# How Large a Shift is Needed in the Shifted Helmholtz Preconditioner for its Effective Inversion by Multigrid?

Pierre-Henri Cocquet, Martin Gander

# HOW LARGE A SHIFT IS NEEDED IN THE SHIFTED HELMHOLTZ PRECONDITIONER FOR ITS EFFECTIVE INVERSION BY MULTIGRID ?

PIERRE-HENRI COCQUET AND MARTIN J. GANDER

**Abstract.** The shifted Helmholtz operator has received a lot of attention over the past decade as a preconditioner for the iterative solution of the Helmholtz equation. The idea is that if one uses a small complex shift, the shifted Helmholtz operator is still close to the original Helmholtz operator and could thus be an effective preconditioner. It was shown in [17] that the shift can be at most $O(k)$ to prove rigorously wave number independent convergence of the preconditioned system solved with GMRES, provided the preconditioner is inverted exactly. In practice however the preconditioner is inverted only approximately, and if one shifts enough, this can be done effectively by standard multigrid methods. We show in this paper that for a finite element discretization, the shift has to be at least $O(k^2)$ to be able to invert the shifted Helmholtz preconditioner using multigrid. There is therefore a gap between being a good preconditioning operator (shift at most $O(k)$) and being able to effectively invert the preconditioner by multigrid (shift at least $O(k^2)$). So what shift should be chosen in practice, and when the preconditioner is not inverted exactly? By studying the numerical range of the preconditioned operator, we show that one can not obtain analytical results for this case with currently available tools. We thus test the preconditioner extensively numerically for a wave guide type square domain in the range of shifts between $O(\sqrt{k})$ and $O(k^2)$ with approximate inversion by one multigrid V-cycle. We find in our experiments that preconditioned GMRES iteration numbers will then inevitably grow like $O(k^2)$. We also see that in contrast to common practice where shifts of $O(k^2)$ are used, it might be beneficial for the wave guide to use a smaller shift, e.g. $O(k^{3/2})$, especially when several smoothing steps are used.

**Key words.** Multigrid methods, Helmholtz equation, Shifted Helmholtz Preconditioner, Finite element.

**AMS subject classifications.** 65N55, 35J05, 65N30, 65F08, 65L20.

## 1. Introduction.

Multigrid methods are iterative methods widely used to solve linear systems coming from the discretization of partial differential equations. Their main feature is their fast convergence, and that convergence does not deteriorate when the mesh size decreases for certain classes of problems, see [24, 30]. The indefinite Helmholtz equation involves the continuous operator $\mathcal{H}_k = (\Delta + k^2)$, where $k > 0$ is the wave number. Since the operator $\mathcal{H}_k$ is not coercive, the extension of multigrid methods to such problems remains challenging, and standard multigrid fails to converge, see for example [1, 12, 14, 15, 5], and references therein. It is worth noting that a convergent multigrid method has been designed using a dispersion correction technique in [15]. Unfortunately, this dispersion correction which matches the discrete and continuous dispersion relations at the coarse level does not extend to higher dimensions.

Many methods for solving numerically Helmholtz-like problems have been designed over the years, for an overview, see [14]. Among these, Krylov subspace methods like GMRES or BiCGStab are mostly used because of their robustness. Unfortunately, these methods are not fast enough without a good preconditioner. Incomplete LU (ILU) [29] preconditioners have been developed for the Helmholtz equation in the form of Analytic ILU (AILU) [20], which is based on a factorization of the operator, but the analysis is difficult to extend to inhomogeneous media. The so-called analytic preconditioners, see [10] and references therein, are based on approximate inverses of some pseudo-differential operators that are localized by Padé approximations, involved in the integral equation coming from the Helmholtz equation in exterior domains. These methods are also well-suited when studying homogeneous

media. Several successful preconditioners rely on domain decomposition methods, see [19, 18, 34, 22] and references therein. These methods are based on a splitting of the computational domain into many subdomains, where each subproblem can then be solved with a direct method.

In this paper, we want to focus on an idea which received a lot of attention over the last decade, namely to use a shifted Helmholtz operator as a preconditioner, see [13, 21, 35, 31, 5, 25, 27, 32, 28, 33] and references therein. The latter is defined as a discretization of the operator $\mathcal{H}_{\widetilde{k}}$ for $\widetilde{k}^2 = k^2 + \mathrm{i}\varepsilon$, where $\varepsilon > 0$ is the so-called shift. The main advantage of the shifted Helmholtz preconditioner is its simplicity, and that it can be used for media with varying wave number (see e.g. [13, 33]). In addition, the method is based on the discretization of a continuous problem and therefore inherits many of its properties. It is well known that the shift has to be large enough for standard multigrid methods to be effective to invert the shifted operator, but not too large to still be a good preconditioner for the underlying original Helmholtz problem, see [14, 17]. Recently, the question of the minimal value for the shift has been explored numerically, see [8, 28]. In [8], the authors show using numerical evaluation of quantities obtained from Fourier analysis that the minimal shift depends in an irregular way on the wave number, the mesh size and on the number of pre- and post-smoothing steps. The strategies for choosing the shift from [28] are based on the compromise that one wants to improve stability by maximizing the diagonal entries of the discrete shifted Helmholtz operator without losing accuracy of the preconditioner. The resulting shifts are then locally chosen with respect to each row of the finite difference discretization of $\mathcal{H}_k$. Unfortunately, shifts obtained in [8, 28] heavily rely on the discretization and therefore cannot be linked to a discretization of a continuous Helmholtz-like problem. A one-dimensional shifted Helmholtz equation discretized with standard finite differences was studied analytically in [6], and it was shown that the multigrid algorithm using a Jacobi smoother with a complex damping parameter converges if the shift behaves like $O(k^2)$. Similar results on the size of the shift needed for solving the shifted Helmholtz problem with an additive Schwarz method also hold, see [22].

The goal of this paper is to understand the influence of the complex shift $\varepsilon$ on the convergence of the standard multigrid algorithm applied to the shifted Helmholtz problem in a finite element context, and its implications for the use of the shifted Helmholtz problem as a preconditioner. From a theoretical point of view, our main result states that for a given wave number $k$, taking $\varepsilon = O(k^2)$ with a large enough constant always ensures convergence of the multigrid method applied to the shifted Helmholtz problem. It has been recently proved in [17] that the shifted Helmholtz operator is a robust preconditioner for the Helmholtz equation provided that the shift is not bigger than $O(k)$. Our results then show that in the shifted Helmholtz preconditioner one has to live with a compromise between having an effective preconditioner (shift at most $O(k)$) and a robust multigrid solver (shift at least $O(k^2)$). This was conjectured already by a preliminary Fourier analysis in [14]. The question of what the best shift is can however not be answered by our analysis, since the preconditioner is usually not inverted exactly, but only approximately using a few iterations of multigrid or domain decomposition. We thus study this question numerically by computing the numerical range of the preconditioned problem, and the performance of preconditioned GMRES for shifts ranging from $O(\sqrt{k})$ to $O(k^2)$ for a square wave guide type problem. The numerical range computations show that it will be difficult to obtain rigorous theoretical results on what power of $k$ should be used in the shift for

best performance when approximately inverting the shifted Helmholtz preconditioner. Our iteration counts from preconditioned GMRES grow like $O(k^2)$ with this preconditioner for all shifts ranging from $O(\sqrt{k})$ to $O(k^2)$. We also often obtain substantially lower iteration counts with a shift $O(k^{3/2})$ for our wave guide problem, compared to the traditionally used shift of $O(k^2)$, for which our analysis shows convergence of the multigrid method. Nevertheless, our analysis is of interest: a shift of $O(k^2)$ is also used in [9], where the solution of the original Helmholtz equation is replaced with an infinite number of inversions of $k^2$-shifted Helmholtz problems using the first resolvent formula. The concept of high-order shifted Helmholtz preconditioners [35], whose design is based on Padé's approximants, also uses a shift behaving like $k^2$.

Our paper is organized as follows: in the rest of the introduction, we define the shifted Helmholtz equation, its finite element discretization and the multigrid algorithm. In section 2, we use Fourier local mode analysis to study a two-grid algorithm with a damped Jacobi smoother for the one dimensional shifted Helmholtz equation with Dirichlet boundary conditions. We show that if the shift is less than $O(k^2)$, the two-grid method will diverge for certain wave number and mesh size combinations. We also prove that if the shift is $O(k^2)$ with a precise estimate for the minimal constant, the standard two-grid method will converge for all wave number and mesh size combinations. To study the performance of a multigrid algorithm for the shifted Helmholtz equation in higher spatial dimensions, we then use techniques for estimating the smoothing and approximation properties from [30], and we state in Section 3 all the properties of the continuous problem needed for this purpose. In Section 4 we introduce a small modification in the smoother, which allows us to prove convergence of a W-cycle applied to a general shifted multi-dimensional Helmholtz operator discretized by finite elements, and this for an arbitrarily small complex shift! Our analysis puts however an upper bound on the damping parameter in the smoother, and this upper bound goes to zero as the wave number grows. The method thus needs too many smoothing iterations for practical purposes, except if again the complex shift is $O(k^2)$. In Section 5 we extend our multigrid results to the case of impedance boundary conditions. In Section 6 we illustrate our multigrid analysis results based on the smoothing and approximation properties for one and two dimensional problems. In Section 7, we study some possible directions for future work by exploring the shifted Helmholtz preconditioner numerically. We first consider the case of non-convex, non star-shaped domains, where our numerical simulations indicate that the shifted Helmholtz preconditioner is again a good preconditioner if the shift behaves like $O(k)$, provided the preconditioner is inverted exactly. We then numerically study the best choice of the shift when using only one V-cycle and various numbers of pre- and post-smoothing steps to approximately invert the shifted Helmholtz preconditioner. We present our conclusions in Section 8.

**1.1. The shifted Helmholtz equation.** Let $\Omega$ be a convex polygon of $\mathbb{R}^d$ with $d = 1, 2, 3$. We consider the shifted Helmholtz equation with homogeneous Dirichlet boundary conditions (for impedance boundary conditions, see Section 5),

$$\begin{cases} -\Delta u(x) - (k^2 + \mathrm{i}\varepsilon)u(x) &= f(x),\ x \in \Omega, \\ u|_{\partial\Omega} &= 0. \end{cases} \tag{1.1}$$

Here, $k$ is the wave number, $\varepsilon > 0$ is the shift and $f \in L^2(\Omega)$ is a source term. A Green's formula leads to the variational formulation

$$\begin{cases} \text{Find } u \in H_0^1(\Omega) \text{ such that :} \\ a(u,v) := \displaystyle\int_\Omega \nabla u \cdot \overline{\nabla v} - (k^2 + \mathrm{i}\varepsilon) u \overline{v} dx = \int_\Omega f \overline{v} dx, \ \forall v \in H_0^1(\Omega). \end{cases} \tag{1.2}$$

We consider a finite element discretization of (1.2) with piece-wise linear polynomials. Let $\{\mathcal{T}_l\}$ be a regular family of triangulations of $\Omega$ consisting of $d$-simplices of characteristic size $h_l$. Let $\mathcal{V}_l$ be the corresponding finite element space,

$$\mathcal{V}_l = \left\{ v \in \mathcal{C}(\overline{\Omega}) \mid v|_T \in \mathbb{P}_1 \text{ for all } T \in \mathcal{T}_l, \ v_l|_{\partial\Omega} = 0 \right\}.$$

This leads to the discrete variational problem

$$\begin{cases} \text{Find } u_l \in \mathcal{V}_l \text{ such that :} \\ a(u_l, v_l) = \displaystyle\int_\Omega f \overline{v_l} dx, \ \forall v_l \in \mathcal{V}_l. \end{cases} \tag{1.3}$$

Let $(\phi_j)_{j=1}^{N_l}$ be the standard nodal basis of $\mathcal{V}_l$. This induces an isomorphism between the vector representation of unknowns and the finite element functions,

$$F_l : \mathbb{C}^{N_l} \longrightarrow \mathcal{V}_l, \ F_l \mathbf{z} = \sum_{j=1}^{N_l} z_j \phi_j. \tag{1.4}$$

Problem (1.3) can now be represented as a linear system

$$A_l \mathbf{z}_l = \mathbf{b}_l, \text{ with } (A_l)_{i,j} = a(\phi_i, \phi_j) \text{ and } (b_l)_j = \int_\Omega f \phi_j dx, \tag{1.5}$$

and the Galerkin finite element solution is then $u_l = F_l \mathbf{z}_l$.

REMARK 1.1. *From the definition of the isomorphism $F_l$ one has, for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{C}^{N_l}$ that*

$$a(F_l \mathbf{z}_1, F_l \mathbf{z}_2) = \langle A_l \mathbf{z}_1, \overline{\mathbf{z}_2} \rangle,$$

*where the brackets stand for the usual dot product in $\mathbb{R}^{N_l}$, $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{j=1}^{N_l} x_j y_j$.*

**1.2. The multigrid method.** Our goal is to solve the linear system (1.5) coming from the discretization of (1.1) with a standard multigrid method like the ones presented in [24, 30]. To do so, we assume that the triangulation at level $l$ is obtained from a coarser one at level $l-1$ by refinement, implying the nesting property

$$\mathcal{V}_{l-1} \subset \mathcal{V}_l,$$

where the mesh size of the coarser grid satisfies $h_{l-1} \leq c h_l$ for some constant $c \geq 2$. Multigrid methods use so-called restriction and prolongation operators. Since $\mathcal{V}_{l-1} \subset \mathcal{V}_l$, the identity operator $I_l : \mathcal{V}_{l-1} \longrightarrow \mathcal{V}_l$ is well-defined. This identity operator represents linear interpolation and acts as

$$\forall\, u_{l-1} \in \mathcal{V}_{l-1}, \ (I_l u_{l-1})(x) = \sum_{j=1}^{N} u_{l-1}(x_j^l) \phi_j(x),$$

where $x_j^l$ are the nodes of the fine mesh. This operator is called the prolongation operator and has the matrix representation

$$P_l : \mathbb{C}^{N_{l-1}} \mapsto \mathbb{C}^{N_l},\ P_l := F_l^{-1} F_{l-1}. \tag{1.6}$$

Note that the matrix representation of the prolongation operator can be obtained by

$$\forall\, \mu \in \{1, \cdots, N_{l-1}\},\ (I_l \phi_\mu^{l-1})(x) = \sum_{j=1}^{N_l} \phi_\mu^{l-1}(x_j^l)\phi_j^l(x),$$

where $\phi_j^l$ are the nodal basis functions related to $\mathcal{V}_l$.

According to [24] the canonical restriction operator is defined as

$$R_l := P_l^*. \tag{1.7}$$

For elliptic problems, this choice ensures that $A_{l-1} = R_l A_l P_l$ (see e.g Lemma 3.2 of [30]). This property still holds for the shifted Helmholtz equation as one can see from Lemma 9.1 in the appendix.

REMARK 1.2. *Prolongation and restriction operators are classically acting on real vectors and can thus be represented as real matrices. In this case, the star in $R_l := P_l^*$ actually stands for the usual transpose. Moreover, when acting on complex vectors, one only has to compute $R_l z = R_l \mathcal{R}z + \mathrm{i} R_l \mathcal{I}z$.*

For a general linear system $A_l \mathbf{z}_l = \mathbf{b}_l$, the multigrid algorithm [30, 24], with smoother $S$, $\nu_1$ pre-smoothing steps and $\nu_2$ post-smoothing steps is given by

$$\left\{\begin{array}{ll}
\text{function } \mathbf{z}_l = MGM_l(\mathbf{z}_l, \mathbf{b}_l) & \\
\quad \text{if } \mathrm{l} = 0 \text{ then } \mathbf{z}_0 = A_0^{-1}\mathbf{b}_0 \text{ else} & \\
\qquad \mathbf{z}_l = S^{\nu_1}(\mathbf{z}_l, \mathbf{b}_l); & \% \text{ pre} - \text{smoothing} \\
\qquad \mathbf{d}_{l-1} = R_l(\mathbf{b}_l - A_l \mathbf{z}_l); & \\
\qquad \mathbf{e}_{l-1}^0 = \mathbf{0}; & \\
\qquad \text{for } j = 1 \text{ to } \tau \text{ do} & \\
\qquad\quad \mathbf{e}_{l-1}^j = MGM_{l-1}(\mathbf{e}_{l-1}^{j-1}, \mathbf{d}_{l-1}); & \\
\qquad \text{end} & \\
\qquad \mathbf{z}_l = \mathbf{z}_l + P_l \mathbf{e}_{l-1}^\tau; & \\
\qquad \mathbf{z}_l = S^{\nu_2}(\mathbf{z}_l, \mathbf{b}_l); & \% \text{ post} - \text{smoothing} \\
\quad \text{end} &
\end{array}\right. \tag{1.8}$$

In (1.8), the classical "V-cycle" is obtained for $\tau = 1$ and the so-called "W-cycle" is obtained for $\tau = 2$.

It is easy to see that the multigrid algorithm (1.8) is a linear stationary iterative method. From [30] Theorem 7.1 p.22 (see also [24] Lemma 7.14), its iteration matrix is given by

$$\begin{array}{ll}
C_{MG,0} & = 0, \\
C_{MG,l} & = S_l^{\nu_2}\left(I - P_l\left(I - C_{MG,l-1}^\tau\right) A_{l-1}^{-1} R_l A_l\right) S_l^{\nu_1}.
\end{array} \tag{1.9}$$

Note that $C_{MG,l} = C_{MG,l}(\nu_2, \nu_1)$ where $\nu_1,\ \nu_2$ are the number of pre- and post-smoothing iterations. Moreover, one has for the spectral radius

$$\rho\left(C_{MG,l}(\nu_2, \nu_1)\right) = \rho\left(C_{MG,l}(0, \nu_1 + \nu_2)\right),$$

so we only study the one parameter case $\nu = \nu_1 + \nu_2$.

**2. Analysis of the two-grid operator for a 1D model problem.** The continuous model problem we consider is

$$\left\{ \begin{array}{rl} -u''(x) - (k^2 + \mathrm{i}\varepsilon)u(x) &= f(x),\ x \in (0,1), \\ u(0) &= 0, \\ u(1) &= 0. \end{array} \right. \tag{2.1}$$

We use a uniform mesh with $N_l$ interior points and mesh size $h_l = 1/(N_l+1)$ to approximate (2.1). The basis of $\mathcal{V}_l$ we use are the hat functions

$$\phi_j(x) = \max\left(0, 1 - \frac{|x - jh_l|}{h_l}\right),\ j = 1, \cdots, N.$$

After some algebra, we obtain the matrix of the discrete problem (1.3)

$$A_l = \text{tridiag}\left(-\frac{1}{h_l} - \frac{h_l}{6}(k^2 + \mathrm{i}\varepsilon), \frac{2}{h_l} - \frac{2h_l}{3}(k^2 + \mathrm{i}\varepsilon), -\frac{1}{h_l} - \frac{h_l}{6}(k^2 + \mathrm{i}\varepsilon)\right).$$

Note that, unlike the original Helmholtz equation, the shifted Helmholtz equation (2.1) is well-posed with Dirichlet conditions as soon as $\varepsilon > 0$, and the matrix $A_l$ is then invertible for any mesh size.

The next coarser mesh is then defined by the uniform mesh with $N_{l-1} = N_l/2$ interior points and has a characteristic size given by $h_{l-1} = 1/(N_{l-1}+1)$. The prolongation and restriction operators (see [15, 24, 30, 7]) are defined as

$$P_l := \begin{pmatrix} 1/2 & 0 & & \\ 1 & 0 & & \\ 1/2 & 1/2 & & \\ 0 & 1/2 & & \\ & & \ddots & \\ & & & 1/2 \\ & & & 1 \\ & & & 1/2 \end{pmatrix}, \quad R_l := P_l^T.$$

We focus here on the two-grid operator, which is according to (1.9) given by

$$T = S_l^{\nu_2}\left(I - P_l A_{l-1}^{-1} R_l A_l\right) S_l^{\nu_1}.$$

To simplify the notation, we drop in what follows the index on $h$, i.e. $h \equiv h_l$. The eigenvectors of $A_l$ are $\mathbf{v}_j^h := [\sin j\ell\pi h]_{\ell=1}^{N_l}$, $j = 1, \ldots, N_l$, and thus the block diagonalization performed for the finite difference scheme in [24, 15] still applies. The latter uses the subspaces

$$\text{span}\{\mathbf{v}_1^h, \mathbf{v}_{N_l}^h\}, \text{span}\{\mathbf{v}_2^h, \mathbf{v}_{N_l-1}^h\}, \ldots, \text{span}\{\mathbf{v}_n^h, \mathbf{v}_{n+2}^h\}, \text{span}\{\mathbf{v}_{n+1}^h\}, \tag{2.2}$$

where $n + 1 = 1/(2h)$. Denoting by $j' := n + 1 - j$, the complementary mode index, $c_j := \cos\frac{j\pi h}{2}$ and $s_j := \sin\frac{j\pi h}{2}$, one gets the eigenvalues of $A_l$ at the finer and coarser levels

$$\begin{array}{rcl} \lambda_j^h &=& \frac{2}{h} - \frac{2h}{3}(k^2 + \mathrm{i}\varepsilon) - 2(\frac{1}{h} + \frac{h}{6}(k^2 + \mathrm{i}\varepsilon))(1 - 2s_j^2), \\ \lambda_{j'}^h &=& \frac{2}{h} - \frac{2h}{3}(k^2 + \mathrm{i}\varepsilon) - 2(\frac{1}{h} + \frac{h}{6}(k^2 + \mathrm{i}\varepsilon))(1 - 2c_j^2), \\ \lambda_j^H &=& \frac{2}{H} - \frac{2H}{3}(k^2 + \mathrm{i}\varepsilon) - 2(\frac{1}{H} + \frac{H}{6}(k^2 + \mathrm{i}\varepsilon))(1 - 8s_j^2 c_j^2), \end{array} \tag{2.3}$$

where $H = 2h$. For the smoother, we consider a standard damped Jacobi iteration,

$$\mathbf{u}_{m+1} = \mathbf{u}_m + \omega D^{-1}(\mathbf{f} - A_l \mathbf{u}_m),$$

where $D = \mathrm{diag}(A_l)$, and $\omega$ is a damping parameter. The latter is chosen as the one giving equi-oscillation on the oscillatory part of the spectrum where the wave number is replaced by the shifted wave number. This requires, when $\varepsilon = 0$, that $|1 - w| = |1 - w\lambda_{N_l}|$ and yields

$$\omega = \frac{12 - 4h^2(k^2 + \mathrm{i}\varepsilon)}{18 - 3h^2(k^2 + \mathrm{i}\varepsilon)}. \tag{2.4}$$

Finally, denoting by $\sigma_j$ the eigenvalues of $S = I - \omega D^{-1} A_l$, the two-grid operator can be written as $\mathrm{diag}(T_1, \ldots, T_n, T_{n+1})$ with

$$T_j = \begin{bmatrix} \sigma_j & 0 \\ 0 & \sigma_{j'} \end{bmatrix}^{\nu_2} \begin{bmatrix} 1 - 2c_j^4 \frac{\lambda_j^h}{\lambda_j^H} & 2c_j^2 s_j^2 \frac{\lambda_{j'}^h}{\lambda_j^H} \\ 2c_j^2 s_j^2 \frac{\lambda_j^h}{\lambda_j^H} & 1 - 2s_j^4 \frac{\lambda_{j'}^h}{\lambda_j^H} \end{bmatrix} \begin{bmatrix} \sigma_j & 0 \\ 0 & \sigma_{j'} \end{bmatrix}^{\nu_1}, \; T_{n+1} = \sigma_{n+1}^{\nu_1+\nu_2}. \tag{2.5}$$

We first study the case of a shift $\varepsilon = Ck^{2-\delta}$, $0 \le \delta \le 2$.

THEOREM 2.1. *Assume that we are performing $\nu = \nu_1 + \nu_2$ smoothing steps and that $\varepsilon = Ck^{2-\delta}$ for $0 < \delta \le 2$. Then, for all wave number and mesh sizes such that*

$$kh = \sqrt{6} + o(1) \text{ as } h \to 0,$$

*one has*

$$\rho(T) \ge \left(\frac{2}{3}\frac{6^{\delta/2}}{Ch^\delta}\right)^\nu + o\left(\frac{1}{h^{\delta\nu}}\right),$$

*and the two-grid algorithm will diverge.*

*Proof.* Because of the block diagonal form of the two-grid iteration matrix, one has

$$\rho(T) \ge \rho(T_j), \; j = 1, \cdots, n+1.$$

Taking $j = n + 1$ yields

$$\rho(T) \ge |1 - \omega|^\nu = \left|\frac{6 + k^2h^2 + \mathrm{i}h^2\varepsilon}{18 - 3kh^2 - 3\mathrm{i}\varepsilon h^2}\right|^\nu = |\sigma_{n+1}(k)|^\nu.$$

We now need to find the maximum of $|1 - \omega|^2$ as a function of the wave number $k$. Its derivative is

$$\partial_k |\sigma_{n+1}(k)|^2 = \frac{16h^2}{3} \frac{(k^{2\delta+5}h^4C^2\delta - k^{2\delta+5}h^4C^2 - k^{4\delta+5}h^4 + 36k^{4\delta+1})}{(h^4C^2k^4 + h^4k^{4+2\delta} - 12h^2k^{2+2\delta} + 36k^{2\delta})^2},$$

and the maximum is thus reached at $k(h)$ satisfying

$$h^4 k(h)^{2\delta+4} - 36k(h)^{2\delta} + C^2h^4(1-\delta)k(h)^4 = 0.$$

Since we can not compute exactly $k(h)$ (except when $\delta = 0, 1$), we compute an asymptotic expansion of $k(h)$ as $h$ goes to 0 which reduces to find a constant $\alpha_0$ such that

$$k(h) = \frac{\alpha_0}{h} + o\left(\frac{1}{h}\right).$$

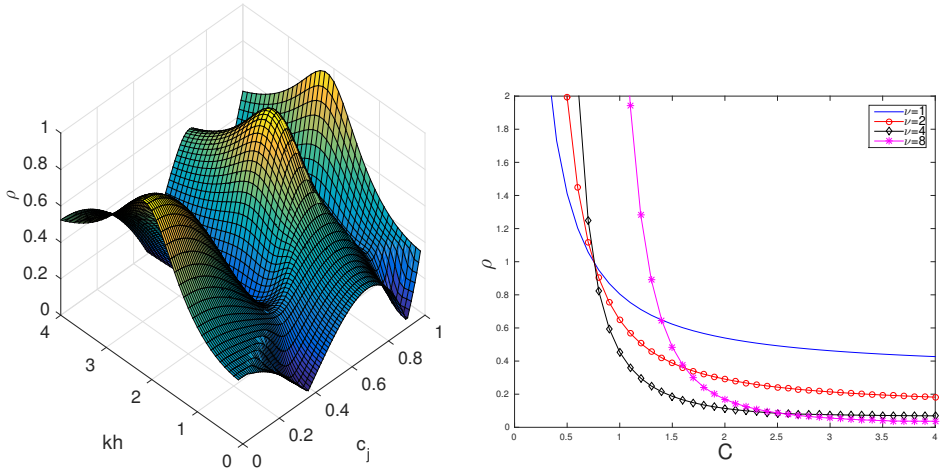


FIG. 2.1. *Using the heuristically optimized damping parameter (2.4) and shift $\varepsilon = Ck^2$. Left: spectral radius of $T_j$ with $\nu = 1$ for $C = 0.8$ as a function of $kh$ and $c_j$. Right: spectral radius of $T$ as a function of $C$ for various numbers of smoothing steps.*

Direct computations then show that for $h$ small enough

$$k(h)^4 = \frac{\alpha_0^4}{h^4} + o\left(\frac{1}{h^2}\right), \; k(h)^{2\delta} = \frac{\alpha_0^{2\delta}}{h^{2\delta}} + o\left(\frac{1}{h^{2\delta}}\right).$$

Replacing these formulas into the equation satisfied by $k(h)$ and identifying terms having same homogeneity in $h$ gives

$$\frac{1}{h^{2\delta}}\left(-36\alpha_0^{2\delta} + \alpha_0^{2\delta+4}\right) + o\left(\frac{1}{h^{2\delta}}\right) = 0.$$

Therefore $\alpha_0 = \sqrt{6}$ and one can check that this is indeed asymptotically a maximum. Inserting the asymptotic formula of $k(h)$ into $|\sigma_{n+1}|$ shows that, for $h$ small enough

$$\rho(T) \geq \left(\frac{1}{9}\frac{C^2h^{2\delta} + 46^\delta}{C^2h^{2\delta}} + o\left(\frac{1}{h^{2\delta}}\right)\right)^{\nu/2} \geq \left(\frac{2}{3}\frac{6^{\delta/2}}{Ch^\delta}\right)^\nu + o\left(\frac{1}{h^{\delta\nu}}\right),$$

which gives the desired result. □

REMARK 2.2. *In the proof of Theorem 2.1, we only give the first term of the asymptotic expansion of $k(h)$ since this suffices to obtain divergence; the asymptotic expansion could be computed to any order without difficulties.*

Theorem 2.1 shows that if the shift is less than $O(k^2)$, the two-grid method, and hence the multigrid method, will diverge if certain wave number-mesh size combinations appear in the multi-grid mesh hierarchy. For the expected convergence when $\varepsilon = Ck^2$, we have the following result:

THEOREM 2.3. *Assume that $\varepsilon = Ck^2$ with large enough constant $C > 0$, then the two-grid algorithm converges.*

To obtain this result, one can substitute $\varepsilon = Ck^2$ and $s_j = \sqrt{1 - c_j^2}$ into the block-diagonal form (2.5) of $T$, which leads to a function that only depends on the product $kh > 0$ and $c_j \in [0, 1]$. We show in Figure 2.1 on the left the spectral radius of the matrix $T_j$ for one smoothing step, $\nu = 1$, as a function of $kh$ and $c_j$ for $C = 0.8$

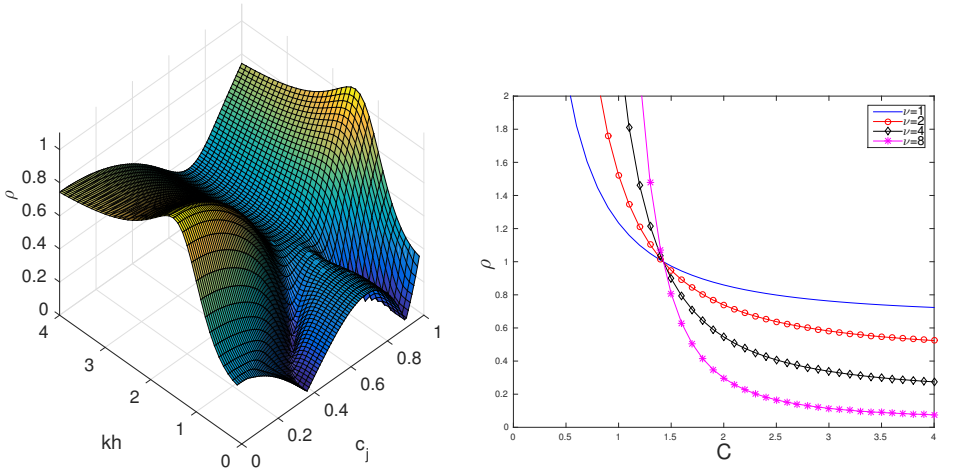


FIG. 2.2. *Using the more classical damping parameter $\omega = 2/3$ and $C = 1.25$ for the same experiment as in Figure 2.1.*

to illustrate that it is uniformly less than one. On the right in Figure 2.1, we show the maximum spectral radius over all $kh$ and $c_j$, i.e. the spectral radius of $T$, as a function of $C$, for various numbers of smoothing steps $\nu = 1, 2, 4, 8$. We clearly see that for $C$ small, the two grid method does not converge, even with a shift of the form $\varepsilon = Ck^2$. For larger values of $C$ however, depending on the number of smoothing steps, we get convergence, the spectral radius is less than one, and for $C > 2.5$, we observe the expected multigrid behavior for Laplace like problems, where increasing the number of smoothing steps improves the performance of the two grid method.

We show in Figure 2.2 for comparison the same results when using the more classical Jacobi damping parameter $\omega = 2/3$. We see from this experiment that for the range of smoothing steps $\nu$ considered, it is better to choose the heuristically optimized damping parameter, instead of $\omega = 2/3$.

We now give some numerical results to illustrate Theorem 2.1 and Theorem 2.3. Consider a uniform grid on $(0, 1)$ with $N_l = 2^{10} - 1$ interior points and the three wave numbers

$$k_1 = \frac{\sqrt{6}}{h_l} = 2508.3, \quad k_2 = \frac{1}{2h_l} = 512, \quad k_3 = h_l^{-2/3} = 101.6.$$

The first wave number $k_1$ corresponds for the given $h_l$ to the troublesome $k$ from Theorem 2.1, $k_2$ corresponds to a minimum rule of thumb resolution, and $k_3$ is the case where no pollution effect occurs, see [26] and also [36]. We show the corresponding spectral radii of $T$ in Table 2.1 for $\nu = 1, 2, 4$. One can see that the two-grid algorithm diverges for $k_1 h_l = \sqrt{6}$, except if the shift is large enough, as expected from Theorem 2.1, which gives a lower bound for $\rho(T)$ that is achieved for $k(h)h = \sqrt{6} + o(1)$ as $h \to 0$. Therefore, $k_1$ is too close to this critical value and yields divergence, except when the shift is large enough, as predicted by Theorem 2.3. The two-grid algorithm also diverges if $k = k_2$ when $\varepsilon \neq k^2$, which means that the first two levels in a multilevel algorithm would already lead to divergence. Finally, for the first two levels of a mesh with no-pollution effect ($k = k_3$), all goes well since the spectral radius of $T$ is smaller than 1 even for $\varepsilon$ that does not depend on $k$. This does however not mean that a multigrid method would also work, since on coarser levels, the algorithm would

| $\varepsilon$ | 1 | $\sqrt{k}$ | $k$ | $k^{3/2}$ | $k^2$ |
|---|---|---|---|---|---|
| $k = k_1,\ \nu = 1$ | $4.1943.10^6$ | $8.3748.10^4$ | $1.6722.10^3$ | 33.3906 | 0.7454 |
| $\nu = 2$ | $1.7592.10^{13}$ | $7.0137.10^9$ | $2.7962.10^6$ | $1.1149.10^3$ | 0.5556 |
| $\nu = 4$ | $3.0949.10^{26}$ | $4.9191.10^{19}$ | $7.8188.10^{12}$ | $1.2433.10^6$ | 0.3221 |
| $k = k_2,\ \nu = 1$ | 40.1940 | 40.1460 | 26.9411 | 1.6269 | 0.3623 |
| $\nu = 2$ | 29.4072 | 29.3721 | 19.7105 | 1.1755 | 0.1623 |
| $\nu = 4$ | 31.8974 | 31.8593 | 21.3795 | 1.2722 | 0.1141 |
| $k = k_3,\ \nu = 1$ | 0.5093 | 0.5089 | 0.4732 | 0.3454 | 0.3344 |
| $\nu = 2$ | 0.1626 | 0.1625 | 0.1605 | 0.1235 | 0.1138 |
| $\nu = 4$ | 0.1202 | 0.1200 | 0.1052 | 0.0633 | 0.0633 |

TABLE 2.1
*Spectral radius of $T$ for $k = k_1, k_2, k_3$.*

encounter wave number-mesh size combinations like in the example for $k_1$ and $k_2$, and thus not converge in general.

**3. Properties of the continuous problem.** We have so far only studied a two-grid algorithm in one spatial dimension. To investigate the properties of a true multi-grid algorithm applied to the shifted Helmholtz equation, also in higher spatial dimensions, we will need to obtain norm estimates for the smoothing and approximation property. We start by proving properties of the shifted Helmholtz problem with Dirichlet boundary conditions which we will be needed in what follows.

THEOREM 3.1. *For all $u, v \in H_0^1(\Omega)$, the following properties hold:*

1. *The bilinear form $a$ is continuous,*

$$\begin{aligned} |a(u,v)| &\leq (1 + C_P^2|k^2 + \mathrm{i}\varepsilon|)\|\nabla u\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} \\ &=: C_{\mathrm{cont}}\|\nabla u\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)}, \end{aligned}$$

   *where $C_P$ is any constant such that $\|u\|_{L^2(\Omega)} \leq C_P\|\nabla u\|_{L^2(\Omega)}$.*
2. *For any complex number $\gamma \in \mathbb{C}$ such that $|\gamma| = 1$, $0 < \mathcal{R}\gamma \leq \varepsilon/|k^2 + \mathrm{i}\varepsilon|$ and $\mathcal{I}\gamma > 0$, one has*

$$\mathcal{R}\gamma a(u,u) \geq \min\{\mathcal{R}\gamma, \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma\}\|\nabla u\|_{L^2(\Omega)}^2 =: C_{\mathrm{coer}}\|\nabla u\|_{L^2(\Omega)}^2. \quad (3.1)$$

3. *The bilinear form is coercive,*

$$|a(u,u)| \geq C_{\mathrm{coer}}\|\nabla u\|_{L^2(\Omega)}^2.$$

4. *There is $C(\Omega)$, a strictly positive constant that depends only on $\Omega$, such that for all $f \in L^2(\Omega)$, the solution of (1.1) satisfies*

$$\|u\|_{H^2(\Omega)} \leq C(\Omega)\left(1 + \frac{|k^2 + \mathrm{i}\varepsilon|}{\varepsilon} + \frac{C_P^2}{C_{\mathrm{coer}}}\right)\|f\|_{L^2(\Omega)} =: C(\Omega)C_{H^2}\|f\|_{L^2(\Omega)}. \quad (3.2)$$

*Proof.*

1. The continuity follows from triangle and Poincaré inequalities.
2. Let $\gamma = \alpha + \mathrm{i}\beta$, with $\alpha^2 + \beta^2 = 1$. A direct computation gives

$$\mathcal{R}\gamma a(u,u) = \alpha\|\nabla u\|_{L^2(\Omega)}^2 + (\varepsilon\beta - k^2\alpha)\|u\|_{L^2(\Omega)}^2.$$

We now impose that $\alpha > 0$ and $\varepsilon\beta - k^2\alpha \geq 0$. Using the relation $\alpha^2 + \beta^2 = 1$ then yields the following restriction on $\gamma$

$$0 < \alpha \leq \frac{\varepsilon}{|k^2 + \mathrm{i}\varepsilon|}.$$

Finally, from

$$\mathcal{R}\gamma a(u,u) \geq \min\{\alpha, \varepsilon\beta - k^2\alpha\}(\|\nabla u\|^2_{L^2(\Omega)} + \|u\|^2_{L^2(\Omega)})$$

the desired result follows.

3. The coercivity estimate follows from the fact that $|\gamma| = 1$. Indeed

$$|a(u,u)| = |\gamma a(u,u)| \geq |\mathcal{R}\gamma a(u,u)| \geq \min\{\alpha, \varepsilon\beta - k^2\alpha\}\|\nabla u\|^2_{L^2(\Omega)}.$$

The well-posedness of (1.2) and (1.3) then follows from the Lax-Milgram lemma.

4. Using elliptic regularity, one gets that the solution to problem (1.1) is in $H^2(\Omega) \cap H^1_0(\Omega)$. From [23] (p.199), we have the estimate

$$\|u\|_{H^2(\Omega)} \leq C(\Omega)\left(\|\Delta u\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)}\right),$$

where $C(\Omega) > 0$ is a constant that depends only on $\Omega$. Since $u$ is solution to (1.1), the previous estimate gives

$$\|u\|_{H^2(\Omega)} \leq C(\Omega)\left(\|f\|_{L^2(\Omega)} + |k^2 + \mathrm{i}\varepsilon|\|u\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)}\right). \tag{3.3}$$

From the coercivity estimate for $a(u,u)$, the fact that $a(u,u) = \int_\Omega f\overline{u}dx$ and a Cauchy-Schwarz inequality, we get

$$C_{\mathrm{coer}}\|\nabla u\|^2_{L^2(\Omega)} \leq \|u\|_{L^2(\Omega)}\|f\|_{L^2(\Omega)} \leq C_P\|\nabla u\|_{L^2(\Omega)}\|f\|_{L^2(\Omega)}.$$

Using once more a Poincaré inequality on the left hand side yields

$$\|u\|_{L^2(\Omega)} \leq C_{\mathrm{P}}\|\nabla u\|_{L^2(\Omega)} \leq \frac{C_P^2}{C_{\mathrm{coer}}}\|f\|_{L^2(\Omega)}. \tag{3.4}$$

Now taking the imaginary part of $a(u,u)$ leads to

$$-\varepsilon\|u\|^2_{L^2(\Omega)} = \mathcal{I}\int_\Omega f\overline{u}dx.$$

Using Cauchy-Schwarz again then shows that

$$\|u\|_{L^2(\Omega)} \leq \frac{\|f\|_{L^2(\Omega)}}{\varepsilon}. \tag{3.5}$$

Combining the previous bounds (3.3), (3.4) and (3.5) then gives the desired estimate (3.2).

□

Theorem 3.1 shows the existence and uniqueness of $u \in H^1_0(\Omega) \cap H^2(\Omega)$ satisfying (1.2) and also the well-posedness of problem (1.3).

Remark 3.2.

1) *The complex number $\gamma$ introduced in Theorem 3.1 must depend on both the wave number $k$ and the shift $\varepsilon$. Otherwise the necessary requirements to have a coercive sesquilinear form can not be satisfied, unless we add some restriction on the shift.*
2) *A sharp bound for the coercivity constant of the sesquilinear form $a$ has been obtained in [17] (see Lemma* 3.1 *page 18). The authors assumed that $\varepsilon \leq C(\Omega)k^2$ and proved*

$$\forall \varphi \in H_0^1(\Omega),\ |a(\varphi,\varphi)| \geq C(\Omega)\frac{\varepsilon}{k^2}\left(\|\nabla\varphi\|^2_{L^2(\Omega)} + k^2\|\varphi\|^2_{L^2(\Omega)}\right),$$

*where $C(\Omega)$ are constants only depending on the domain. We now compare this estimate with the one from Theorem 3.1. Choosing*

$$\gamma = \frac{\varepsilon + \mathrm{i}k^2}{|\varepsilon + \mathrm{i}k^2|} = \frac{\varepsilon + \mathrm{i}k^2}{|\mathrm{i}\varepsilon + k^2|}$$

*meets all the requirements for the following estimate to hold:*

$$\forall \varphi \in H_0^1(\Omega),\ |a(\varphi,\varphi)| \geq |\mathcal{R}\gamma a(\varphi,\varphi)| = \mathcal{R}\gamma\|\nabla\varphi\|^2_{L^2(\Omega)} = \frac{\varepsilon}{|k^2 + \mathrm{i}\varepsilon|}\|\nabla\varphi\|^2_{L^2(\Omega)}.$$

*Thus, if $\varepsilon \leq \xi k^2$, one gets*

$$\forall \varphi \in H_0^1(\Omega),\ |a(\varphi,\varphi)| \geq \frac{\varepsilon}{k^2(1+\xi)}\|\nabla\varphi\|^2_{L^2(\Omega)}.$$

*We now proceed as in [17]. Let $\phi_j \in H_0^1(\Omega)$ be the Dirichlet eigenfunctions of $-\Delta$ and $\lambda_j > 0$ their associated eigenvalues. Taking $k = \sqrt{\lambda_j}$ gives*

$$\frac{|a(\phi_j,\phi_j)|}{\|\nabla\phi_j\|^2_{L^2(\Omega)}} = \frac{|\lambda_j - (k^2 + \mathrm{i}\varepsilon)|}{\lambda_j} = \frac{\varepsilon}{k^2},$$

*and thus our bound for the coercivity constant is also sharp in its $(k,\varepsilon)$-dependence.*

**4. Convergence analysis of the multigrid method.** To study the convergence of the multigrid algorithm, we prove that the assumptions of the next theorem are satisfied for the shifted Helmholtz equation (1.1).

THEOREM 4.1 (Theorem 7.20 [30]). *Assume that there are constants $C_A$, $C_S$ and a monotonically decreasing function $g(\nu)$ with $g(\nu) \to 0$ for $\nu \to +\infty$ such that for all $h$*

$$\begin{aligned} \|A_l^{-1} - P_l A_{l-1}^{-1} R_l\|_2 &\leq C_A\|A_l\|_2^{-1}, && \text{(The approximation property)} \\ \|A_l S^\nu\|_2 &\leq g(\nu)\|A_l\|_2,\ \nu \geq 1, && \text{(The smoothing property)} \\ \|S^\nu\|_2 &\leq C_S, \nu \geq 1. && \text{(The multigrid contraction number)} \end{aligned}$$

*If $\tau \geq 2$, then for any $\xi \in (0,1)$ there exists a $\nu_\xi$ such that for all $\nu \geq \nu_\xi$*

$$\|C_{MG,l}\|_2 \leq \xi,$$

*where $C_{MG,l}$ is the iteration matrix of the multigrid method.*

The rest of this section is dedicated to show that the approximation and smoothing property and the multigrid contraction number hold for the shifted Helmholtz

equation for any $\varepsilon > 0$, provided a small modification in the smoother is applied. Theorem 4.1 then ensures that, for a large enough number of smoothing steps, the iteration matrix of the multigrid algorithm is a contraction, and hence this iterative method converges.

REMARK 4.2.

1. *Theorem 4.1 only gives necessary conditions for the convergence of the multigrid algorithm. This result can also be found in [24] (see Theorem 7.1.2 page 161) and both require a large enough number of smoothing steps to get a convergent multigrid solver.*
2. *The proof of Theorem 4.1 (see also Theorem 7.1.2 page 161 from [24]) is algebraic since it only uses the approximation property, the smoothing property and the multigrid contraction number. These results can thus be used for a large class of continuous problems. In addition, both proofs use the fact that the multigrid iteration matrix can be considered as a perturbation of the two-grid operator where the perturbation is small if one does enough smoothing steps. As a result, this states that if the two-grid algorithm converges and the multigrid contraction number holds, then the multigrid algorithm converges. We therefore state all our theoretical results based on this theorem for the multigrid algorithm.*
3. *One could wonder about the assumptions $A_l$ needs to satisfy so that those of Theorem 4.1 hold. These are actually open questions whose answer depends on the problem under study. For instance, the symmetric coercive case is well-understood and a precise estimate on $\|C_{MG,l}\|_2$ can be computed (see e.g. [30] page 35 Theorem 7.29 or [24] page 165 Theorems 7.22-7.23). The non-coercive case [3] is harder and requires, for example, a small enough coarse grid mesh size, which would not be of interest here.*
4. *We prove below that the approximation property, smoothing property and multigrid contraction number hold for the shifted Helmholtz equation. Nevertheless, $C_A$, $C_S$ and $g(\nu)$ depend on $k$ and $\varepsilon$. Therefore, the multigrid algorithm can have very slow convergence or even be divergent for some values of these parameters (see Section 6 for an illustration).*

**4.1. The approximation property.** We first prove that the approximation property holds for the shifted Helmholtz equation.

THEOREM 4.3. *There exists $C(\Omega) > 0$ such that*

$$\|A_l^{-1} - P_l A_{l-1}^{-1} R_l\|_2 \leq C(\Omega) C_{\text{reg}} C_{\text{cont}} \|A_l\|_2^{-1},$$

*where*

$$C_{\text{reg}} = \left(\frac{C_{\text{cont}}}{C_{\text{coer}}}\right)^2 C_{\text{cont}} C_{H^2}^2 \|f\|_{L^2(\Omega)},$$

*and the other constants are defined in Theorem 3.1.*

*Proof.* Let $F_l^* : \mathcal{V}_l \longrightarrow \mathbb{C}^{N_l}$ be the adjoint of $F_l$ defined in (1.4), such that $(F_l \mathbf{z}, \overline{v_l})_{L^2(\Omega)} = \langle \mathbf{z}, F_l^* \overline{v_l} \rangle = \left\langle \mathbf{z}, \overline{F_l^T v_l} \right\rangle$ holds for all $\mathbf{z} \in \mathbb{C}^{N_l}$ and $v_l \in \mathcal{V}_l$. Note also that the inverse of $F_l$ satisfies $F_l^{-1} : \mathcal{V}_l \longrightarrow \mathbb{C}^{N_l}$.

Let $\mathbf{b} \in \mathbb{C}^{N_l}$ be given. We consider the three variational problems

$$\begin{aligned}
\text{find } \varphi \in H_0^1(\Omega):\ a(\varphi,\psi) &= \left((F_l^*)^{-1}\mathbf{b}, \overline{\psi}\right)_{L^2(\Omega)}, \forall \psi \in H_0^1(\Omega),\\
\text{find } \varphi_l \in \mathcal{V}_l:\ a(\varphi_l,\psi_l) &= \left((F_l^*)^{-1}\mathbf{b}, \overline{\psi_l}\right)_{L^2(\Omega)}, \forall \psi_l \in \mathcal{V}_l,\\
\text{find } \varphi_{l-1} \in \mathcal{V}_{l-1}:\ a(\varphi_{l-1},\psi_{l-1}) &= \left((F_l^*)^{-1}\mathbf{b}, \overline{\psi_{l-1}}\right)_{L^2(\Omega)}, \forall \psi_{l-1} \in \mathcal{V}_{l-1}.
\end{aligned}$$

Note that all the variational problems above are well-posed due to Theorem 3.1. Since $\left((F_l^*)^{-1}\mathbf{b}, \overline{\psi}\right)_{L^2(\Omega)} = \left\langle \mathbf{b}, \overline{F_l^{-1}\psi} \right\rangle$, the second variational problem can be rewritten as

$$a(\varphi_l,\psi_l) = \left\langle A_l F_l^{-1}\varphi_l, \overline{F_l^{-1}\psi_l} \right\rangle = \left\langle \mathbf{b}, \overline{F_l^{-1}\psi_l} \right\rangle,$$

from which one gets $A_l F_l^{-1}\varphi_l = \mathbf{b}$ and thus $A_l^{-1}\mathbf{b} = F_l^{-1}\varphi_l$. Using a similar argument for the third variational problem yields

$$\begin{aligned}
\left((F_l^*)^{-1}\mathbf{b}, \overline{\psi_{l-1}}\right)_{L^2(\Omega)} &= a(\varphi_{l-1},\psi_{l-1})\\
&= \left\langle A_{l-1}F_{l-1}^{-1}\varphi_{l-1}, \overline{F_{l-1}^{-1}\psi_{l-1}} \right\rangle\\
&= \left((F_{l-1}^*)^{-1}A_{l-1}F_{l-1}^{-1}\varphi_{l-1}, \overline{\psi_{l-1}}\right)_{L^2(\Omega)},
\end{aligned}$$

from which we infer $F_{l-1}^{-1}\varphi_{l-1} = A_{l-1}^{-1}(F_l^{-1}F_{l-1})^*\mathbf{b} = A_{l-1}^{-1}R_l\mathbf{b}$, see definition (1.7) for the restriction $R_l$. Now multiplying by the prolongation $P_l := F_l^{-1}F_{l-1}$, see (1.6), we get $P_l A_{l-1}^{-1}R_l\mathbf{b} = P_l F_{l-1}^{-1}\varphi_{l-1} = F_l^{-1}\varphi_{l-1}$, and combining the results for both variational problems gives

$$\|A_l^{-1}\mathbf{b} - P_l A_{l-1}^{-1}R_l\mathbf{b}\|_2 = \|F_l^{-1}(\varphi_l - \varphi_{l-1})\|_2.$$

Now using Lemma 9.5 from the Appendix, we get the estimate

$$\|A_l^{-1}\mathbf{b} - P_l A_{l-1}^{-1}R_l\mathbf{b}\|_2 \le C(\Omega)h_l^{-\frac{d}{2}}\|\varphi_l - \varphi_{l-1}\|_{L^2(\Omega)}. \tag{4.1}$$

The $H^2$-regularity of $\varphi$ and standard $L^2$ error bounds for finite element methods on regular meshes [2] give

$$\|\varphi_p - \varphi\|_{L^2(\Omega)} \le C(\Omega)C_{\text{reg}}h_l^2,\ p = l, l-1,$$

where the exact value $C_{\text{reg}}$ of the constant can be found in Lemma 9.2 and reads

$$C_{\text{reg}} = \left(\frac{C_{\text{cont}}}{C_{\text{coer}}}\right)^2 C_{\text{cont}}C_{H^2}^2\|f\|_{L^2(\Omega)}.$$

A triangle inequality, the previous estimate and since we consider nested meshes satisfying $h_{l-1} \le ch_l$ then yield

$$\begin{aligned}
\|\varphi_l - \varphi_{l-1}\|_{L^2(\Omega)} &\le \|\varphi_l - \varphi\|_{L^2(\Omega)} + \|\varphi - \varphi_{l-1}\|_{L^2(\Omega)}\\
&\le C(\Omega)C_{\text{reg}}\|(F_l^*)^{-1}b\|_{L^2(\Omega)}(h_l^2 + h_{l-1}^2)\\
&\le C(\Omega)C_{\text{reg}}(1+c^2)\|(F_l^*)^{-1}b\|_{L^2(\Omega)}h_l^2.
\end{aligned}$$

Using this estimate for the right hand side of (4.1) and again Lemma 9.5 gives

$$\|A_l^{-1}\mathbf{b} - P_l A_{l-1}^{-1}R_l\mathbf{b}\|_2 \le C(\Omega)C_{\text{reg}}(1+c^2)h_l^{2-d}\|b\|_2. \tag{4.2}$$

To finish the proof, we need to estimate $\|A_l\|_2$. We recall first an inverse inequality [2] coming from the analysis of finite element methods,

$$\|\nabla\psi_l\|_{L^2(\Omega)} \leq \frac{C(\Omega)}{h_l}\|\psi_l\|_{L^2(\Omega)},\ \forall\psi\in\mathcal{V}_l.$$

From the continuity of $a$, Lemma 9.5 and this inverse inequality, one then gets

$$\begin{aligned}\|A_l\|_2 = \max_{\mathbf{z}_1,\mathbf{z}_2\in\mathbb{C}^{N_l}} \frac{|\langle A_l\mathbf{z}_1,\overline{\mathbf{z}_2}\rangle|}{\|\mathbf{z}_1\|_2\|\mathbf{z}_2\|_2} &\leq C(\Omega)h_l^d \max_{u_l,v_l\in\mathcal{V}_l}\frac{|a(u_l,v_l)|}{\|u_l\|_{L^2(\Omega)}\|v_l\|_{L^2(\Omega)}}\\ &\leq C(\Omega)C_{\mathrm{cont}}h_l^d\max_{u_l,v_l\in\mathcal{V}_l}\frac{\|\nabla u_l\|_{L^2(\Omega)}\|\nabla v_l\|_{L^2(\Omega)}}{\|u_l\|_{L^2(\Omega)}\|v_l\|_{L^2(\Omega)}}\\ &\leq C(\Omega)C_{\mathrm{cont}}h_l^{d-2}.\end{aligned}$$

We can thus replace the dependence on $h_l^{2-d}$ by a dependence on $\|A_l\|_2$ in (4.2) and obtain

$$\|A_l^{-1}\mathbf{b} - P_lA_{l-1}^{-1}R_l\mathbf{b}\|_2 \leq C(\Omega)C_{\mathrm{reg}}C_{\mathrm{cont}}\|A_l\|_2^{-1}\|\mathbf{b}\|_2,$$

which concludes the proof, since our argument holds for any $\mathbf{b}\in\mathbb{C}^{N_l}$. ∎

**4.2. The smoothing property.** We now show that the smoothing property holds for the shifted Helmholtz equation with a small modification in the Jacobi smoother.

THEOREM 4.4. *There exists $C(\Omega)>0$ such that for all*

$$\omega\in(0,\frac{2C(\Omega)C_{\mathrm{coer}}^2}{C_{\mathrm{cont}}^2}), \tag{4.3}$$

*one has with the modification of putting a modulus on $D=\mathrm{diag}(A_l)$ in the Jacobi smoother*

$$\|I-\omega\gamma|D|^{-1}A_l\|_D \leq 1, \tag{4.4}$$

$$\left\|A_l\left(I-\frac{\omega}{2}\gamma|D|^{-1}A_l\right)^{\nu}\right\|_2 \leq 2C(\Omega)\sqrt{\frac{2}{\pi\omega^2\nu}}\sqrt{\frac{C_{\mathrm{cont}}}{C_{\mathrm{coer}}}}\|A_l\|_2, \tag{4.5}$$

*where $I$ is the identity operator, $\gamma$, $C_{\mathrm{cont}}$, and $C_{\mathrm{coer}}$ are defined in Theorem 3.1, and $\forall\mathbf{z}\in\mathbb{C}^{N_l}$, $\|\mathbf{z}\|_D := \|D^{\frac{1}{2}}\mathbf{z}\|_2$, whose induced matrix norm is $\|B\|_D = \|D^{\frac{1}{2}}BD^{-\frac{1}{2}}\|_2$.*

*Proof.* Let $\mathbf{z}\in\mathbb{C}^{N_l}$. We begin to estimate

$$\begin{aligned}\|\overline{D}^{-\frac{1}{2}}A_lD^{-\frac{1}{2}}\mathbf{z}\|_2 &= \max_{\mathbf{z}_1\in\mathbb{C}^{N_l}}\frac{\left|\left\langle\overline{D}^{-\frac{1}{2}}A_lD^{-\frac{1}{2}}\mathbf{z},\overline{\mathbf{z}_1}\right\rangle\right|}{\|\mathbf{z}_1\|_2}\\ &= \max_{\mathbf{z}_1\in\mathbb{C}^{N_l}}\frac{\left|\left\langle A_lD^{-\frac{1}{2}}\mathbf{z},\overline{D^{-\frac{1}{2}}\mathbf{z}_1}\right\rangle\right|}{\|\mathbf{z}_1\|_2}\\ &= \max_{\mathbf{z}_1\in\mathbb{C}^{N_l}}\frac{\left|a\left(F_lD^{-\frac{1}{2}}\mathbf{z},F_lD^{-\frac{1}{2}}\mathbf{z}_1\right)\right|}{\|\mathbf{z}_1\|_2}.\end{aligned}$$

Now using the continuity of the bilinear form together with the inverse inequality and finally Lemma 9.5, one gets

$$\begin{aligned}\|\overline{D}^{-\frac{1}{2}}A_lD^{-\frac{1}{2}}\mathbf{z}\|_2 &\leq C_{\mathrm{cont}}\|\nabla F_lD^{-\frac{1}{2}}\mathbf{z}\|_{L^2(\Omega)}\frac{C(\Omega)}{h_l}\max_{\mathbf{z}_1\in\mathbb{C}^{N_l}}\frac{\|F_lD^{-\frac{1}{2}}\mathbf{z}_1\|_{L^2(\Omega)}}{\|\mathbf{z}_1\|_2}\\ &\leq C(\Omega)C_{\mathrm{cont}}h_l^{\frac{d}{2}-1}\|\nabla F_lD^{-\frac{1}{2}}\mathbf{z}\|_{L^2(\Omega)}\|D^{-\frac{1}{2}}\|_2.\end{aligned}$$

The coercivity (3.1) of the bilinear form then yields

$$\|\overline{D}^{-\frac{1}{2}} A_l D^{-\frac{1}{2}} \mathbf{z}\|_2 \leq C(\Omega) C_{\text{cont}} h_l^{\frac{d}{2}-1} \|D^{-\frac{1}{2}}\|_2 \sqrt{\frac{1}{C_{\text{coer}}}} \sqrt{\mathcal{R}\gamma a(F_l D^{-\frac{1}{2}}\mathbf{z}, F_l D^{-\frac{1}{2}}\mathbf{z})}.$$

We now derive an upper bound for $h_l^{\frac{d}{2}-1}\|D^{-\frac{1}{2}}\|_2$. Since this matrix is diagonal, its 2-norm is nothing but its largest value. From the coercivity of $a$ and the fact that $|\gamma| = 1$, one can infer that

$$|D_{jj}^{\frac{1}{2}}| = |(\gamma D_{jj})^{\frac{1}{2}}| = |(\gamma a(\phi_j, \phi_j))^{\frac{1}{2}}| \geq \sqrt{C_{\text{coer}}} \|\nabla \phi_j\|_{L^2(\Omega)} \geq C(\Omega)\sqrt{C_{\text{coer}}} h_l^{\frac{d}{2}-1},$$

where the last inequality can be shown by using for $T \subset \text{supp}(\phi_j)$ the affine transformation from the unit simplex to $T$. We then arrive at

$$h_l^{\frac{d}{2}-1}\|D^{-\frac{1}{2}}\|_2 \leq \frac{1}{C(\Omega)\sqrt{C_{\text{coer}}}}, \tag{4.6}$$

and thus we have the bound

$$\begin{aligned}\|\gamma \overline{D}^{-\frac{1}{2}} A_l D^{-\frac{1}{2}} \mathbf{z}\| &\leq C(\Omega)\frac{C_{\text{cont}}}{C_{\text{coer}}} \sqrt{\mathcal{R}\gamma a(F_l D^{-\frac{1}{2}}\mathbf{z}, F_l D^{-\frac{1}{2}}\mathbf{z})} \\ &= C(\Omega)\frac{C_{\text{cont}}}{C_{\text{coer}}} \sqrt{\mathcal{R}\left\langle \gamma \overline{D}^{-\frac{1}{2}} A_l D^{-\frac{1}{2}}\mathbf{z}, \overline{\mathbf{z}}\right\rangle}.\end{aligned}$$

Now we apply Lemma 9.3 which states that for all $\omega \in (0, 2C(\Omega)C_{\text{coer}}^2/C_{\text{cont}}^2)$, we have the estimate

$$\|I - \omega\gamma \overline{D}^{-\frac{1}{2}} A_l D^{-\frac{1}{2}}\|_2 \leq 1. \tag{4.7}$$

Using that $|D|^{-1} = D^{-1/2}\overline{D}^{-1/2}$, inequality (4.7) proves the first result (4.4) of the theorem, because

$$\|I - \gamma\omega |D|^{-1} A_l\|_D = \|I - \omega\gamma \overline{D}^{-\frac{1}{2}} A_l D^{-\frac{1}{2}}\|_2 \leq 1. \tag{4.8}$$

Applying Corollary 9.4 with $M_l = (\omega\gamma)^{-1}|D|$ and using that $|\gamma| = 1$ gives

$$\left\| A_l \left( I - \frac{1}{2} M_l^{-1} A_l \right)^{\nu} \right\|_D \leq 2\sqrt{\frac{2}{\pi\nu}} \|M_l\|_D = 2\sqrt{\frac{2}{\pi\nu}} \frac{\|D\|_2}{\omega} \leq 2\sqrt{\frac{2}{\pi\nu\omega^2}} \|A_l\|_2, \tag{4.9}$$

where the last inequality holds, because ($\mathbf{e}_j$ denotes the canonical basis of $\mathbb{R}^{N_l}$)

$$\|D\|_2 = \max_{j=1,\cdots.N_l} |a(\phi_j, \phi_j)| = \max_{j=1,\cdots.N_l} |\langle A_l \mathbf{e}_j, \overline{\mathbf{e}}_j\rangle| \leq \max_{\mathbf{z}_1,\mathbf{z}_2 \in \mathbb{C}^{N_l}} \frac{|\langle A_l \mathbf{z}_1, \overline{\mathbf{z}_2}\rangle|}{\|\mathbf{z}_1\|_2 \|\mathbf{z}_2\|_2} = \|A_l\|_2.$$

To relate the $D$-norm in (4.9) to a 2-norm, we use the definition of $\|.\|_D$ and obtain

$$\begin{aligned}\left\| A_l \left( I - \frac{\omega}{2}\gamma |D|^{-1} A \right)^{\nu} \right\|_2 &= \left\| D^{-\frac{1}{2}} D^{\frac{1}{2}} A_l \left( I - \frac{\omega}{2}\gamma |D|^{-1} A \right)^{\nu} D^{-\frac{1}{2}} D^{\frac{1}{2}} \right\|_2 \\ &\leq \left\| A_l \left( I - \frac{\omega}{2}\gamma |D|^{-1} A \right)^{\nu} \right\|_D \|D^{-\frac{1}{2}}\|_2 \|D^{\frac{1}{2}}\|_2. \end{aligned} \tag{4.10}$$

Following a similar argument as for (4.6), one can prove that

$$\|D^{\frac{1}{2}}\|_2 \leq C(\Omega)\sqrt{C_{\text{cont}}} h_l^{\frac{d}{2}-1}. \tag{4.11}$$

Using now estimates (4.11) and (4.6) in (4.10), we obtain from (4.9) the desired estimate (4.5). ∎

**4.3. Multigrid contraction number.** We now give the multigrid contraction number with the modified smoother for the shifted Helmholtz equation (1.1).

THEOREM 4.5. *For all $\nu \geq 1$, one has*

$$\|S^\nu\|_2 \leq C(\Omega)\sqrt{\frac{C_{\text{cont}}}{C_{\text{coer}}}}, \tag{4.12}$$

*where $S = I - \frac{\omega}{2}\gamma|D|^{-1}A_l$ and $C(\Omega)$ depends only on $\Omega$.*

*Proof.* Using (4.8), we can estimate

$$\|S\|_D \leq \frac{1}{2}\|I\|_D + \frac{1}{2}\|I - \gamma\omega|D|^{-1}A_l\|_D \leq 1. \tag{4.13}$$

Now we use the definition of the $D$-norm to get

$$\|S^\nu\|_2 \leq \|S^\nu\|_D\|D^{\frac{1}{2}}\|_2\|D^{-\frac{1}{2}}\|_2 \leq C(\Omega)\sqrt{\frac{C_{\text{cont}}}{C_{\text{coer}}}},$$

where we used (4.11), (4.6) and (4.13) for the last inequality. ∎

**4.4. The convergence result.** Collecting results from Theorems 4.1, 4.7, 4.4, 4.3 and 4.5, we have finally proved the following convergence result for the multigrid method with modified Jacobi smoother applied to the shifted Helmholtz equation (1.1) for any shift $\varepsilon > 0$.

THEOREM 4.6. *If all the assumptions made in this section hold, then the multigrid algorithm (1.8) applied to the shifted Helmholtz equation (1.1) with modified Jacobi smoother $S$ defined in Theorem 4.5 converges, and for all $\xi \in (0,1)$, there exists $\nu_\xi$ such that $\|C_{MG,l}\|_2 \leq \xi$ for all $\nu \geq \nu_\xi$.*

Theorem 4.6 shows that the multigrid method converges for any $\varepsilon > 0$ if one does a large enough number of smoothing steps. The number $\nu_{\min}$ required to reach the threshold $\|C_{MG,l}\|_2 < 1$ from which the multigrid algorithm actually converges depends on the wave number and on $\varepsilon$. As a result, it can happen that an extremely large number of smoothing iterations is needed in order for multigrid to be convergent for solving the discrete problem (1.3). Some explicit values for $\nu_{\min}$ are given in Section 6.

**4.5. Discussion on the choice of the damping parameter.** The results we proved so far are valid for all $\varepsilon > 0$, and it thus seems that one can use multigrid with the modified Jacobi smoother successfully to solve the shifted Helmholtz problem (1.1) for any $\varepsilon > 0$ in an efficient manner. This is however not the case, since the choice of the damping parameter $\omega$ depends on $k$ and $\varepsilon$ and must satisfy

$$\omega \in \left(0, \frac{2C(\Omega)C_{\text{coer}}^2}{C_{\text{cont}}^2}\right),$$

where $C(\Omega)$ only depends on the geometry of $\Omega$. Using the explicit values of the continuity and coercivity constants of the bilinear form $a$, see Theorem 3.1, we get for the damping parameter the upper bound

$$0 < \omega < \omega_{\text{sup}}$$

with

$$\omega_{\text{sup}} = \begin{cases} 2C(\Omega)\left(\dfrac{\min\{\mathcal{R}\gamma, \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma\}}{1 + C_P|k^2 + \mathrm{i}\varepsilon|}\right)^2 & \text{if } \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma > 0, \\ 2C(\Omega)\left(\dfrac{\mathcal{R}\gamma}{1 + C_P|k^2 + \mathrm{i}\varepsilon|}\right)^2 & \text{otherwise,} \end{cases} \tag{4.14}$$

where $\gamma$ is a complex number such that $|\gamma| = 1$, $\mathcal{I}\gamma > 0$ and $0 < \mathcal{R}\gamma \leq \varepsilon/|k^2 + \mathrm{i}\varepsilon|$.

If $\omega$ is becoming too small, the multigrid algorithm (1.8) will have very poor convergence properties, and the smoothing property given by the estimates of Theorem 4.4 could even become numerically false, which we show now by carefully studying the upper bound $\omega_{\text{sup}}$ for $\omega$. First we fix $\varepsilon > 0$ and study the values $\omega$ can take. Assuming first that $\varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma = 0$, we see that $\gamma = \gamma_0$ where

$$\gamma_0 = \frac{\varepsilon + \mathrm{i}k^2}{|\varepsilon + \mathrm{i}k^2|}.$$

As a result, the upper bound for the damping parameter (4.14) becomes

$$\omega_{\text{sup}} = 2C(\Omega)\left(\frac{\varepsilon}{|k^2 + \mathrm{i}\varepsilon|(1 + C_P|k^2 + \mathrm{i}\varepsilon|)}\right)^2, \tag{4.15}$$

and a direct computation gives the estimate

$$\omega_{\text{sup}} = O\left(\frac{\varepsilon^2}{(k^4 + \varepsilon^2)^2}\right).$$

The above formula goes to zero when $\varepsilon$ goes to zero and/or $k$ grows to infinity. Therefore, for all $\varepsilon > 0$, there exists $k_{0,\varepsilon}$ such that for all $k > k_{0,\varepsilon}$, one has

$$0 < \omega < \omega_{\text{sup}} < \texttt{macheps},$$

where `macheps` is the machine precision. For such wave numbers, Theorem 4.4 does not hold any more numerically, since

$$\forall \varepsilon > 0 \ \forall k > k_{0,\varepsilon}, \ \|A_l\left(I - \frac{\omega}{2}\gamma_0|D|^{-1}A\right)^{\nu}\|_2 = \|A_l\|_2, \ \nu \geq 1.$$

We illustrate this numerically in Section 6.1.

Now assume that $\varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma > 0$ and that $\mathcal{R}\gamma \geq \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma$. A direct calculation shows that $\gamma$ then satisfy the two sided inequality

$$\frac{\varepsilon}{|(k^2 + 1) + \mathrm{i}\varepsilon|} \leq \mathcal{R}\gamma < \frac{\varepsilon}{|k^2 + \mathrm{i}\varepsilon|}, \qquad \mathcal{I}\gamma = \sqrt{1 - \mathcal{R}\gamma^2}. \tag{4.16}$$

From (4.16), we see that for all $\varepsilon > 0$, $\lim_{k\to+\infty} \mathcal{R}\gamma = 0$ and $\lim_{k\to+\infty} \mathcal{I}\gamma = 1$, and as a consequence, for all $\varepsilon > 0$, there exists $k_{0,\varepsilon}$ such that for all $k > k_{0,\varepsilon}$, one has again

$$\omega_{\text{sup}} = 2C(\Omega)\left(\frac{\varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma}{1 + C_P|k^2 + \mathrm{i}\varepsilon|}\right)^2 < \texttt{macheps},$$

and hence Theorem 4.4 does not hold any more numerically. For a given $\varepsilon > 0$, it is therefore not possible to have a convergent multigrid method (1.8) for all wave

numbers $k$, because this would involve a damping parameter which is too small, and even numerically zero. A way to proceed is to restrict $\varepsilon$ to get a coercivity constant that behaves like $O(k^2)$. With such a behavior, the ratio $C_{\text{coer}}/C_{\text{cont}}$ is going to remain bounded away from zero for large wave numbers.

We therefore now assume that $\gamma$ does not depend on these parameters, and use the upper bound

$$\omega_{\text{sup}} = 2C(\Omega)\left(\frac{\varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma}{1 + C_P|k^2 + \mathrm{i}\varepsilon|}\right)^2,$$

which meets all the desired requirements on $\omega$ since this upper bound does not go to zero. The next result gives an estimate on $\varepsilon$ to have this desired property.

THEOREM 4.7. *Let $\gamma \in \mathbb{C}$ satisfying all the constraint for which the coercivity inequality (3.1) holds. Then for all $\varepsilon$ satisfying*

$$\sqrt{\frac{(\mathcal{R}\gamma)^2}{1-(\mathcal{R}\gamma)^2}}k^2 < \varepsilon \leq \sqrt{\frac{(\mathcal{R}\gamma)^2}{1-(\mathcal{R}\gamma)^2}}(k^2+1),$$

*the upper bound for the damping parameter of the Jacobi smoother given by formula (4.14) is bounded from above and away from zero for all wave numbers.*

*Proof.* According to the definition of $C_{\text{coer}}$ in Theorem 3.1 and $\omega_{\text{sup}}$ in (4.14), we only need to check that $\varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma > 0$ and that $\mathcal{R}\gamma \geq \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma$. Since $|\gamma| = 1$ with $\mathcal{I}\gamma > 0$ and $\mathcal{R}\gamma > 0$, the first statement is equivalent to

$$\varepsilon > k^2\frac{\mathcal{R}\gamma}{\mathcal{I}\gamma}. \tag{4.17}$$

Using again that $|\gamma| = 1$ with $\mathcal{I}\gamma > 0$, the condition $\mathcal{R}\gamma \geq \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma$ now reads

$$(1+k^2)\mathcal{R}\gamma \geq \varepsilon\mathcal{I}\gamma = \varepsilon\sqrt{1-(\mathcal{R}\gamma)^2},$$

and gives the upper bound for $\varepsilon$. ∎

According to Theorem 4.7, one can find $\varepsilon = O(k^2)$ and $\gamma$ that does neither depend on $k$ nor on $\varepsilon$ such that the coercivity inequality (3.1) holds and the damping parameter of the Jacobi smoother does not go to zero for any wave number.

REMARK 4.8. *The upper bound for $\varepsilon$ given in Theorem 4.7 is artificial since for large wave numbers, one always has for all $\gamma$ satisfying $\mathcal{R}\gamma > 0$ and $\mathcal{I}\gamma > 0$ that $\mathcal{R}\gamma \geq \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma$.*

Note that the function

$$y \in (0,1) \longmapsto \sqrt{\frac{y^2}{1-y^2}} \in \mathbb{R}^+$$

is bijective. Therefore, from Theorem 4.7, one can get a convergent multigrid method for $\varepsilon = \mathrm{c}k^2$ with a constant $\mathrm{c} > 0$, independent of $k$ and $\varepsilon$, being as small as we want. The main practical interest of this property is that one can apply multigrid with a shift smaller than the one used in most applications (which is $\mathrm{c} \geq 0.5$, see for example [8], and also results from Section 2 where $\mathrm{c} = 0.8$). This comes however again at a cost of a large number of smoothing steps for this variant of the method.

REMARK 4.9. *A similar upper bound for $\omega$ can be obtained if one follows the proof of Theorem 4.4 using the $k$-dependent $H^1$-norm and the coercivity estimate from Lemma* 3.1 *in [17]. This requires to set $\gamma = \mathrm{i}\overline{\sqrt{k^2+\mathrm{i}\varepsilon}}/|\sqrt{k^2+\mathrm{i}\varepsilon}|$ and yields*

$$\omega \in \left(0, 2C(\Omega)\left(\frac{\varepsilon}{k^2}\right)^2\right).$$

*This bound is simpler, but the behavior of the multigrid algorithm remains the same, since its convergence still needs $\varepsilon = O(k^2)$.*

**5. Extension to the case of impedance boundary conditions.** The Helmholtz problem with impedance boundary conditions is

$$\begin{cases} -\Delta u(x) - (k^2+\mathrm{i}\varepsilon)u(x) &= f(x),\ x\in\Omega, \\ \partial_{\mathbf{n}} u - \mathrm{i}\eta u &= 0,\ \text{on } \partial\Omega, \end{cases} \tag{5.1}$$

where $\mathbf{n}$ is the unit outward normal vector to $\partial\Omega$, and $\eta > 0$ is the impedance parameter. Denoting by $d\sigma$ the measure associated with $\partial\Omega$, we obtain (5.1) in variational form,

$$\begin{cases} \text{Find } u\in H^1(\Omega) \text{ such that for all } v\in H^1(\Omega): \\ a_\eta(u,v) := \displaystyle\int_\Omega \nabla u\cdot\overline{\nabla v} - (k^2+\mathrm{i}\varepsilon)u\overline{v}dx - \mathrm{i}\eta\int_{\partial\Omega} u\overline{v}d\sigma = \int_\Omega f\overline{v}dx. \end{cases} \tag{5.2}$$

Let $\mathcal{V}_l$ be the finite element space obtained with piece-wise linear polynomials

$$\mathcal{V}_l = \left\{v\in\mathcal{C}(\overline{\Omega}) \mid v|_T\in\mathbb{P}_1 \text{ for all } T\in\mathcal{T}_l\right\}.$$

The discrete problem we then obtain is

$$\begin{cases} \text{Find } u_l\in\mathcal{V}_l \text{ such that}: \\ a_\eta(u_l,v_l) = \displaystyle\int_\Omega f\overline{v_l}dx,\ \forall v_l\in\mathcal{V}_l. \end{cases} \tag{5.3}$$

This is equivalent to the linear system $A_l\mathbf{z}_l = \mathbf{b}_l$ where $u_l = F_l\mathbf{z}_l$ is the Galerkin solution. Algorithm (1.8) can be applied to this linear system. Apart from the fact that we use the $H^1$ semi-norm in the proof of Theorem 4.6, the Dirichlet boundary condition does not affect our analysis, since we only used the continuity and the coercivity of the bilinear form $a$ associated with problem (4.6). Therefore, the analysis of multigrid methods for (5.3) easily follows from our previous analysis, simply replacing $\|\nabla\phi\|_{L^2(\Omega)}$ by $\|\phi\|_{H^1(\Omega)}$ and using the inverse inequality

$$\|\psi_l\|_{H^1(\Omega)} \le \frac{C(\Omega)}{h_l}\|\psi_l\|_{L^2(\Omega)},\ \forall\psi_l\in\mathcal{V}_l,$$

which holds since $h_l\le 1$. We also have to replace the correct value for the coercivity, continuity and $H^2$-regularity constants of the Helmholtz impedance problem. The latter are given below.

THEOREM 5.1. *For all $u,v\in H^1(\Omega)$, we have the properties:*

1. *The bilinear form $a_\eta$ is continuous*

$$|a_\eta(u,v)| \le (1+|k^2+\mathrm{i}\varepsilon|+\eta C_t)\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)} = C_{\mathrm{cont},\eta}\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)},$$

*where $C_t$ is any constant such that $\|u\|_{L^2(\partial\Omega)}\le C_t\|u\|_{H^1(\Omega)}$.*

2. *For any* $\gamma \in \mathbb{C}$ *satisfying* $|\gamma| = 1$, $\mathcal{I}\gamma > 0$ *and* $0 < \mathcal{R}\gamma < \varepsilon/|k^2 + \mathrm{i}\varepsilon|$ *one has*

$$\mathcal{R}\gamma a_\eta(u,u) \geq \min\{\mathcal{R}\gamma, \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma, \eta\mathcal{I}\gamma\}\|u\|^2_{H^1(\Omega)} = C_{\mathrm{coer},\eta}\|u\|^2_{H^1(\Omega)}.$$

3. *The bilinear form* $a_\eta$ *is coercive*

$$|a_\eta(u,v)| \geq C_{\mathrm{coer},\eta}\|u\|^2_{H^1(\Omega)}.$$

4. *There exists a strictly positive constant* $C(\Omega)$ *that depends only on* $\Omega$*, such that for all* $f \in L^2(\Omega)$*, the solution to (5.1) satisfies*

$$\|u\|_{H^2(\Omega)} \leq C(\Omega)\|f\|_{L^2(\Omega)}\left(1 + \frac{|k^2 + \mathrm{i}\varepsilon|}{\varepsilon} + \frac{1 + |\eta|}{C_{\mathrm{coer},\eta}}\right) = C_{H^2,\eta}\|f\|_{L^2(\Omega)}.$$

*Proof.* The proof is very similar to the one of Theorem 3.1. The first item is easy. For the second point, we take a complex number $\gamma = \alpha + \mathrm{i}\beta$ such that $|\gamma| = 1$. For any $\varphi \in H^1(\Omega)$, a direct computation gives

$$\begin{aligned}\gamma a_\eta(\varphi,\varphi) = \gamma\|\nabla\varphi\|^2_{L^2(\Omega)} + (\beta\varepsilon - \alpha k^2 + \mathrm{i}(\beta k^2 + \alpha\varepsilon))\|\varphi\|^2_{L^2(\Omega)} \\ +(\beta\eta - \mathrm{i}\eta\alpha)\|\varphi\|^2_{L^2(\partial\Omega)}.\end{aligned}$$

Taking the real part then yields

$$\mathcal{R}\gamma a_\eta(\varphi,\varphi) = \alpha\|\nabla\varphi\|^2_{L^2(\Omega)} + (\beta\varepsilon - \alpha k^2)\|\varphi\|^2_{L^2(\Omega)} + \beta\eta\|\varphi\|^2_{L^2(\partial\Omega)}.$$

Now choosing $\gamma$ as above leads to

$$\mathcal{R}\gamma a_\eta(\varphi,\varphi) \geq C_{\mathrm{coer},\eta}(\|\varphi\|^2_{H^1(\Omega)} + \|\varphi\|^2_{L^2(\partial\Omega)}),$$

which proves the coercivity estimate.

The third item now follows from $|\gamma| = 1$, $|a_\eta(\varphi,\varphi)| \geq |\mathcal{R}\gamma a_\eta(\varphi,\varphi)|$ and the coercivity of the bilinear form.

For the elliptic regularity estimate, one can use inequality (2.10) page 12 from [17]. The latter holds for convex polygons and states that

$$\|v\|_{H^2(\Omega)} \leq C(\Omega)\left(\|\Delta v\|_{L^2(\Omega)} + \|v\|_{H^1(\Omega)} + \|\partial_{\mathbf{n}} v\|_{H^{1/2}(\partial\Omega)}\right), \tag{5.4}$$

for any $v \in H^1(\Omega)$ such that $\Delta v \in L^2(\Omega)$ with $\partial_{\mathbf{n}} v \in H^{1/2}(\partial\Omega)$. Note that estimate (5.4) can be applied with the solution to (5.1) using trace regularity together with the impedance boundary condition and that $f \in L^2(\Omega)$. One gets

$$\|u\|_{H^2(\Omega)} \leq C(\Omega)\left(\|f - (k^2 + \mathrm{i}\varepsilon)u\|_{L^2(\Omega)} + \|u\|_{H^1(\Omega)} + |\eta|\|u\|_{H^1(\Omega)}\right).$$

The bound for the $H^1$ norm comes from the coercivity estimate. The $L^2$ norm of $u$ can be bounded as follows:

$$\mathcal{I}a_\eta(u,u) = -\varepsilon\|u\|^2_{L^2(\Omega)} - \eta\|u\|^2_{L^2(\partial\Omega)} = \mathcal{I}\int_\Omega f\overline{u}dx,$$

and the Cauchy-Schwarz inequality then yields

$$\|u\|_{L^2(\Omega)} \leq \frac{\|f\|_{L^2(\Omega)}}{\varepsilon}.$$

Collecting the previous estimates then gives the desired result. ∎

Note that the $\gamma$ defined above depends on both the wave number and the shift. In addition, for a suitable choice of $\gamma$, one can show that the $(k, \varepsilon)$-dependence of the coercivity constant of $a_\eta$ is optimal hence behaving like $\varepsilon/k^2$ (see Remark 3.2).

From the above observations, Theorems 4.4, 4.3 and 4.5 hold for problem (5.3) with $C_{str,\eta}$ replacing $C_{\text{str}}$ for $\text{str} \in \{\text{Cont}, \text{Coer}\}$. The upper bound for the damping parameter of the Jacobi smoother is now given by

$$\omega_{\text{sup}} = 2C(\Omega)\left(\frac{C_{\text{coer},\eta}}{C_{\text{cont},\eta}}\right)^2 = 2C(\Omega)\left(\frac{\min\{\mathcal{R}\gamma, \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma, \eta\mathcal{I}\gamma\}}{(1 + |k^2 + \mathrm{i}\varepsilon| + \eta C_t)}\right)^2.$$

A theorem similar to Theorem 4.6 thus holds for the impedance Helmholtz problem (5.1).

The results of Section 4.5 also apply here: in order to have a multigrid method that converges for all $k$, one needs to restrict $\varepsilon$ and $\eta$ such that $\omega_{\text{sup}}$ does not tend to zero when $k$ goes to infinity. This requires $\gamma$ not to depend on $k$ and $\varepsilon$, and $\eta \leq Ck^2$ for a positive constant $C$. A theorem similar to Theorem 4.7 thus holds for problem (5.1):

THEOREM 5.2. *Let $\gamma \in \mathbb{C}$ be defined as in Theorem 5.1. Then for all $\varepsilon$ and $\eta$ satisfying*

$$0 \leq \eta \leq Ck^2, \ \sqrt{\frac{(\mathcal{R}\gamma)^2}{1-(\mathcal{R}\gamma)^2}}k^2 \leq \varepsilon \leq \sqrt{\frac{(\mathcal{R}\gamma)^2}{1-(\mathcal{R}\gamma)^2}}(k^2+1),$$

*the upper bound for the damping parameter of the Jacobi smoother used in the multigrid algorithm for the impedance Helmholtz problem is bounded from above and away from zero for all wave numbers.*

Theorem 5.2 gives bounds on the shift and on the impedance parameter to have a robust multigrid algorithm for all wave numbers.

REMARK 5.3. *The typical case $\eta = k$ of an approximate radiation boundary condition can be treated by the multigrid method for all $k$, provided the shift $\varepsilon$ is large enough, see Theorem 5.2.*

REMARK 5.4. *From the proof of Theorem 5.1, one can see that the coercivity constant for Neumann boundary conditions (that is $\eta = 0$) is given by*

$$C_{\text{coer},0} = \min\{\mathcal{R}\gamma, \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma\}.$$

*The upper bound for the damping parameter of the modified Jacobi smoother is now given by*

$$\omega_{\text{sup}} = 2C(\Omega)\left(\frac{C_{\text{coer},0}}{C_{\text{cont},0}}\right)^2 = 2C(\Omega)\left(\frac{\min\{\mathcal{R}\gamma, \varepsilon\mathcal{I}\gamma - k^2\mathcal{R}\gamma\}}{(1 + |k^2 + \mathrm{i}\varepsilon|)}\right)^2.$$

**6. Numerical experiments.** We present now some numerical illustrations of Theorems 4.4, 4.3, 4.5 in a one and two dimensional setting on $\Omega = (0,1)^d$. We apply the multigrid method (1.8) with the Jacobi-like smoother given in Theorem 4.4. A comparison between the $1D$ Fourier analysis and the general analysis is also carried out. We finish this section with two dimensional numerical experiments.

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $\varepsilon = 1$ | $9.8145.10^{-9}$ | $3.9101.10^{-11}$ | $1.5622.10^{-13}$ | $6.6265.10^{-16}$ | $4.2251.10^{-18}$ |
| $\varepsilon = k$ | $9.6240.10^{-7}$ | $1.5563.10^{-8}$ | $2.4964.10^{-10}$ | $4.2396.10^{-12}$ | $1.0815.10^{-13}$ |
| $\varepsilon = k^{3/2}$ | $8.1203.10^{-6}$ | $2.8376.10^{-7}$ | $9.5163.10^{-9}$ | $3.3091.10^{-10}$ | $1.7030.10^{-11}$ |

TABLE 6.1
*Spectral radius of* $\omega\gamma \left|D\right|^{-1} A_l$.

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $\varepsilon = 1$ | 64.9 | 69.4 | 277.8 | 2010.6 | 557.2 |
| $\varepsilon = k$ | 35.6033 | 56.9249 | 237.8041 | 810.4252 | 352.7963 |
| $\varepsilon = k^{3/2}$ | 6.2924 | 15.1921 | 37.2161 | 61.7139 | 34.3143 |

TABLE 6.2
$\|A_l^{-1} - P_l A_{l-1}^{-1} R_l\|_2 \|A_l\|$ *as a function of* $k$.

**6.1. One dimensional numerical results.** We use $N_l = 2^7 - 1 = 127$ points for the fine mesh of $[0, 1]$ and $N_{l-1} = 2^4 - 1 = 63$ points for the coarse mesh. The wave number varies from $k = 10$ to $k = 160$. As a result, one has

$$kh = 0.0781,\ 0.1563,\ 0.3125,\ 0.6250,\ 1.2500.$$

Note that the bilinear form $a$ is coercive for all wave numbers such that $k < \frac{1}{C_P}$, where $C_P$ is the constant in the Poincaré inequality. Below, we use $C_P = 1$ for the Poincaré constant and set $C(\Omega) = 1$.

**The case** $\varepsilon < O(k^2)$**.** This case requires $\gamma$ to depend on $k$ and $\varepsilon$. According to Theorem 3.1 and (4.14), we use the parameters

$$\gamma = \frac{\varepsilon + \mathrm{i}k^2}{|\varepsilon + \mathrm{i}k^2|},\ \omega = \left(\frac{\mathcal{R}\gamma}{1 + |k^2 + \mathrm{i}\varepsilon|}\right)^2,\ \varepsilon = 1,\ k,\ k^{3/2}.$$

We show in Table 6.1 the spectral radius of $\omega\gamma \left|D\right|^{-1} A_l$, which indicates that the smoothing property fails numerically for large wave numbers, because one obtains numerically $S = I - \frac{\omega}{2}\gamma|D|^{-1}A_l = I$. This behavior can be explained: because $\gamma = O(1)$ and the damping parameter behaves like $\omega = O(k^{-4})$, the eigenvalues of $\omega\gamma \left|D\right|^{-1} A_l$ are, for growing $k$, close to or smaller than the machine precision, which in double precision is `macheps`$= 2.2204e-16$. Thus increasing the number of smoothing steps does not yield numerically the expected smoothing property any more. Table 6.2 shows that the approximation property holds in this case with a constant that decreases as the shift increases. Table 6.3 gives the values of the spectral radius $\rho(T)$ of the two-grid operator for $\nu = 1, 3$. We see that $\rho(T)$ becomes numerically equal to 1 for growing $k$. This is because the two-grid operator reduces to the coarse-grid operator since the smoother is close to $S = I$ in finite precision arithmetic. As the coarse grid operator is a projection, one thus obtains that the spectral radius of $T$ is close to 1. We observe nevertheless that the spectral radius of the two-grid operator decreases as $\nu$ increases, so a large number of smoothing steps could still yield a nice contraction factor in the cases considered. Note however also that for $k \in \{40, 80, 160\}$ and $\varepsilon = 1$, and $k = 160$ and $\varepsilon = k$ the contraction factor numerically is bigger than one for small $\nu$ and the algorithm diverges.

| $\nu$ | $\varepsilon$ | $k=10$ | $k=20$ | $k=40$ | $k=80$ | $k=160$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 0.999999999950900 | 0.999999999999955 | 1.000000000000005 | 1.000000000000018 | 1.000000000000007 |
| 3 | 1 | 0.999999999852697 | 0.999999999999854 | 1.000000000000004 | 1.000000000000022 | 1.000000000000006 |
| 1 | $k$ | 0.999999952087386 | 0.999999999610567 | 0.999999999996854 | 0.999999999999975 | 1.000000000000001 |
| 3 | $k$ | 0.999999856262145 | 0.999999998831708 | 0.999999999990563 | 0.999999999999920 | 1.000000000000002 |
| 1 | $k^{3/2}$ | 0.999998775023979 | 0.999999968971335 | 0.999999999250695 | 0.999999999980993 | 0.999999999999213 |
| 3 | $k^{3/2}$ | 0.999996325076436 | 0.999999906914004 | 0.999999997752078 | 0.999999999942982 | 0.999999999997680 |

TABLE 6.3
*Spectral radius $\rho(T)$ of the two-grid operator for $\nu = 1, 3$.*

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $\nu=1$ | 0.9996 | 0.9993 | 0.9990 | 0.9987 | 0.9984 |
| $\nu=3$ | 0.9952 | 0.9906 | 0.9860 | 0.9814 | 0.9768 |
| $\nu=5$ | 0.9349 | 0.8741 | 0.8173 | 0.7642 | 0.7145 |

TABLE 6.4
*Multigrid contraction number $\|S^\nu\|_2$ for $\varepsilon = O(k^2)$.*

**The case $\varepsilon = \boldsymbol{O}(k^2)$.** Theorem 4.7 shows that we need a $\gamma$ that does neither depend on $k$ nor on $\varepsilon$ for robust smoothing, and thus the shift has to satisfy specific bounds. We use the parameter choice

$$\gamma = \frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2},\ \varepsilon = k^2(\frac{\sqrt{3}}{3} + 1) \sim 1.5k^2. \tag{6.1}$$

According to Theorem 4.7, the damping parameter for the smoother can then be chosen as

$$\omega = \left(\frac{-(\mathcal{R}\gamma)k^2 + (\mathcal{I}\gamma)\varepsilon}{1 + |k^2 + \mathrm{i}\varepsilon|}\right)^2.$$

The numerical results are represented in Table 6.4 for the multigrid contraction number, and Table 6.5 shows the values of the smoothing property.

It might appear strange at first that $\|S^\nu\|_2$ is smaller than 1, although Theorem 4.5 gives $\|S^\nu\|_2 \leq C(\Omega)\sqrt{C_{\text{cont}}/C_{\text{coer}}}$. However, in this one dimensional setting, the two norms $\|.\|_2$ and $\|.\|_D$ are the same since $D = I(2/h - 2(k^2 + \mathrm{i}\varepsilon)h/3)$. The bound (4.13) then gives $\|S^\nu\|_2 \leq 1$ as we observe in Table 6.4. Table 6.5 shows that $\|A_l\left(I - \frac{\omega}{2}\gamma|D|^{-1}A\right)^\nu\|_2\|A_l\|_2^{-1}$ is decreasing when the number of smoothing steps increases. The precise rate is $\sqrt{\nu}^{-1}$, as we proved in Theorem 4.4.

Table 6.6 shows the value of the constant of the approximation property. This shows that this constant weakly depends on $k$ when the shift behaves like $O(k^2)$.

Table 6.7 shows the spectral radius of the two-grid iteration matrix (1.9). The latter is less than 1 for any number of smoothing steps and decreases as $\nu$ increases as expected from the convergence theorem. Similar results can be found in Table 6.8 for the W-cycle with 3 levels and various smoothing steps.

**Comparison of the two methods.** We now compare the method with the classical damped Jacobi smoother to the one with the modified Jacobi smoother, by computing numerically the spectral radius of each block $T_j$ given by the Fourier analysis. We use

$$\Omega = (0,1),\ N_l = 2^5 - 1,\ h_l = \frac{1}{N_l + 1},\ N_{l-1} = 2^4 - 1,\ h_{l-1} = \frac{1}{N_{l-1} + 1},$$

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $\nu = 1$ | 0.9177 | 0.8371 | 0.7636 | 0.6965 | 0.6354 |
| $\nu = 3$ | 0.9066 | 0.8228 | 0.7467 | 0.6776 | 0.6149 |
| $\nu = 5$ | 0.8728 | 0.7637 | 0.6694 | 0.6259 | 0.5852 |

TABLE 6.5
*Smoothing property* $\frac{\|A_l(I-\frac{\omega}{2}\gamma|D|^{-1}A)^\nu\|_2}{\|A_l\|_2}$ *for* $\varepsilon = O(k^2)$.

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| | 2.2010 | 2.1934 | 2.1695 | 2.1677 | 1.8790 |

TABLE 6.6
$\|A_l^{-1} - P_l A_{l-1}^{-1} R_l\|_2 \|A_l\|$ *as a function of* $k$ *for* $\varepsilon = O(k^2)$.

and the wave number goes from 1 to $4/h_l$ so that $kh_l \in [1, 4]$, which is according to Theorem 2.1 (see also Figure 2.1 on the left) the range where the two-grid cycle (and thus any multigrid cycle getting to this coarse resolution in its hierarchy) will have problems for a shift that is not large enough. We denote by $\omega_{\text{op}}$ the damping parameter used for the classical Jacobi smoother, and by $\omega_{\text{gen}}$ the one for the modified smoother.

We first choose a shift $\varepsilon = Ck^2$, and take $C = 0.8$. The general analysis gives for the damping parameter $\omega_{\text{gen}} = \left(\frac{-(\mathcal{R}\gamma)k^2+(\mathcal{I}\gamma)\varepsilon}{1+|k^2+\mathrm{i}\varepsilon|}\right)^2$, where $\gamma$ is such that $\mathcal{R}\gamma > 0$ and $\sqrt{\frac{(\mathcal{R}\gamma)^2}{1-(\mathcal{R}\gamma)^2}}k^2 < \varepsilon$, which yields $0 < \mathcal{R}\gamma < 4/\sqrt{41} \sim 0.625$, and hence we can take $\gamma = \frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}$. We show in Figure 6.1 the numerical results for one smoothing step. Clearly both multigrid methods converge, but the modified smoother, for which we presented a complete convergence analysis, leads to a less effective solver than the classical one.

We now choose a smaller shift, $\varepsilon = 0.8k$. Figure 6.2 shows the spectral radius of each $T_j$ and $\log(\rho(T))$ as function of the wave number when one smoothing step is used for the case of the classical Jacobi smoother. We see that this multigrid method fails to converge for some wave number and mesh size, as expected from our analysis. When using the modified Jacobi smoother from our general analysis, with the parameters

$$\gamma = \frac{\varepsilon + \mathrm{i}k}{|\varepsilon + \mathrm{i}k|}, \omega_{\text{gen}} = \left(\frac{\mathcal{R}\gamma}{1 + |k^2 + \mathrm{i}\varepsilon|}\right)^2,$$

we obtain Figure 6.3, where we see that the spectral radius is less than 1, but arbitrary close to 1 resulting in very bad convergence for the multigrid method. This is because $\omega_{\text{gen}} = O(k^{-4})$, so when $k$ increases, $\sigma_j \to 1$ (that is the eigenvalues of the iteration matrix of the smoother) and thus only the coarse grid operator remains. As a result, the spectral radius of each $T_j$ converges to 1. Indeed, the coarse grid operator is a projection so it has eigenvalues 0 and 1.

We show in Table 6.9 that increasing the number of smoothing steps helps for the modified smoother, but one can not obtain an efficient method when the shift is smaller than $O(k^2)$.

**6.2. Two dimensional numerical results.** We apply in this section the multigrid algorithm to the shifted-Helmholtz equation on the square $\Omega = (0, 1)^2$ equipped

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $\nu = 1$ | 0.9510 | 0.9044 | 0.8600 | 0.8179 | 0.7778 |
| $\nu = 3$ | 0.9457 | 0.8944 | 0.8458 | 0.7999 | 0.7565 |
| $\nu = 5$ | 0.8925 | 0.7965 | 0.7109 | 0.6345 | 0.5663 |

TABLE 6.7

*Spectral radius of the two-grid operator for increasing number of smoothing steps for $\varepsilon = O(k^2)$.*

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $\nu = 1$ | 0.9549 | 0.9537 | 0.9503 | 0.9362 | 0.9018 |
| $\nu = 3$ | 0.8706 | 0.8675 | 0.8581 | 0.8207 | 0.7334 |
| $\nu = 5$ | 0.7938 | 0.7890 | 0.7749 | 0.7194 | 0.5965 |

TABLE 6.8

*Spectral radius of the three-grid operator for increasing number of smoothing steps for $\varepsilon = O(k^2)$.*

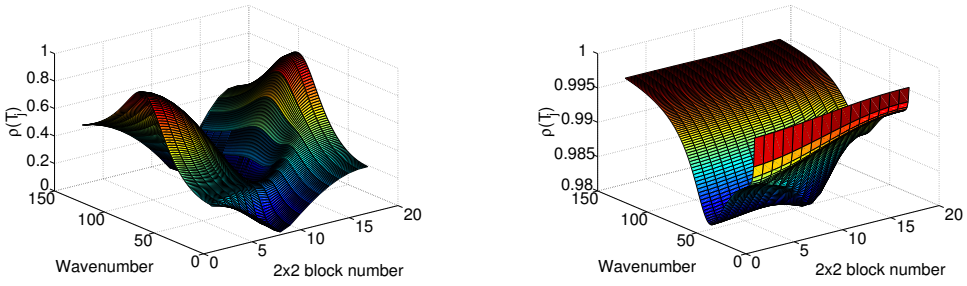


FIG. 6.1. *Left: Spectral radius of $T_j$ with $\omega = \omega_{\text{op}}$. Right: $\omega = \omega_{\text{gen}}$ for $C = 0.8$ and $\nu = 1$.*

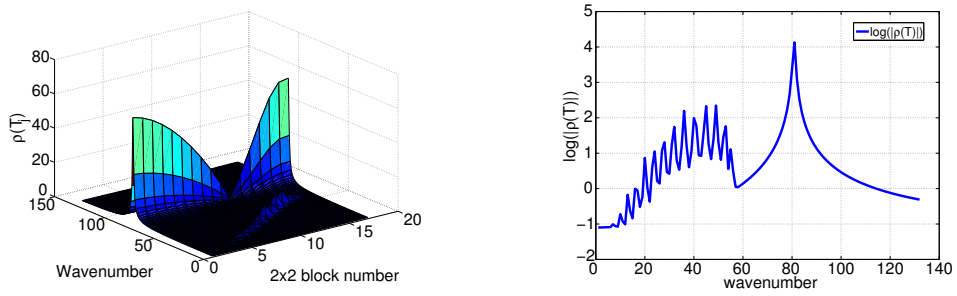


FIG. 6.2. *Spectral radius of $T_j$ with $\omega = \omega_{\text{op}}$ (left) and $log(|\rho(T)|)$ as a function of the wave number (right) for $C = 0.8$ and $\nu = 1$.*

| $k$ | 1 | 3 | 5 | 7 | 9 |
|---|---|---|---|---|---|
| $\nu = 1$ | 0.9770 | 0.9999 | 1 | 1 | 1 |
| $\nu = 10$ | 0.7924 | 0.9922 | 1 | 1 | 1 |
| $\nu = 1000$ | 0.0079 | 0.9226 | 0.9971 | 0.9997 | 0.9999 |

TABLE 6.9

*Spectral radius $\rho(T)$ when using the modified smoother and increasing the number of smoothing steps $\nu$.*

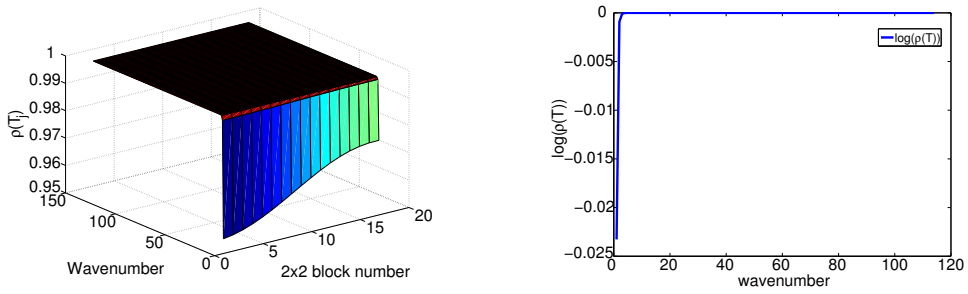


FIG. 6.3. *Spectral radius of $T_j$ with $\omega = \omega_{\text{gen}}$ (left) and $\log(\rho(T))$ as a function of the wave number (right) for $C = 0.8$ and $\nu = 1$.*

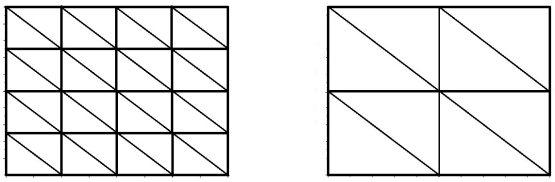

FIG. 6.4. *Examples of fine and coarse meshes for $N = 5$*

with Neumann boundary conditions, i.e.

$$\begin{cases} -\Delta u(x) - (k^2 + \mathrm{i}\varepsilon)u(x) &= f(x),\ x \in \Omega, \\ \partial_{\mathbf{n}} u &= 0,\ \text{on } \partial\Omega. \end{cases} \tag{6.2}$$

We consider a uniform mesh obtained from a uniform $N_l \times N_l$ grid. The mesh size is thus $h_l = 1/(N_l - 1)$. The nodes are counted in lexicographical order and the coarse mesh is obtained by considering a uniform $N_{l-1} \times N_{l-1}$ grid with $N_{l-1} = (N_l + 1)/2$. The coarse mesh size is then

$$h_{l-1} = 1/(N_{l-1} - 1) = 2h_l.$$

Figure 6.4 shows the fine mesh for $N = 5$ and the corresponding coarse one (see also [24] p.72 for other constructions of coarse meshes).

One needs the matrix representation of the prolongation and restriction operators. To get these, let $\phi_j^l$ be the nodal basis functions related to $\mathcal{V}_l$. The matrix $P_l$ of the prolongation operator (1.6) is obtained by interpolating the coarse nodal basis functions on the fine mesh. This gives

$$\forall\, \mu \in \{1, \cdots, N_{l-1}\},\ (I_l \phi_\mu^{l-1})(x) = \sum_{j=1}^{N_l} \phi_\mu^{l-1}(x_j^l)\phi_j^l(x).$$

As result, one gets

$$P_l = \left(\phi_\mu^{l-1}(x_j^l)\right)_{\mu,j} \in \mathrm{Hom}(\mathbb{C}^{N_{l-1}}, \mathbb{C}^{N_l}),\ \mu = 1, \cdots, N_{l-1},\ j = 1, \cdots, N_l.$$

Note that $\phi_\mu^{l-1}(x_j^l)$ is non-vanishing only on the fine nodes corresponding to the neighborhoods of $x_{l(j)}^{l-1} = x_j^l$. The restriction operator is $R_l = P_l^T$.

| $\nu$ | $\varepsilon$ | $k=10$ | $k=20$ | $k=40$ | $k=80$ | $k=160$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 0.999999999975240 | 1.000000000000025 | 1.000000000000046 | 1.000000000001082 | 1.000000000000053 |
| 3 | 1 | 0.999999999925725 | 0.999999999999980 | 1.000000000000049 | 1.000000000001164 | 1.000000000000047 |
| 1 | $k$ | 0.999999975841890 | 0.999999999801596 | 0.999999999998355 | 1.000000000000026 | 1.000000000000046 |
| 3 | $k$ | 0.999999927525641 | 0.999999999404773 | 0.999999999994992 | 0.999999999999972 | 1.000000000000064 |
| 1 | $k^{3/2}$ | 0.999999382351815 | 0.999999984191422 | 0.999999999600494 | 0.999999999983235 | 0.999999999999554 |
| 3 | $k^{3/2}$ | 0.999998147056534 | 0.999999952574267 | 0.999999998801486 | 0.999999999949603 | 0.999999999998608 |

TABLE 6.10
*Spectral radius of the two-grid operator for the* 2*d problem.*

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $kh$ | 0.1000 | 0.2000 | 0.4000 | 0.8000 | 1.6000 |
| $\nu=1$ | 0.9764 | 0.9757 | 0.9736 | 0.9648 | 0.9306 |
| $\nu=3$ | 0.9309 | 0.9289 | 0.9228 | 0.8982 | 0.8059 |
| $\nu=5$ | 0.8875 | 0.8843 | 0.8747 | 0.8361 | 0.6980 |

TABLE 6.11
*Spectral radius of the two-grid operator for increasing number of smoothing steps* $\nu$*.*

Below, we only give the numerical values of the spectral radius of the two-grid operator because the two-dimensional situation presents the same features as the 1D case.

**The case** $\varepsilon < O(k^2)$**.** We chose here $N_l = 49$ and $\varepsilon = 1, k, k^{3/2}$. This gives $49 \times 49 = 2401$ degrees of freedom for the fine mesh and 625 for the coarse mesh. With Theorem 3.1 and Remark (5.4), one can use the same parameters as for the 1D problem,

$$\gamma = \frac{\varepsilon + \mathrm{i}k^2}{|\varepsilon + \mathrm{i}k^2|},\ \omega = \left(\frac{\mathcal{R}\gamma}{1+|k^2+\mathrm{i}\varepsilon|}\right)^2,\ \varepsilon = 1.$$

Table 6.10 gives the spectral radius of the two-grid operator in this setting, very similar to the results in one dimension.

**The case** $\varepsilon = O(k^2)$**.** We consider a uniform mesh of $\Omega = (0,1)^2$ with $101 \times 101 = 10201$ degrees of freedom. Note that the coarse mesh has $51 \times 51$ degrees of freedom. We use the same parameters as for the one-dimensional numerical experiments, shown in (6.1). Since we have Neumann boundary conditions, Remark 5.4 ensures that the damping parameter for the smoother can be chosen as

$$\omega = \left(\frac{-(\mathcal{R}\gamma)k^2 + (\mathcal{I}\gamma)\varepsilon}{1+|k^2+\mathrm{i}\varepsilon|}\right)^2.$$

The spectral radius of the two-grid operator is shown in Table 6.11. These numerical results are in good agreement with the theory, and the convergence of the multigrid algorithm with the modified Jacobi smoother is improved by increasing the number of smoothing steps. Also note that this algorithm behaves better when the number of points per wavelength increases. Similar results can be found in Table 6.12 for the W-cycle with 3 levels and various smoothing steps.

**7. Directions for future work.** In this section, we present several numerical experiments in order to explore properties of the shifted Helmholtz preconditioner solved by multigrid which are not covered by the present theory.

| $k$ | 10 | 20 | 40 | 80 | 160 |
|---|---|---|---|---|---|
| $kh$ | 0.1000 | 0.2000 | 0.4000 | 0.8000 | 1.6000 |
| $\nu = 1$ | 0.9764 | 0.9757 | 0.9736 | 0.9648 | 0.9306 |
| $\nu = 3$ | 0.9309 | 0.9289 | 0.9229 | 0.8982 | 0.8058 |
| $\nu = 5$ | 0.8876 | 0.8844 | 0.8748 | 0.8362 | 0.6978 |

TABLE 6.12
*Spectral radius of the three-grid operator for increasing number of smoothing steps $\nu$.*

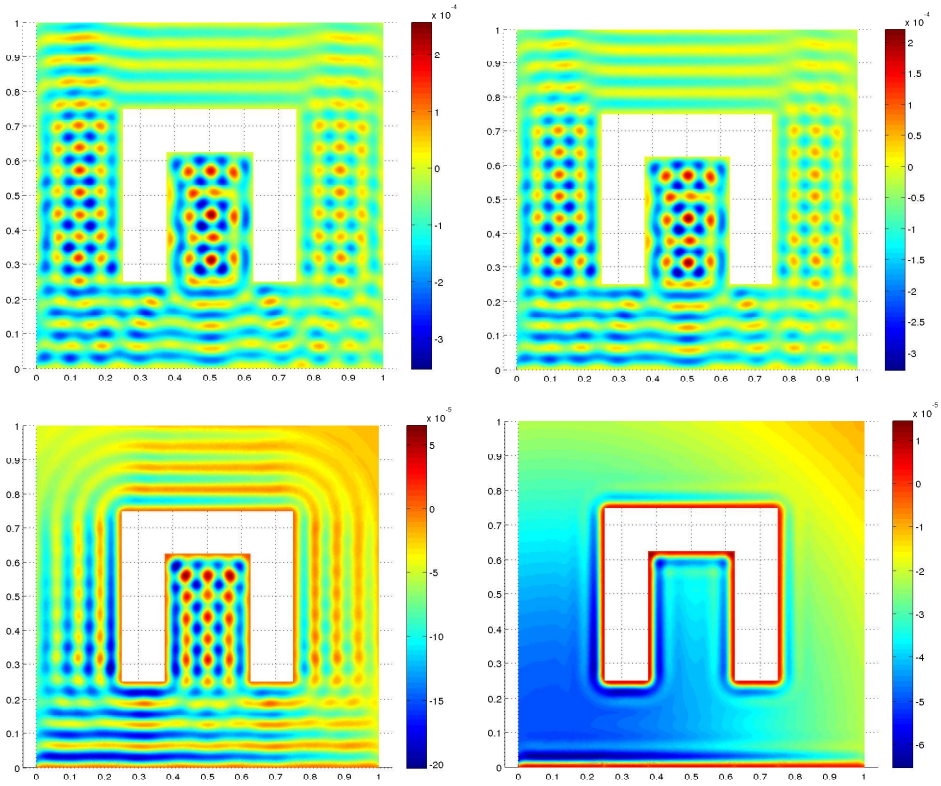

FIG. 7.1. *Top left: Solution of a scattering problem. Top right: solution computed with the operator shifted by $\varepsilon = k$. Bottom left: solution computed with the operator shifted by $\varepsilon = k^{3/2}$. Bottom right: solution computed with the operator shifted by $\varepsilon = k^2$.*

**7.1. Neither convex nor star-shaped domains.** The results from [17] hold for star-shaped domains and our theorems are valid for convex domains only. We illustrate here numerically why taking $\varepsilon = O(k)$ in the shifted Helmholtz preconditioner still yields a good preconditioner for such domains, and why a shift $\varepsilon = O(k^2)$ is too large.

We show in Figure 7.1 a scattering example on an open cavity in two spatial dimensions. On the obstacle, we impose Dirichlet conditions, and also on the wall represented by the bottom line, whereas on the other outer boundaries we impose a Robin radiation condition. For the right hand side, we use a Gaussian of the form $f(x, y) = e^{-((x-0.1)^2+(y-0.1)^2)}$, and the wave number is $k = 100$. We use a finite element discretization with 147456 P1 finite elements. In the top left panel, we show

the solution of the original, undamped Helmholtz equation, which we are in general interested in. In the top right panel, we show the solution when using the shifted Helmholtz operator with the maximal shift $\varepsilon = k$ such that GMRES would still converge independently of the wave number if the hypotheses in [17] were satisfied. The solution with this shift is very similar to the original solution we are interested in, and we thus have a graphical illustration why this could still be a good preconditioner. This property actually comes from Theorem 6.1 in [17] which shows that, if $\varepsilon/k$ is small enough, the solution to the original Helmholtz equation is close to the solution to the shifted problem. In the bottom left panel we show the solution using the shifted Helmholtz operator with the shift $\varepsilon = k^{3/2}$ for which wave number independent convergence can not be guaranteed any more according to [17]. We clearly see that the solution obtained now is quite different from the one in the top left panel we are interested in. We finally show in the bottom right panel the solution of this problem obtained when using the shifted Helmholtz operator with shift $\varepsilon = k^2$, which is the case we can solve effectively with multigrid according to our results. Unfortunately this solution does not have much in common any more with the one in the top left panel we are interested in.

These numerical simulations illustrate that a shift of order $k$ is probably the maximum to get a qualitatively effective shifted Helmholtz preconditioner. Nevertheless a shift behaving like $O(k^2)$ was used in practice successfully to improve the iterative solver, see [9, 35, 4, 33]. It is thus of interest to study the convergence of multigrid methods for $\varepsilon = O(k^2)$ which is, according to the results of this paper, the limiting case yielding convergence. Theorem 4.1 is still valid for non-star shaped domains, but one has to check the approximation property (see Theorem 4.3), whose proof uses the full elliptic regularity of problems (1.1,5.1). This kind of regularity does not hold for non-smooth domains [23], and thus the convergence of the multigrid algorithm does not readily follow from the analysis in this paper in this case.

**7.2. Preconditioned Problem with Multigrid.** We have seen that the shift required for multigrid to be an effective solver is too big for the shifted operator to be an effective preconditioner. One can however argue that one does not need to really invert the preconditioner, one would just use one or a few V-cycles, and then solve the preconditioned system with a Krylov method. It is therefore of interest to investigate the spectrum and the numerical range of such a preconditioned operator, to see if it might be theoretically possible to obtain an optimal choice for the shift and a convergence result for GMRES, as in [17], where it was shown that for small enough shifts ($\varepsilon/k$ small), the numerical range of the preconditioned operator stays away from zero, and thus with a result from [11] one can obtain robust convergence for preconditioned GMRES, independent of the wave number $k$.

We show in Figures 7.2-7.5 such a sequence of numerical experiments. We computed the spectrum and numerical range of the Helmholtz operator on a square, with Dirichlet conditions on top and bottom, and Robin radiation conditions on the left and right, i.e. the continuous wave guide like problem reads

$$\begin{cases} (\Delta + k^2)u &= -f \quad \text{in } (0,1)^2, \\ u &= 0 \quad \text{on } (0,1)\times\{0\}, \\ u &= 0 \quad \text{on } (0,1)\times\{1\}, \\ \partial_{\mathbf{n}} u - \mathrm{i}ku &= 0 \quad \text{on } \{0\}\times(0,1), \\ \partial_{\mathbf{n}} u - \mathrm{i}ku &= 0 \quad \text{on } \{1\}\times(0,1). \end{cases}$$

We discretized the problem using the classical five point finite difference stencil. From

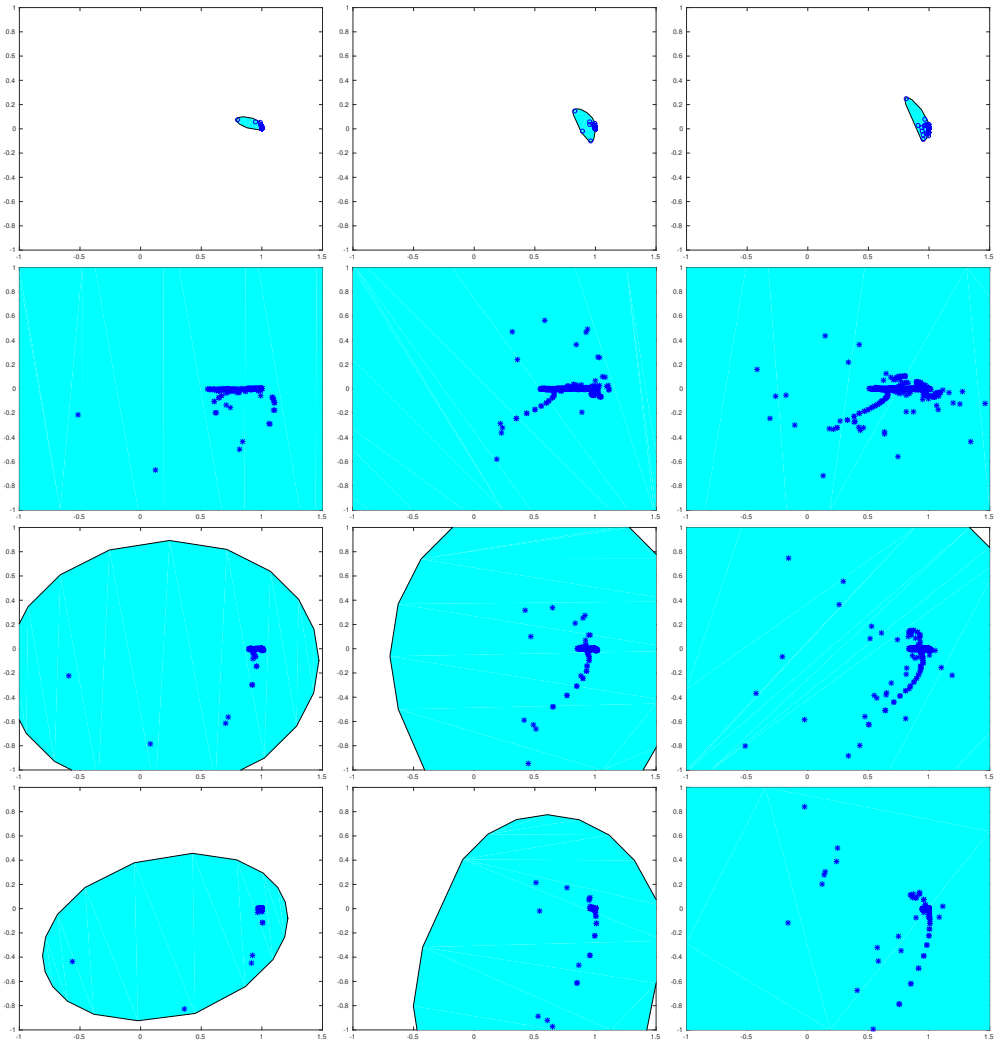

FIG. 7.2. *Case $\varepsilon = \sqrt{k}$: from left to right spectrum and numerical range of the preconditioned Helmholtz operator with increasing wave number $k \in \{\pi, 2\pi, 4\pi\}$ and corresponding mesh size $h \in \{\frac{1}{16}, \frac{1}{32}, \frac{1}{64}\}$. Top row using the exact inverse of the shifted preconditioner, and rows below performing one V-cycle with 1, 3 and 10 pre- and post-smoothing steps to approximately invert the shifted preconditioner.*

left to right in these figures, we increase the wave number $k \in \{\pi, 2\pi, 4\pi\}$ and diminish the corresponding mesh size $h \in \{\frac{1}{16}, \frac{1}{32}, \frac{1}{64}\}$, so that we have 32 points per wavelength on the fine grid. We invert the shifted preconditioner exactly, and also use one V-cycle to approximately invert it, using two, three and four levels for the corresponding values of $k$ and $h$, and a Jacobi smoother with damping parameter $2/3$, like in Figure 2.2. In Figure 7.2, we show the case of a shift $\varepsilon = \sqrt{k}$. In the top row we inverted the preconditioner exactly, and one can see that the numerical range stays away from zero for the resolutions tested, and thus GMRES convergence would be robust. This was proved also in [17], albeit under assumptions on the geometry that are not verified in the present setting. In the next rows, where we applied one V-cycle with 1, 3

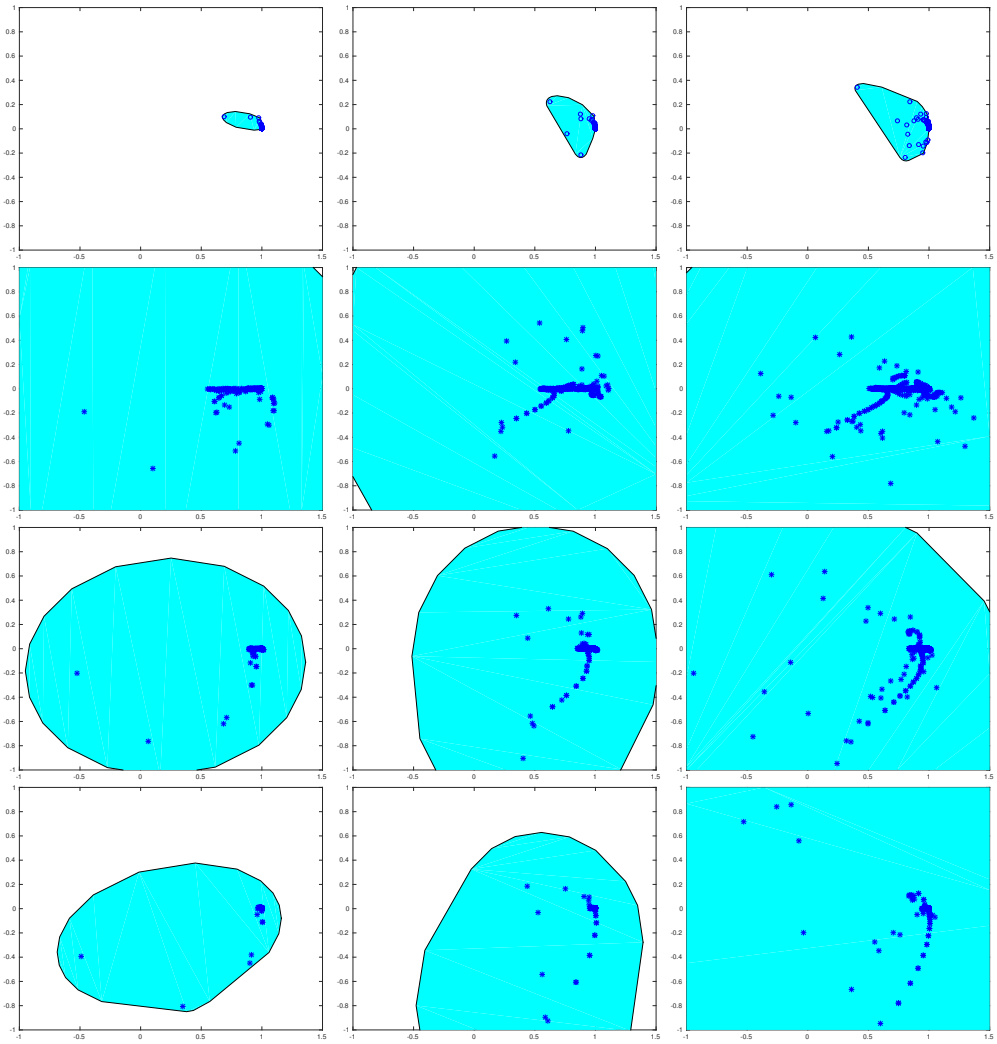

FIG. 7.3. *Case $\varepsilon = k$ for the same configuration as in Figure 7.2.*

and 10 pre- and post-smoothing steps to invert the preconditioner approximately, we see that immediately the clustering of the spectrum is lost, and the numerical range contains zero. From left to right, we also see that the situation gets worse as the wave number increases, the approximate inversion by the V-cycle is not effective. In Figure 7.3, we show the case of a shift $\varepsilon = k$, which is according to [17] the boundary where exact inversion will still lead to an effective preconditioner under suitable assumptions. We see in the top row of Figure 7.3 that exact inversion still leads to a well clustered spectrum, however now with a numerical range that slowly approaches 0, since for this wave guide problem, the assumptions in [17] are not met. The approximate inversion by a V-cycle produces unfavorable spectra for GMRES, and the situation gets worse as the wave number increases. In Figure 7.4, we show the case of a shift $\varepsilon = k^{3/2}$, too large according to [17] for exact inversion to lead to a good preconditioner. We see in the top row of Figure 7.4 that exact inversion now leads to

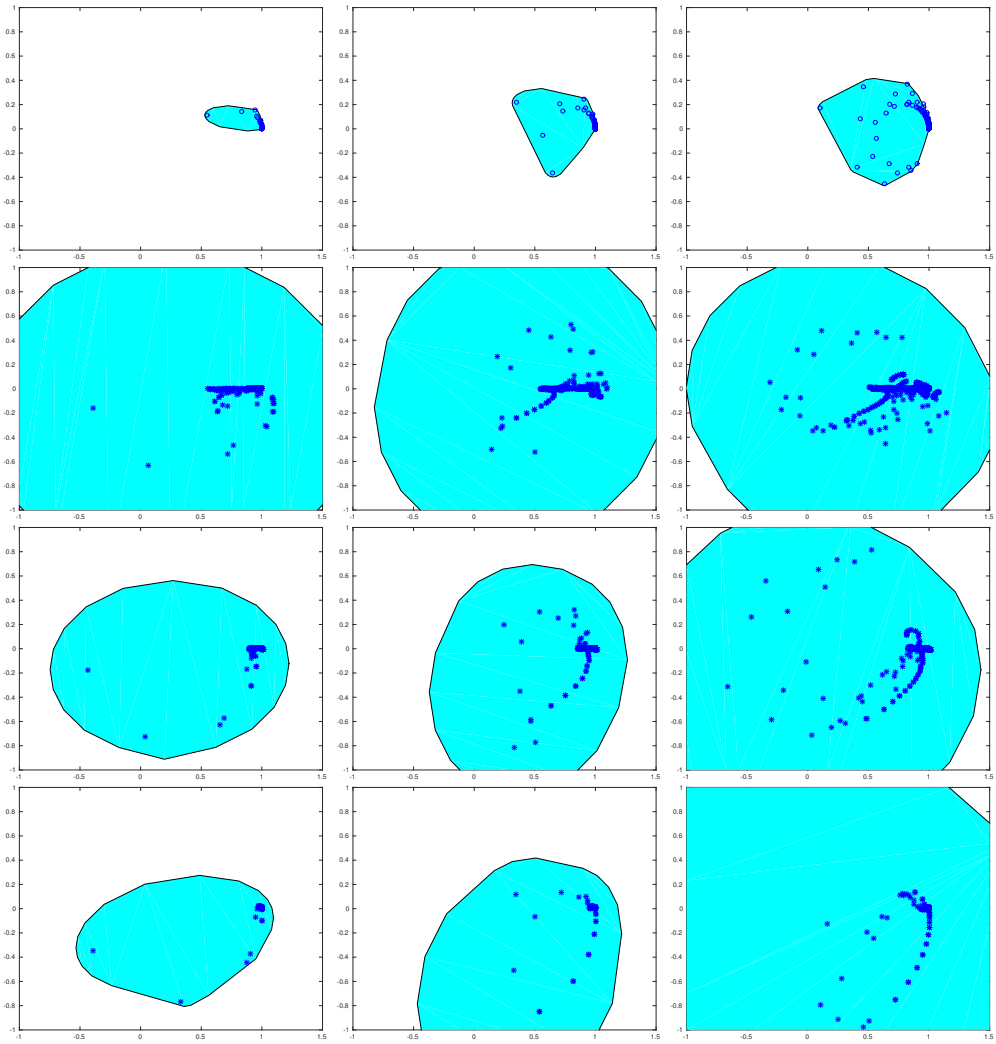

FIG. 7.4. *Case $\varepsilon = k^{3/2}$ for the same configuration as in Figure 7.2.*

a spectrum and numerical range that approaches zero as the wave number increases, which is not good for GMRES convergence. The approximate inversion by V-cycles also still produces unfavorable spectra for GMRES. We finally show in Figure 7.4 the case of a shift $\varepsilon = k^2$, much too large according to [17] for exact inversion to lead to a good preconditioner, as we observe in the top row: the spectrum and field of values is getting rapidly very close to zero as the wave number grows. The approximate inversion by V-cycles now better approximates the inverse of the preconditioner, but it is the preconditioned system itself that has an unfavorable spectrum for GMRES. These numerical experiments show that it is very difficult to obtain analytical results on the behavior of this type of preconditioner in the range of shifts $\varepsilon \in \{\sqrt{k}, k^2\}$, and that that iteration numbers of preconditioned GMRES in this case will depend on the wave number. This is illustrated in Table 7.1 when using as a right hand

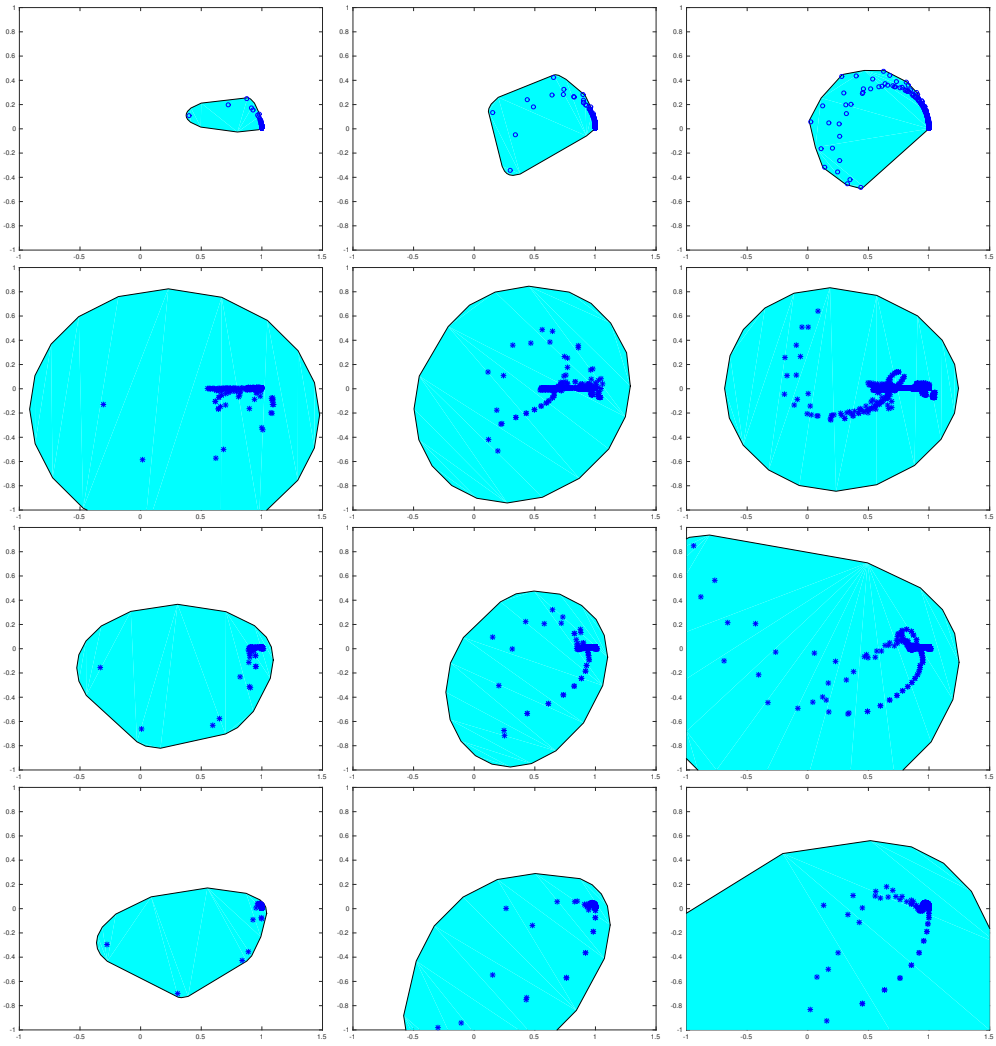

FIG. 7.5. *Case $\varepsilon = k^2$ for the same configuration as in Figure 7.2.*

side a random vector[1], and a zero initial guess for GMRES, run with a standard tolerance of $1e-6$. We see here from the first four lines with exact inversion that it is necessary to have a small shift to get low iteration numbers for preconditioned GMRES, and iteration numbers still grow slowly, as estimated in Figure 7.6 in the top left panel, since the hypotheses of [17] are not met in this example; otherwise for shifts $\varepsilon = \sqrt{k}$ and $\varepsilon = k$ iteration numbers would be constant, see [17] for the corresponding experiment. This changes however drastically as soon as we use multigrid V-cycles: iteration numbers grow rapidly for higher wave numbers using one V-cycle with one pre- and post-smoothing step, and the higher wave number problems can only be solved with many preconditioned GMRES iterations. With one and three pre- and

[1]To reproduce these results, use `rand('state',0)` in Matlab, and for the importance of the random initial guess containing all frequencies, see [16, Section 5.1], and the result at the end of this section.

| $\varepsilon$ | $\nu$ | $k=\pi$ | $k=2\pi$ | $k=4\pi$ | $k=8\pi$ | $k=16\pi$ | $k=32\pi$ |
|---|---|---|---|---|---|---|---|
| $\sqrt{k}$ | invert | 4 | 5 | 6 | 6 | 7 | 8 |
| $k$ | invert | 5 | 7 | 9 | 11 | 16 | 21 |
| $k^{3/2}$ | invert | 5 | 9 | 16 | 28 | 58 | 120 |
| $k^2$ | invert | 6 | 12 | 33 | 99 | 313 | 990 |
| $\sqrt{k}$ | $\nu=1$ | 160 | 462 | 567 | 1855 | 7171 | - |
| $k$ | $\nu=1$ | 175 | 628 | 548 | 1524 | 5873 | - |
| $k^{3/2}$ | $\nu=1$ | 189 | 810 | 483 | 1164 | 4267 | - |
| $k^2$ | $\nu=1$ | 190 | 860 | 950 | 775 | $>10000$ | - |
| $\sqrt{k}$ | $\nu=3$ | 87 | 317 | 532 | 764 | 3296 | - |
| $k$ | $\nu=3$ | 88 | 287 | 425 | 601 | 2126 | - |
| $k^{3/2}$ | $\nu=3$ | 88 | 237 | 270 | 336 | 1180 | - |
| $k^2$ | $\nu=3$ | 88 | 194 | 171 | 293 | 1000 | - |
| $\sqrt{k}$ | $\nu=10$ | 37 | 94 | 89 | 33 | 126 | 415 |
| $k$ | $\nu=10$ | 37 | 92 | 78 | 30 | 117 | 392 |
| $k^{3/2}$ | $\nu=10$ | 37 | 89 | 61 | 28 | 91 | 275 |
| $k^2$ | $\nu=10$ | 36 | 78 | 52 | 112 | 380 | 1322 |
| no preconditioner | | 63 | 148 | 426 | 1494 | 5711 | - |

TABLE 7.1

*GMRES iteration numbers for the experiments in Figures 7.2-7.5 going also to higher wave numbers and corresponding resolutions. Dash means the problem was too big to run to completion.*

post-smoothing steps, using the shift $\varepsilon = k^2$ as it is often done in practice is a little better than using a smaller shift, but this changes when ten pre- and post-smoothing steps are used: now clearly smaller shifts are better, the winning one for this example is $\varepsilon = k^{3/2}$, where e.g. for $k = 8\pi$, with $\nu = 10$ and 28 iterations a total of 280 pre and post smoothing iterations were used, compared to 775 pre- and post-smoothing iterations with $\nu = 1$ and $\varepsilon = k^2$, a factor of almost 3 less, in addition to the gain of the much smaller Krylov space.

We also see from Figure 7.6 that while initially it seems that the shifted Helmholtz preconditioner approximately inverted by multigrid is somehow effective, as soon as the wave number is large, iteration numbers grow like $O(k^2)$ for this wave guide like problem, like unpreconditioned GMRES shown in the last line of Table 7.1. The initial phase where the shifted Helmholtz preconditioner approximately inverted by multigrid seems to work is due to the high resolution chosen in this example, namely 32 points per wave length on the finest grid. In the small wave number cases $k \in \{\pi, 2\pi, 4\pi\}$, where the numerical range computations already indicate problems, using only $\{2, 3, 4\}$ levels in the multigrid hierarchy avoids the substantial difficulty indicated in Theorem 2.1. For larger wave numbers which require more levels, this is not the case any more and leads to the deterioration shown in Figure 7.6. This is confirmed in Table 7.2 and Figure 7.7, where the same experiments were run now with a resolution of 8 points per wave length on the finest grid. We see that while the exact inversion leads to very similar iteration numbers as in the high resolution case, iteration numbers when using multigrid to approximately invert the preconditioner start now growing right from the beginning. Also choosing a smaller shift than $k^2$ can be beneficial if more than one pre- and post-smoothing iteration are used for the approximate inversion in the V-cycle. Nevertheless, as we see in Figure 7.7, iteration numbers with this type of preconditioner grow like $O(k^2)$ for this wave guide like problem. Note that

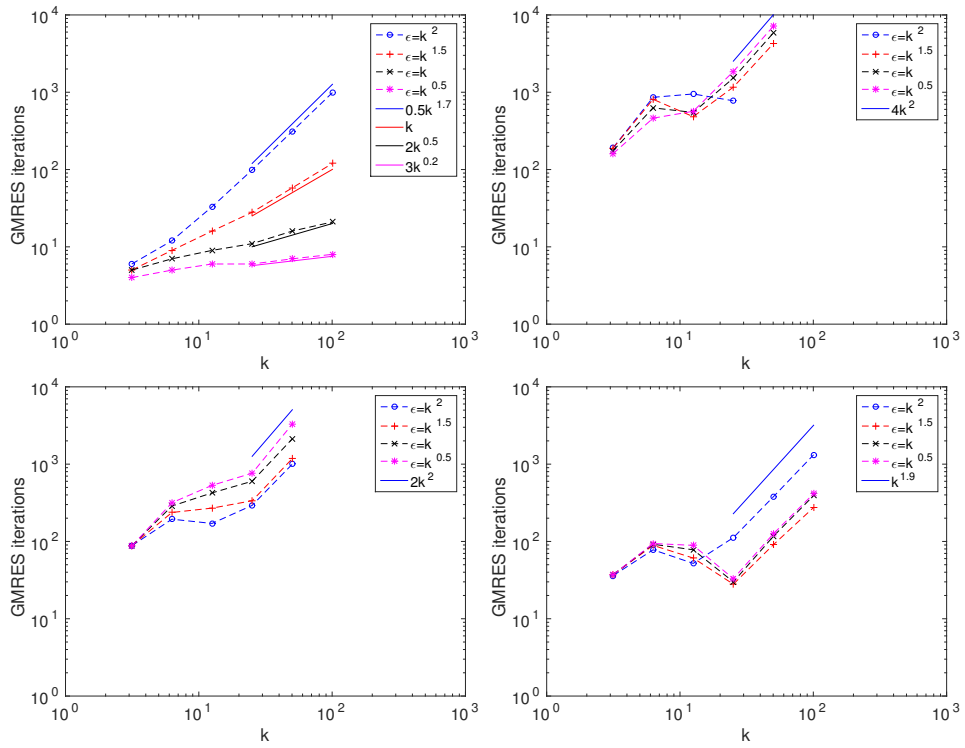


FIG. 7.6. *Graphical representation of the preconditioned GMRES iteration numbers from Table 7.1, with numerically estimated growth rates. Top left: exact inversion. Top right: multigrid preconditioner with* $\nu = 1$ *(*$\varepsilon = k^2$ *for* $k = 32\pi$ *did not converge in less than 10000 iterations, see Table 7.1). Bottom left: multigrid preconditioner with* $\nu = 3$*. Bottom right: multigrid preconditioner with* $\nu = 10$*.*

the last measurement for $\varepsilon = k^2$ is a bit better than expected. To investigate this, we ran the next higher resolution for $k = 64\pi$ which led to 7144 iterations, and we also ran the same sequence with a right hand side equal 1, and obtained the iteration numbers $4, 13, 58, 116, 314, 1123$, which show again the asymptotic growth $O(k^2)$ as in the other cases. The lower iteration counts than with the random right hand side also shows that the random right hand side is a much harder test for the multigrid preconditioner, since it contains all possible frequencies, see also the footnote[1].

In [31], a growth of only $O(k)$ was observed when exactly inverting the shifted Helmholtz preconditioner for a shift $O(k^2)$ for a model problem with Robin conditions all around. We obtain for this case the iteration numbers $6, 11, 21, 41, 71, 132$, and comparing with the corresponding fourth line of Figure 7.7, we see that the problem is indeed easier to solve, and growth seems to be only linear in $k$.

**8. Concluding remarks.** We have shown that in the shifted Helmholtz problem discretized by finite elements, the shift must be of the order of the wave number squared in order for multigrid to be a robust solver. Our results are based on two different types of analysis: Fourier analysis for a model problem with a standard damped Jacobi smoother, and functional analysis for the general case with a modified smoother, which gives a convergent multigrid method for all positive shifts. Convergence for a shift smaller than the order of the wave number squared comes however

| $\varepsilon$ | $\nu$ | $k=\pi$ | $k=2\pi$ | $k=4\pi$ | $k=8\pi$ | $k=16\pi$ | $k=32\pi$ |
|---|---|---|---|---|---|---|---|
| $\sqrt{k}$ | invert | 4 | 5 | 6 | 6 | 6 | 6 |
| $k$ | invert | 4 | 6 | 9 | 12 | 14 | 18 |
| $k^{3/2}$ | invert | 5 | 8 | 17 | 30 | 55 | 111 |
| $k^2$ | invert | 6 | 12 | 32 | 104 | 300 | 898 |
| $\sqrt{k}$ | $\nu=1$ | 9 | 49 | 222 | 870 | 2843 | 7331 |
| $k$ | $\nu=1$ | 9 | 49 | 217 | 878 | 3217 | 7633 |
| $k^{3/2}$ | $\nu=1$ | 9 | 49 | 211 | 882 | 3690 | $>10000$ |
| $k^2$ | $\nu=1$ | 9 | 47 | 194 | 784 | 3214 | $>10000$ |
| $\sqrt{k}$ | $\nu=3$ | 7 | 31 | 83 | 336 | 1270 | 4496 |
| $k$ | $\nu=3$ | 7 | 32 | 63 | 232 | 837 | 3046 |
| $k^{3/2}$ | $\nu=3$ | 7 | 37 | 59 | 224 | 532 | 1837 |
| $k^2$ | $\nu=3$ | 7 | 38 | 148 | 547 | 1781 | 3183 |
| $\sqrt{k}$ | $\nu=10$ | 3 | 6 | 7 | 33 | 128 | 416 |
| $k$ | $\nu=10$ | 3 | 6 | 7 | 30 | 123 | 401 |
| $k^{3/2}$ | $\nu=10$ | 3 | 9 | 7 | 29 | 105 | 274 |
| $k^2$ | $\nu=10$ | 3 | 11 | 34 | 74 | 257 | 1013 |
| no preconditioner | | 9 | 34 | 105 | 372 | 1413 | 5109 |

TABLE 7.2

*GMRES iteration numbers for the same experiment as in Table 7.1 but now with a finest resolution of only 8 points per wave length instead of 32.*

at the prize of many smoothing steps, and this can only be avoided with a shift of the size of the wave number squared. Our results apply as soon as one has a complex coercive variational form like for instance shifted Helmholtz problems with inhomogeneous media, and they also hold for other discretizations, like for example high order finite elements.

The fact that the shift must be of the order of the wave number squared for multigrid to be effective, together with the results in [17] which show that the shift should be at most order of the wave number for the shifted Helmholtz problem to be effective as preconditioner, do however not answer the most important question in practice: what shift should one choose as the best compromise? We showed with extensive numerical experiments on a difficult problem that iteration numbers with this type of preconditioning will in general grow like $O(k^2)$, and that it might be better to choose a shift smaller than $O(k^2)$ in contrast to current practice, especially if more pre- and post-smoothing iterations are used for the approximate inversion of the preconditioner.

One can see from our analysis that the constraint on the shift comes from the necessity to have a damping parameter for the smoother that is uniformly bounded with respect to the wave number. This condition is ensured if the ratio $C_{\text{coer}}/C_{\text{cont}}$ is uniformly bounded with respect to $k$ (see section 4.5), and thus the restriction on the shift seems to be inherent to the shifted Helmholtz equation. One possible continuation of the present work is the study of other smoothing procedures, satisfying the assumptions of Theorem 4.1, which could then give convergence of a multigrid algorithm for shifts smaller that $O(k^2)$.

The results of the present paper are also interesting since they can be associated to methods that improve the numerical solution of the original Helmholtz equation by preconditioning with a $k^2$-shifted Helmholtz operator [35, 9, 31, 4]. An efficient

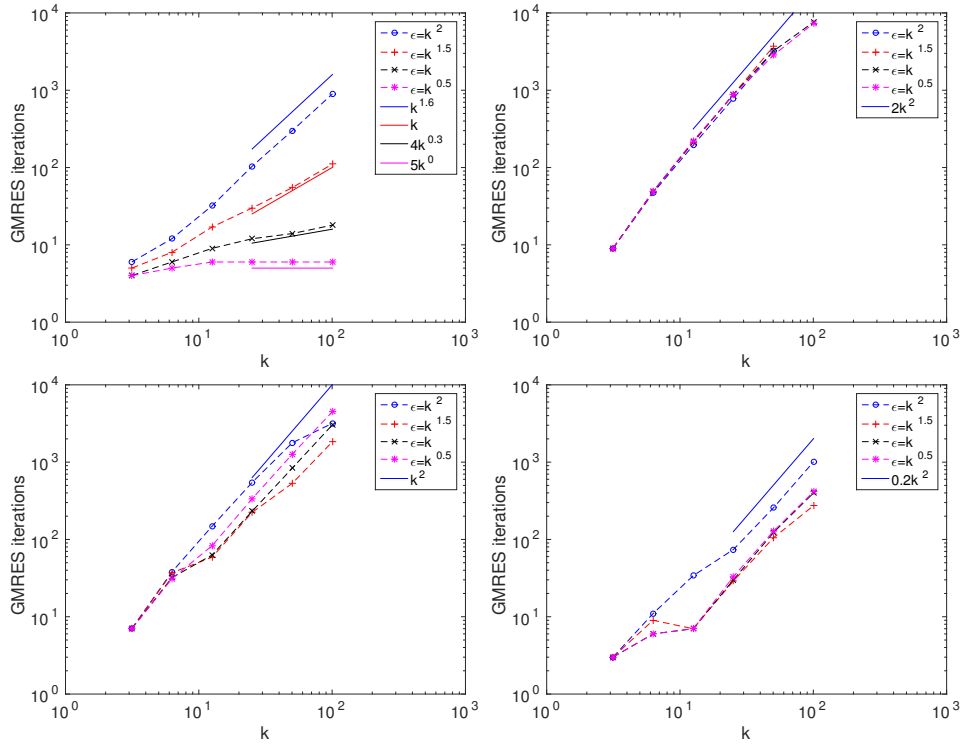


FIG. 7.7. *Graphical representation of the results from Table 7.2 with 8 points per wave length finest resolution, with numerically estimated growth rates. Top left: exact inversion. Top right: multigrid preconditioner with* $\nu = 1$ *(*$\varepsilon = k^2$ *and* $\varepsilon = k^{1.5}$ *for* $k = 32\pi$ *did not converge in less than 10000 iterations, see Table 7.2). Bottom left: multigrid preconditioner with* $\nu = 3$*. Bottom right: multigrid preconditioner with* $\nu = 10$*.*

inversion of the preconditioner is possible with the multigrid method presented here.

## 9. Appendix.

**Verification of the Galerkin condition.** We prove that choosing the canonical restriction operator ensures that the so called Galerkin condition $A_{l-1} = R_l A_l P_l$ is satisfied.

LEMMA 9.1. *Assume that* $\varepsilon > 0$*. Then one has*

$$A_{l-1} = R_l A_l P_l \text{ if and only if } R_l^* = P_l.$$

*Proof.* Recall that $\langle A_l \mathbf{z}_1, \overline{\mathbf{z}_2} \rangle = a(F_l \mathbf{z}_1, F_l \mathbf{z}_2)$ for all $\mathbf{z}_1,\ \mathbf{z}_2 \in \mathbb{C}^{N_l}$. From the definition of the prolongation operator $P_l = F_l^{-1} F_{l-1}$ and the well-posedness of problem (1.1), see Theorem 3.1, one has

$$\begin{aligned} R_l A_l P_l = A_{l-1} &\iff a(F_{l-1}\mathbf{z}_1, F_l R_l^* \mathbf{z}_2) = a(F_{l-1}\mathbf{z}_1, F_{l-1}\mathbf{z}_2),\ \forall \mathbf{z}_1,\ \mathbf{z}_2 \in \mathbb{C}^{N_l}, \\ &\iff F_l R_l^* \mathbf{z}_2 = F_{l-1}\mathbf{z}_2,\ \forall \mathbf{z}_2 \in \mathbb{C}^{N_l} \\ &\iff R_l^* = F_l^{-1} F_{l-1} = P_l. \end{aligned}$$

□

**Explicit bound for the $L^2$ error.** We give below the explicit constant appearing in the $L^2$ finite element error bound which is used to prove the approximation property.

LEMMA 9.2. *Let $a$ be a sesquilinear form on a closed subset $\mathcal{V} \subset H^1(\Omega)$ such that*

$$\forall u, v \in \mathcal{V},\ |a(u,v)| \leq C_{\text{cont}} \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)},$$
$$\forall u \in \mathcal{V},\ |a(u,u)| \geq C_{\text{coer}} \|u\|^2_{H^1(\Omega)}.$$

*Assume that, for any $f \in L^2(\Omega)$, the solution $u$ to $a(u,v) = (f, \overline{v})_{L^2(\Omega)}$ is in $H^2(\Omega)$ and that the solution $\varphi$ to the adjoint problem $\overline{a(v,\varphi)} = (f, \overline{\varphi})_{L^2(\Omega)}$ uniquely exists and they both satisfy*

$$\|u\|_{H^2(\Omega)} \leq C_{H^2} \|f\|_{L^2(\Omega)},\ \|\varphi\|_{H^2(\Omega)} \leq C_{H^2} \|f\|_{L^2(\Omega)}.$$

*Then, the $\mathbb{P}^1$ finite element approximation $u_h \in \mathcal{V}_h \subset \mathcal{V}$ of $u$ satisfies the error bound*

$$\|u - u_h\|_{L^2(\Omega)} \leq C(\Omega) h^2 C_{\text{cont}} \left(\frac{C_{\text{cont}}}{C_{\text{coer}}}\right)^2 C^2_{H^2} \|f\|_{L^2(\Omega)},$$

*where $C(\Omega) > 0$ is a constant that depends only on $\Omega$.*

*Proof.* We first recall the existence of an orthogonal projection $\Pi_h : L^2(\Omega) \mapsto \mathcal{V}_h$ that satisfies the estimate (see [2])

$$\forall u \in H^2(\Omega),\ \|u - \Pi_h u\|_{L^2(\Omega)} + h\|\nabla(u - \Pi_h u)\|_{L^2(\Omega)} \leq C(\Omega) h^2 \|u\|_{H^2(\Omega)}. \tag{9.1}$$

From the Galerkin orthogonality, $\forall v_h \in \mathcal{V}_h,\ a(u - u_h, v_h) = 0$, one has

$$a(u - \Pi_h u, u_h - \Pi_h u) = a(u_h - \Pi_h u, u_h - \Pi_h u).$$

Using the coercivity, the continuity of $a$, (9.1) and a triangle inequality, one gets

$$\|u - u_h\|_{H^1(\Omega)} \leq C(\Omega) \frac{C_{\text{cont}}}{C_{\text{coer}}} \|u\|_{H^2(\Omega)} \leq C(\Omega) \frac{C_{\text{cont}}}{C_{\text{coer}}} C_{H^2} \|f\|_{L^2(\Omega)}, \tag{9.2}$$

which gives the standard $H^1$ error estimate. Note that this bound also holds for the solution of the adjoint problem.

Now consider $\varphi \in \mathcal{V}$ solution to the adjoint problem $\overline{a(v,\varphi)} = (f, \overline{v})_{L^2(\Omega)}$ for all $v \in \mathcal{V}$. Let $\varphi_h \in \mathcal{V}_h$ be the finite element approximation of $\varphi$ satisfying $\overline{a(v_h, \varphi_h)} = (f, \overline{v_h})_{L^2(\Omega)}$ for all $v_h \in \mathcal{V}_h$. Now choosing $v = u - u_h$, one has

$$\begin{aligned}
\|u - u_h\|_{L^2(\Omega)} &= \sup_{f \in L^2(\Omega),\ f \neq 0} \frac{|(f, \overline{u - u_h})_{L^2(\Omega)}|}{\|f\|_{L^2(\Omega)}} \\
&= \sup_{\{f \in L^2(\Omega),\ f \neq 0\}} \frac{|a(u - u_h, \varphi)|}{\|f\|_{L^2(\Omega)}} \\
&= \sup_{\{f \in L^2(\Omega),\ f \neq 0\}} \frac{|a(u - u_h, \varphi - \varphi_h)|}{\|f\|_{L^2(\Omega)}} \\
&\leq C_{\text{cont}} \|u - u_h\|_{H^1(\Omega)} \sup_{\{f \in L^2(\Omega),\ f \neq 0\}} \frac{\|\varphi - \varphi_h\|_{H^1(\Omega)}}{\|f\|_{L^2(\Omega)}} \\
&\leq C(\Omega) h^2 C_{\text{cont}} \left(\frac{C_{\text{cont}}}{C_{\text{coer}}}\right)^2 C^2_{H^2} \|f\|_{L^2(\Omega)},
\end{aligned}$$

where we used Galerkin orthogonality and estimate (9.2) twice. ∎

From Theorems 3.1 and 5.1, one can see that all the assumptions of Lemma 9.2 are satisfied for the shifted Helmholtz equation with Dirichlet or impedance boundary conditions.

**Some results used to prove the smoothing property.** We now give the lemma used in the proof of Theorem 4.4.

LEMMA 9.3. *Let* $B \in \mathrm{Hom}(\mathbb{C}^m)$ *such that* $\|Bz\|_2^2 \leq C_B \mathcal{R}\langle Bz, \overline{z}\rangle$, $\forall z \in \mathbb{C}^m$. *Then* $\|I - \omega B\|_2 \leq 1$, $\forall \omega \in (0, \frac{2}{C_B})$.

*Proof.* Let $\omega$ be a positive real number. A direct computation gives

$$\|z - \omega Bz\|_2^2 = \|z\|_2^2 + w^2\|Bz\|_2^2 - 2\omega\mathcal{R}\langle Bz, \overline{z}\rangle .$$

The assumption then yields

$$\|z - \omega Bz\|_2^2 \leq \|z\|_2^2 + \omega\left(\omega - \frac{2}{C_B}\right)\|Bz\|_2^2,$$

which concludes the proof. ∎

COROLLARY 9.4 (Corollary 7.13 p.29 [30]). *Let* $\|.\|$ *be any induced matrix norm. Assume that for a linear iterative method with iteration matrix* $I - M^{-1}A_l$ *one has* $\|I - M^{-1}A_l\| \leq 1$. *Then for* $S := I - \frac{1}{2}M^{-1}A_l$, *we have the smoothing property*

$$\|A_l S^\nu\| \leq 2\sqrt{\frac{2}{\pi\nu}}\|M\|, \ \nu \geq 1.$$

**Bounds for the operator $F_l$.** We present some results from [30] used in our paper.

LEMMA 9.5 (Lemma 7.13 p.24 [30]). *There are two constants* $C_1$ *and* $C_2$ *that depend only on* $\Omega$ *such that*

$$C_1\|F_l z\|_{L^2(\Omega)} \leq h^{\frac{d}{2}}\|z\|_2 \leq C_2\|F_l z\|_{L^2(\Omega)}, \ \forall z \in \mathbb{C}^{N_l}.$$

*The same type of estimate holds for* $F_l^*$,

$$C_1\|F_l^* v_l\|_{L^2(\Omega)} \leq h^{\frac{d}{2}}\|v_l\|_{L^2(\Omega)} \leq C_2\|F_l^* v_l\|_{L^2(\Omega)}, \ \forall v_l \in \mathcal{V}_h.$$

## REFERENCES


[1] A. Brandt, S. Ta'asan. Multigrid method for nearly singular and slightly indefinite problems, *Springer Berlin Heidelberg*, 1986.

[2] S.C. Brenner, R. Scott. The mathematical theory of finite element methods, Vol. 15, *Springer*, 2008.

[3] J.H. Bramble, D. Y. Kwak, J.E. Pasciak. Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems, *SIAM journal on numerical analysis*, 31(6), 1746-1763, 1994.

[4] H. Calandra, S. Gratton, X. Pinel, X. Vasseur. An improved two-grid preconditioner for the solution of three-dimensional Helmholtz problems in heterogeneous media, *Numerical Linear Algebra with Applications*, vol. 20, no 4, p. 663-688, 2013.

[5] H. Chen, H. Wu, X. Xu. Multilevel preconditioner with stable coarse grid corrections for the Helmholtz equation, *SIAM Journal on Scientific Computing*, 37(1), A221-A244, 2015.

[6] P.H. Cocquet, M.J Gander. On the minimal shift in the shifted Laplacian preconditioner for multigrid to work, *To appear in Proceedings of the 22nd international conference on Domain Decomposition Methods*, 2015.
[7] S. Cools, B. Reps, W. Vanroose. A new level-dependent coarse grid correction scheme for indefinite Helmholtz problems, *Numerical Linear Algebra with Applications*, 2013.
[8] S. Cools, W. Vanroose. Local Fourier analysis of the complex shifted Laplacian preconditioner for Helmholtz problems, *Numerical linear algebra with applications*, 2013.
[9] S. Cools, W. Vanroose. Generalization of the complex shifted Laplacian: on the class of expansion preconditioners for Helmholtz problems, *arXiv preprint, arXiv:1501.04445*, 2015.
[10] M. Darbas, E. Darrigrand, Y. Lafranche. Combining analytic preconditioner and fast multipole method for the 3-D Helmholtz equation, *Journal of Computational Physics*, vol. 236, p. 289-316, 2013.
[11] S. C. Eisenstat, H.C. Elman, M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations, *SIAM Journal on numerical analysis*, 20(2), p. 345-357, 1983.
[12] H.C. Elman, O.G. Ernst, D. P. O'Leary. A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations, *SIAM Journal on scientific computing*, 23(4), p. 1291-1315, 2001.
[13] Y.A Erlangga, C. Vuik, C.W. Oosterlee. On a class of preconditioners for solving the discrete Helmholtz equation, *Applied Numerical Mathematics*, p. 409-425, 2004.
[14] O.G. Ernst and M.J. Gander. Why it is difficult to solve Helmholtz problems with classical iterative methods, Numerical Analysis of Multiscale Problems, I. Graham, T. Hou, O. Lakkis and R. Scheichl, Editors, *Springer Verlag*, p. 325-363 2012.
[15] O.G. Ernst and M.J. Gander. Multigrid Methods for Helmholtz Problems: A Convergent Scheme in 1D Using Standard Components, Direct and inverse problems in wave propagation and applications, I. Graham, U. Langer, M. Melenk and M. Sini eds, DeGruyter, p. 135-186, 2013.
[16] M.J. Gander. Schwarz Methods over the Course of Time, *ETNA*, Vol. 31, pp. 228–255, 2008.
[17] 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?, *Numerische Mathematik*, Vol. 53, No. 1, pp. 573–579, 2015.
[18] M.J. Gander, L. Halpern, Frédéric Magoules. An optimized Schwarz method with two-sided Robin transmission conditions for the Helmholtz equation, *International journal for numerical methods in fluids*, vol 55, no 2, p. 163-175, 2007.
[19] M.J. Gander, Frédéric Magoules, Frédéric Nataf. Optimized Schwarz methods without overlap for the Helmholtz equation, *SIAM Journal on Scientific Computing*, vol. 24, no. 1, p. 38-60, 2002.
[20] M.J. Gander, F. Nataf, F, AILU for Helmholtz problems: a new preconditioner based on the analytic parabolic factorization, *Journal of Computational Acoustics*, vol. 9, no. 04, p. 1499-1506, 2001.
[21] M.B. Van Gijzen, B. Martin, Y.A. Erlangga, C. Vuik. Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted Laplacian, *SIAM Journal on Scientific Computing*, 29(5), p.1942-1958, 2007.
[22] I.G. Graham, E.A. Spence, E. Vainikko, Domain decomposition preconditioning for high-frequency Helmholtz problems using absorption, *Submitted 7th July*, 2015
[23] P. Grisvard. Elliptic problems in non-smooth domains, *SIAM*, 2011.
[24] W. Hackbusch. Multi-grid methods and applications, 2. printing. - Berlin [u.a.] : Springer, 2003. - XIV, 377 p. (Springer series in computational mathematics ; 4), ISBN 978-3-540-12761-1 ISBN 3-540-12761-5 ISBN 0-387-12761-5.
[25] A. Hannukainen. Field of values analysis of a two-level preconditioner for the Helmholtz equation, *SIAM Journal on Numerical Analysis*, vol. 51, no 3, p. 1567-1584, 2013.
[26] F. Ihlenburg, I. Babuska, Finite Element Solution of the Helmholtz Equation with High Wave Number Part I: The h-Version of the FEM, *Computers Math. Applic.*, vol. 30, no. 9, p. 9-37, 1995.
[27] H. Knibbe, C.W. Oosterlee, C. Vuik. GPU implementation of a Helmholtz Krylov solver preconditioned by a shifted Laplace multigrid method, *Journal of Computational and Applied Mathematics*, vol. 236, no 3, p. 281-293, 2011.
[28] D. Osei-Kuffuor, Y. Saad. Preconditioning Helmholtz linear systems, *Applied numerical mathematics*, vol. 60, no 4, p. 420-431, 2010.
[29] D. Osei-Kuffuor, L. Ruipeng, Y. Saad. Matrix reordering using multilevel graph coarsening for ILU preconditioning, *SIAM Journal on Scientific Computing*, vol 37, no 1, p. A391-A419, 2015.

[30] A. Reusken. Introduction to multigrid methods for elliptic boundary value problems, *Multiscale Simulation Methods in Molecular Sciences*, p. 467-506, 2009.

[31] A.H. Sheikh, D. Lahaye, C. Vuik. On the convergence of shifted Laplace preconditioner combined with multilevel deflation, *Numerical Linear Algebra with Applications*, 20, p. 645-662, 2013.

[32] P. Tsuji, R. Tuminaro. Augmented AMG?shifted Laplacian preconditioners for indefinite Helmholtz problems, *Numerical Linear Algebra with Applications*, 2015.

[33] N. Umetani, S.P. MacLachlan, C.W. Oosterlee. A multigrid-based shifted Laplacian preconditioner for a fourth-order Helmholtz discretization, *Numerical Linear Algebra with Applications*, vol. 16, no 8, p. 603-626, 2009.

[34] A. Vion, C. Geuzaine. Double sweep preconditioner for optimized Schwarz methods applied to the Helmholtz problem. *Journal of Computational Physics*, vol. 266, p. 171-190, 2014.

[35] I. Zangré, X. Antoine and C. Geuzaine. High-Order Shifted Laplace Preconditioners for the Helmholtz Exterior Problem, *Submitted*, 2012.

[36] L. Zhu, H. Wu, Preasymptotic Error Analysis of CIP-FEM and FEM for Helmholtz Equation with High Wave Number. Part II: hp Version, *SIAM Journal on Numerical Analysis*, 51(3), p. 1828-1852, 2013.