcal{Q}_h$. One expects that the overall growth with increasing κ is slight, and this was indeed experienced in a similar situation with complex arithmetics [5].

4.2 Mesh independent superlinear convergence estimates

Proposition 4.1. *The operator $N_S^{-1}Q_S$ is compact in $H^1(\Omega)^2$.*

Proof. Let us introduce the bounded linear operators Q_1 and $Q_2 : H^1(\Omega) \to H^1(\Omega)$, defined by

$$\langle Q_1 u, v\rangle_{H^1} := \int_\Omega uv \qquad \text{and} \qquad \langle Q_2 u, v\rangle_{H^1} := \int_{\partial\Omega} uv \qquad (v \in H^1(\Omega)).$$

The operators Q_1 and Q_2 are compact in $H^1(\Omega)$ (see, e.g., [4, Prop. 3.1]; this in fact follows from the compact embeddings of $H^1(\Omega)$ into $L^2(\Omega)$ and of $H^1(\Omega)_{|\partial\Omega}$ into $L^2(\partial\Omega)$). The operator Q_S can then be expressed as an operator matrix with compact entries:

$$Q_S = \begin{bmatrix} 0 & -\varepsilon Q_1 + (\kappa-\mu)Q_2 \\ \varepsilon Q_1 - (\kappa-\mu)Q_2 & -2(\varepsilon Q_1 + \mu Q_2) \end{bmatrix}$$

(as an operator analogue of (16)), i.e. for all (u, v) and $(z, w) \in H^1(\Omega)^2$ we have

$$\left\langle Q_S \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} z \\ w \end{pmatrix} \right\rangle_{(H^1)^2}$$

$$= -\varepsilon\langle Q_1 v, z\rangle + (k-\mu)\langle Q_2 v, z\rangle + \varepsilon\langle Q_1 u, w\rangle - (k-\mu)\langle Q_2 u, w\rangle - 2\varepsilon\langle Q_1 v, w\rangle - 2\mu\langle Q_2 u, w\rangle$$

from (15). This implies that Q_S is also compact.

Further, one can see that N_S has a bounded inverse in $H^1(\Omega)^2$. Namely, as an operator analogue of (8), we can write N_S also in an operator matrix form:

$$N_S = \begin{bmatrix} A_S & C_S \\ -C_S & A_S + 2C_S \end{bmatrix}$$

with

$$\langle A_S u, v\rangle := \int_\Omega (\nabla u \cdot \nabla z - \kappa^2 uz) \quad (\forall u, v \in H^1(\Omega)$$

and with

$$C_S := \varepsilon Q_1 + \mu Q_2,$$

where Q_1, Q_2 have been defined above. Similarly as in the matrix case, mentioned in subsection 3.2, one can see as a special case that N_S is injective, i.e., the solution of a homogeneous system is only the pair of zeros. Indeed, considering the system

$$\begin{aligned} A_S u + C_S v &= 0 \\ -C_S u + (A_S + 2C_S)v &= 0, \end{aligned}$$

substraction shows that $(A_S + C_S)(v - u) = 0$, hence the regularity of $(A_S + C_S)$ implies $u = v$ and then the first equation yields $u = v = 0$. Further, the form of N_S shows that it is a compact perturbation of the identity. Therefore, using the Fredholm alternative, injectivity implies that N_S has a bounded inverse.

Finally, since N_S^{-1} is bounded and Q_S is compact, we obtain that $N_S^{-1}Q_S$ is also compact. □

Remark 4.3. The above proof includes the result that N_S has a bounded inverse. As mentioned in the introduction, we have a well-posedness result for the original Helmholtz problem via Fredholm theory. This means, by the definition of the operator L_S, that L_S has a bounded inverse too. Also, similarly as seen above for N_S, the operator L_S is also a compact perturbation of the identity.

In order to formulate the next estimates, we will use the constant

$$M := \|N_S\| \tag{19}$$

and we need the following:

ASSUMPTION 4.2: There exist constants $m_0, m_1 > 0$ such that for all considered subspaces V_h,

$$(b) \qquad \|\mathcal{A}_h^{-1}\|_{\mathcal{S}_h} \leq \frac{1}{m_0}, \qquad \|\tilde{\mathcal{A}}_h^{-1}\|_{\mathcal{S}_h} \leq \frac{1}{m_1}.$$

Note that Assumption 4.2 is a natural requirement. Namely, here $\mathcal{A}_h$ and $\tilde{\mathcal{A}}_h$ are Galerkin discretizations of the operators L_S and N_S, respectively, which have bounded inverses and are compact perturbations of the identity on $H^1(\Omega)^2$ by Remark 4.3. For such operators, it was shown in [5] that their Galerkin discretizations also have uniformly bounded inverses for small enough discretization parameter h, i.e. Assumption 4.2 always holds if h is small enough.

Now we are in the position to readily derive the main estimates and then the final theorem.

Proposition 4.2. *Let Assumption 4.2 hold, let $s_j(Q_S)$ $(j = 1, 2, \dots)$ denote the singular values of the compact operator Q_S and let $M, m_0, m_1 > 0$ be as defined above. Then the following relations hold:*

$$(a) \qquad \sum_{j=1}^{k} \lambda_j(\mathcal{S}_h^{-1}\mathcal{Q}_h^T\, \tilde{\mathcal{A}}_h^{-T}\mathcal{S}_h\tilde{\mathcal{A}}_h^{-1}\mathcal{Q}_h)^{1/2} \leq \frac{1}{m_1}\sum_{j=1}^{k} s_j(Q_S) \qquad (j = 1, 2, \dots, n),$$

$$(b) \qquad \|(\tilde{\mathcal{A}}_h^{-1}\mathcal{A}_h)^{-1}\|_{\mathcal{S}_h} \leq \frac{M}{m_0}.$$

Proof. The desired estimates hold in a Hilbert space whenever $\mathcal{A}_h, \tilde{\mathcal{A}}_h, \mathcal{Q}_h$ and $\mathcal{S}_h$ are Galerkin projections of operators L_S, N_S, Q_S and of the inner product of the space, respectively, such that L_S and N_S have bounded inverses and Q_S is compact. This has been proved in [4, Proposition 4.5] in the case of coercive preconditioners for the sum of eigenvalues without square roots in (a), further, it has been pointed out in [5, Proposition 5.4] that coercivity can be replaced here just by bounded invertibility and the estimates hold for each term of the above sum in (a). Hence the result follows, since the required properties for such a setting have been verified above, see Proposition 4.1 and Remark 4.3. □

Theorem 4.1. *Let us consider a family of FEM subspaces $V_h = span\{\varphi_1, \dots, \varphi_n\} \subset H^1(\Omega)$ $(h > 0)$ and the discretization in $V_h \times V_h$ described in Section 3, such that Assumption 4.2 holds. Then the GMRES iteration for the $n \times n$ preconditioned systems (10) provides a mesh independent superlinear convergence estimate, i.e., we have*

$$\left(\frac{\|r_k\|_{\mathcal{S}_h}}{\|r_0\|_{\mathcal{S}_h}}\right)^{1/k} \leq \varepsilon_k \qquad (k = 1, 2, ..., n) \tag{20}$$

where $(\varepsilon_k)_{k\in\mathbf{N}^+} \to 0$ and it is a sequence independent of n and V_h. Namely,

$$\varepsilon_k \leq \frac{M}{m_0 m_1} \cdot \frac{1}{k}\sum_{j=1}^{k} s_j(Q_S) \to 0 \qquad (as\ k \to \infty), \tag{21}$$

Proof. The estimate follows directly from (17) and Proposition 4.2. The convergence of ε_k to zero follows from the compactness of Q_S, since it is constant times the arithmetic mean of a sequence that converges to 0. □

5 Conclusions

We have considered the real system formulation of complex Helmholtz equations by rewriting them in real-valued block form. We have introduced and analyzed a preconditioner in a special two-by-two block form. This block preconditioner can be readily factorized and thus reduced to two standard systems with symmetric positive definite matrices, which leads to a small condition number bound. We have shown that the corresponding preconditoned Krylov iteration converges at a mesh independent superlinear rate.

Acknowledgements. The research of O. Axelsson was supported by the Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project "IT4 Innovations excellence in science LQ1602". The research of J. Karátson was supported by the Hungarian Scientific Research Fund OTKA, No. 112157.

References

[1] Axelsson, O., *Iterative Solution Methods,* Cambridge University Press, 1994.

[2] Axelsson, O., Farouq, S., Neytcheva, M., A preconditioner for optimal control problems, constrained by Stokes equation with a time-harmonic control, *J. Comput. Appl. Math.* 310 (2017), 5-–18.

[3] Axelsson, O., Karátson J., Superlinearly convergent CG methods via equivalent preconditioning for nonsymmetric elliptic operators, *Numer. Math.* 99 (2004), No. 2, 197–223.

[4] Axelsson, O., Karátson J., Mesh independent superlinear PCG rates via compact-equivalent operators, *SIAM J. Numer. Anal.*, 45 (2007), No.4, pp. 1495–1516.

[5] Axelsson, O., Karátson J., Magoules, F., Superlinear convergence under complex shifted Laplace preconditioners for Helmholtz equations, www.cs.elte.hu/˜karatson/Helmholtz-preprint.pdf

[6] Axelsson, O., Neytcheva, M., Ahmad, B., A comparison of iterative methods to solve complex valued linear algebraic systems, *Numer. Algor.* 66 (2014), 811-–841.

[7] Erlangga, Y.A., Vuik, C. and Oosterlee, C.W., A novel multigrid based preconditioner for heterogeneous Helmholtz problems, *SIAM J. Sci. Comput.* 27 (2006), no. 4, 1471—1492.

[8] O. G. Ernst and M. J. Gander, Why it is difficult to solve Helmholtz problems with classical iterative methods, *in:* Numerical Analysis of Multiscale Problems, I.G. Graham, T.Y. Hou, O. Lakkis and R. Scheichl, editors, Proceedings of an LMS Durham Symposium 2010, LNCS 83, Springer Verlag, 2012.

[9] M. J. Gander, I. G. Graham, 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.

[10] M. Gander, F. Magoulès and F. Nataf, Optimized Schwarz methods without overlap for the Helmholtz equation, *SIAM J. Sci. Comp.* 24 (2002), 38—60. rlin, 1985.

[11] R. Hiptmair, Operator preconditioning, *Computers and Mathematics with Applications*, 52 (2006), pp. 699—706.

[12] Krutitskii, P. A., The impedance problem for the propagative Helmholtz equation in interior multiply connected domain, *Comp. Math. Appl.* 46(10-11) (2003), 1601–1610.

[13] Livshits, I., Use of Shifted Laplacian Operators for Solving Indefinite Helmholtz Equations, *Numerical Mathematics: Theory, Methods and Applications*, 8 (01) 2015, pp. 136–148.

[14] F. Magoulès, K. Meerbergen, and J.-P. Coyette, Application of a domain decomposition method with Lagrange multipliers to acoustic problems arising from the automotive industry. *J. Comput. Acoustics*, 8(3) (2000), 503–521.

[15] Moret, I., A note on the superlinear convergence of GMRES, *SIAM J. Numer. Anal.* 34 (1997), 513–516.

[16] Saad, Y., *Iterative Methods for Sparse Linear Systems*, Second Edition, SIAM, 2003.

[17] Simoncini, V., Szyld, D. B., On the Occurrence of Superlinear Convergence of Exact and Inexact Krylov Subspace Methods, *SIAM Review* , 2005, Vol. 47, No. 2, pp. 247-272.

[18] Winter, R., Some superlinear convergence results for the conjugate gradient method, *SIAM J. Numer. Anal.*, 17 (1980), 14–17.

[19] Widlund, O., A Lanczos method for a class of non-symmetric systems of linear equations, *SIAM J. Numer. Anal.,* 15 (1978), 801–812.