On the minimal shift in the shifted Laplacian preconditioner for multigrid to work

Pierre-Henri Cocquet[1] , Martin J. Gander[1]

1 Introduction

Multigrid is an excellent iterative solver for discretized elliptic problems with diffusive nature, see [12] and the references therein. It is natural that substantial research was devoted to extend the multigrid method for solving the Helmholtz equation

$$-\Delta u - k^2 u = f \quad \text{in } \Omega \tag{1}$$

with the same efficiency, but it turned out that this is a very difficult task. Textbooks mention that there are substantial difficulties, see [3, page 72], [11, page 212], [12, page 400], and also the review [7] for why in general iterative methods have difficulties when applied to the Helmholtz equation (1).

Motivated by the early proposition in [2] to use the Laplacian to precondition the Helmholtz equation, the shifted Laplacian has been advocated over the past decade as a way of making multigrid work for the indefinite Helmholtz equation, see [6, 10, 1, 5, 4] and references therein. The idea is to shift the wave number into the complex plane to obtain a good preconditioner for a Krylov method when solving (1). The hope is that due to the shift, it becomes possible to use standard multigrid to invert the preconditioner, and if the shift is not too big, it is still an effective preconditioner for the Helmholtz equation with a real wave number. This implies however two conflicting requirements: the shift should be not too large for the shifted preconditioner to be a good preconditioner, and it should be large enough for multigrid to work. It was already indicated in [7] that it is not possible to satisfy both these requirements, see also [4]. It was then rigorously proved in [9] that the preconditioner is effective, i.e. iteration numbers stay bounded independently of the wave number k if the shift is at most of the size of the wavenumber. We prove here rigorously for a one dimensional model problem that if the complex shift is less than the size of the wavenumber squared, multigrid will not work. It is therefore not pos-

University of Geneva, 2-4 rue du Lièvre, CP 64, 1211 Genève, Switzerland, {martin.gander}{pierre-henri.cocquet}@unige.ch

sible to solve the shifted Laplace preconditioner with multigrid in the regime where it is a good preconditioner. We also show that if the complex shift is of the size of the wave number squared and the constant is large enough, then multigrid will solve the preconditioner independently of the wave number k. For a different shift idea as a dispersion correction, where the shift is real and one obtains in one dimension a multigrid solver with standard components that solves the original Helmholtz problem (1) independently of the wave number, see [8].

2 Model problem and discretization

To study the shifted Laplacian preconditioner for the Helmholtz equation (1) in 1d, we consider the 1d shifted Helmholtz equation

$$-u''(x) - (k^2 + i\varepsilon)u(x) = f(x) \quad x \text{ in } (0,1) \tag{2}$$

with homogeneous Dirichlet boundary conditions $u(0) = u(1) = 0$. We discretize (2) using a standard 3-point centered finite difference approximation on a uniform mesh with n interior grid points and mesh size $h = 1/(n+1)$ to get a linear system $A_h\mathbf{u} = \mathbf{f}$ with system matrix

$$A_h = \frac{1}{h^2}\text{tridiag}(-1, 2-(k^2+i\varepsilon)h^2, -1). \tag{3}$$

It is this system matrix which is used as a preconditioner for solving (1), and therefore following the idea of the shifted Laplacian preconditioner, systems with this matrix have to be solved effectively using multigrid. We analyze here in detail a two grid method: we use a Jacobi smoother,

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

where $D = \text{diag}(A)$, and ω is a relaxation parameter, which we choose here based on the optimal choice of the case without relaxation, see [8], to be

$$\omega := \frac{2-(k^2+i\varepsilon)h^2}{3-(k^2+i\varepsilon)h^2}.$$

For the coarse correction, we assume n to be a power of two minus one, and use the extension operator based on interpolation,

$$I_H^h = \frac{1}{2}\begin{bmatrix} 2 & & & & \\ 1 & 1 & & & \\ & 2 & & & \\ & 1 & & & \\ & & \ddots & 1 & \\ & & & 2 & \\ & & & 1 & 1 \\ & & & & 2 \end{bmatrix} \in \mathbb{R}^{(n+2)\times(N+2)}, \quad N := (n+1)/2 - 1,$$

and for the restriction $I_h^H = \frac{1}{2}(I_H^h)^T$, with the coarse grid matrix obtained by Galerkin projection, $A_H := I_h^H A_h I_H^h$. The resulting two grid operator with ν_1 pre-smoothing and ν_2 post-smoothing steps is then of the form

$$T := (I - \omega D^{-1} A_h)^{\nu_1} (I - I_H^h A_H^{-1} I_h^H A_h)(I - \omega D^{-1} A_h)^{\nu_2}. \tag{4}$$

Using the subspaces

$$\operatorname{span}\{\mathbf{v}_1^h, \mathbf{v}_n^h\},\ \operatorname{span}\{\mathbf{v}_2^h, \mathbf{v}_{n-1}^h\},\ \ldots,\ \operatorname{span}\{\mathbf{v}_N^h, \mathbf{v}_{N+2}^h\},\ \operatorname{span}\{\mathbf{v}_{N+1}^h\} \tag{5}$$

defined by the eigenfunctions of A_h given by $\mathbf{v}_j^h := [\sin j\ell\pi h]_{\ell=1}^n$, $j = 1, \ldots, n$, one can block diagonalize the two grid operator (4), see [8]. The action of T on these one- and two-dimensional subspaces is represented by the block diagonal matrix $\operatorname{diag}(T_1, \ldots, T_N, T_{N+1})$ with

$$T_j = \begin{bmatrix} \sigma_j & 0 \\ 0 & \sigma_{j'} \end{bmatrix}^{\nu_2} \begin{bmatrix} 1 - c_j^4 \frac{\lambda_j^h}{\lambda_j^H} & c_j^2 s_j^2 \frac{\lambda_{j'}^h}{\lambda_j^H} \\ c_j^2 s_j^2 \frac{\lambda_j^h}{\lambda_j^H} & 1 - s_j^4 \frac{\lambda_{j'}^h}{\lambda_j^H} \end{bmatrix} \begin{bmatrix} \sigma_j & 0 \\ 0 & \sigma_{j'} \end{bmatrix}^{\nu_1},\ T_{N+1} = \sigma_{N+1}^{\nu_1+\nu_2}, \tag{6}$$

where $c_j := \cos\frac{j\pi h}{2}$, $s_j := \sin\frac{j\pi h}{2}$, $j = 1, \ldots, N$, $\sigma_j := 1 - \omega(1 - \frac{2\cos(j\pi h)}{2-(k^2+i\varepsilon)h^2})$, $j = 1, \ldots, n$, and

$$\lambda_j^h := \frac{4}{h^2}\sin^2\frac{j\pi h}{2} - (k^2 + i\varepsilon), \qquad j = 1, \ldots, n, \tag{7}$$

$$\lambda_j^H := \frac{4}{H^2}\sin^2\frac{j\pi H}{2} - (k^2 + i\varepsilon), \qquad j = 1, \ldots, N, \tag{8}$$

are the eigenvalues of A_h and A_H, with $j' := N + 1 - j$ denoting the complementary mode index. To prove convergence of the two grid method, one has to prove that the spectral radius of T_j is smaller than one for all $j = 1, \ldots, N+1$, since this implies that the spectral radius of the two grid operator T is less than one. We will show in the next section that if the shift is not large enough, the spectral radius of T will actually be bigger than one, and hence the two grid method can not be convergent.

3 Analysis

We first study the case of a shift $\varepsilon = Ck^{2-\delta}$, $0 \leq \delta < 2$. The following theorem shows that with such a shift, it is not possible to obtain robust multigrid convergence, because for any small mesh parameter h, there exists a wavenumber of the Helmholtz equation for which the two grid method diverges.

Theorem 1 (Divergence with too small shift). *Assume that we are performing* $\nu = \nu_1 + \nu_2$ *smoothing steps and that* $\varepsilon = Ck^{2-\delta}$ *for* $2 > \delta \geq 0$. *Then, for* h *small enough, there exists a wavenumber* $k(h)$ *such that the spectral radius of the two grid method satisfies*

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

and hence the two grid method diverges for this mesh size and wavenumber.

Proof. Denoting by μ_j the eigenvalues of the iteration operator T we have

$$\rho(T) \geq |\mu_j|, \ j = 1, \cdots, n.$$

Using the block diagonal form of the two-grid iteration matrix we have obtained in (6), we have in particular

$$\rho(T) \geq |\sigma_{N+1}^{\nu_1+\nu_2}| = |1-\omega|^{\nu} = |\mu_{N+1}|^{\nu},$$

with

$$|\mu_{N+1}| := \frac{1}{\sqrt{(3-k^2h^2)^2 + C^2h^4k^{4-2\delta}}}. \tag{9}$$

We now wish to find the maximum of $|\mu_{N+1}|$ as a function of the wavenumber k. Taking a derivative with respect to k, we obtain

$$\partial_k(|\mu_{N+1}(k)|^2) = k^{2\delta+1}\frac{2h^2(C^2k^2\delta h^2 - 2C^2k^2h^2 - 2k^{2\delta+2}h^2 + 6k^{2\delta})}{(C^2k^4h^4 + h^4k^{4+2\delta} - 6k^{2+2\delta}h^2 + 9k^{2\delta})^2},$$

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

$$C^2k^2(\delta-2)h^2 - 2k^{2\delta+2}h^2 + 6k^{2\delta} = 0. \tag{10}$$

Since this equation can not be solved in closed form, we compute an asymptotic expansion of $k(h)$ for small mesh size h. We make the Ansatz

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

and obtain for h small enough the expansions

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

Substituting the above expressions into the equation (10) satisfied by $k(h)$ and considering only the leading order terms, we find

$$\frac{1}{h^{2\delta}}\left(6\alpha_0^{2\delta+2} - 2\alpha_0^{2\delta}\right) + o\left(\frac{1}{h^{2\delta}}\right) = 0,$$

and therefore

$$\alpha_0 = \sqrt{3},$$

and one can check that this is indeed asymptotically a maximum. We now replace the asymptotic expansion of $k(h)$ into the expression for $|\mu_{N+1}(k(h))|$ given in (9). Since $k(h)h = \sqrt{3} + o(1)$, a Taylor expansion shows that

$$\rho(T) \geq |\mu_{N+1}(k(h))| = \frac{1}{\sqrt{(3-k(h)^2h^2)^2 + C^2h^4k(h)^{4-2\delta}}} = \frac{3^{\delta/2}}{3Ch^{\delta}} + o\left(\frac{1}{h^{\delta}}\right),$$

which gives the result.

Remark 1. In our proof, we only gave the first term of the asymptotic expansion of $k(h)$, since this was sufficient to obtain divergence. One could however compute the asymptotic expansion also to any order without additional difficulties.

Now we study the case $\varepsilon = Ck^2$. Substituting this value into the blocks (6) of the block diagonal representation, we notice that the matrices become homogeneous functions of the product kh. One can therefore study the spectral radius directly as a function of $kh > 0$ and $c_j \in (0,1)$, using trigonometric formulas to replace the dependency on s_j. We show in Figure 1 on the left for $\nu_1 = 1$, $\nu_2 = 0$ the maximum over all kh of the spectral radius of the matrix T as a function of C for $\varepsilon = Ck^2$. We clearly see that for C small, multigrid does not converge. For C larger however, we get convergence. The value C^* where the spectral radius equals one can be computed, it is $C^* = 0.3850$. We show on the right in Figure 1 the spectrum of the blocks T_j, represented as a continuous function of $c_j \in (0,1)$ and kh for $C = C^*$, and one can clearly see where the maximum value one is reached.

Remark 2. The value C^* is larger than the limiting value $C = 1/3$ found from the limiting case as δ goes to zero in Theorem 1 for which divergence can be guaranteed. This is because Theorem 1 only provides a lower bound for which divergence can be guaranteed. As we see from the sharper analysis above, divergence even set in a bit earlier.

Remark 3. From Figure 1 on the left, we also see that making C very large will eventually not lead to further improvement, the curve has an asymptote which one can compute to be at $1/3$. Hence, the best contraction factor one can achieve with the two grid method applied to the shifted Helmholtz equation with shift $\varepsilon = Ck^2$ for C large in our example is $1/3$. Note also that the two grid convergence is uniform in k as soon as $C > C^*$.

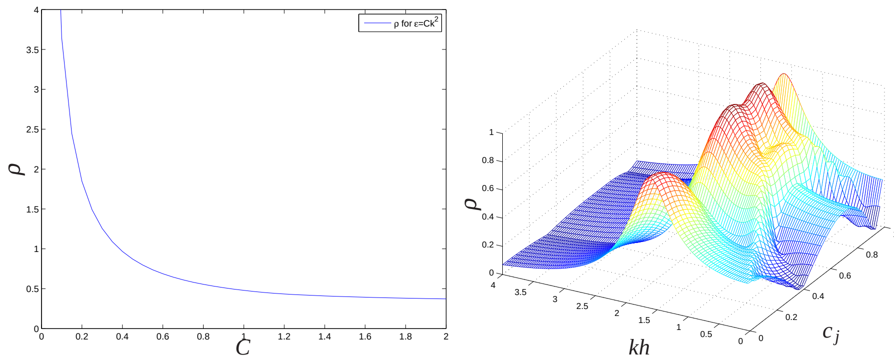

Fig. 1 Maximum over kh of the spectral radius of the two grid operator for shift $\varepsilon = Ck^2$ as a function of C on the left, and for $C = 0.3850$ the spectrum as a function of kh and c_j on the right

4 Numerical Experiments

We present in this section several numerical illustrations of Theorem 1 and our additional estimate for the shift $\varepsilon = Ck^2$. We assume that the source term in the shifted Helmholtz equation (2) is $f = 0$ giving $u = 0$ as the unique solution. We use for our simulations the parameters

$$n = 511, \quad h = \frac{1}{512}, \quad k = \frac{\sqrt{3}}{h}, \quad \nu = 1,$$

so that we are in the regime of Theorem 1 where divergence should be observed if the shift is not sufficient. We perform twenty iterations of the two grid method applied to the shifted problem, starting with a random initial guess.

We first illustrate the result of Theorem 1. We choose $C = 0.45$ in the shift $\varepsilon = Ck^{2-\delta}$. Figure (2) shows the relative error of the two-grid scheme versus the number of iterations for various values of δ. We see that the two grid method converges for $\delta = 0$, but diverges for all other values $\delta > 0$. For the value of $h = 1/512$ in our experiment, and the constant $C = 0.45$, we see that the two grid method would still converge for a very small, but positive value of δ. This is not in disagreement with Theorem 1, which only makes a statement for h small enough.

We next show an experiment to illustrate that even with the shift $\varepsilon = Ck^2$, the constant still needs to be bigger than $C^* = 0.3850$ for the two grid method to converge, see also Remark 2. In Figure 3 we show the relative error versus the iteration index for various values of C in this case. We observe that for $C < C^*$ the multigrid method does not converge, the shift is not enough. For $C > C^*$ however the multigrid method converges, and convergence gets faster as C increases, as expected. There is however a limit, as we have seen in Remark 3, the contraction factor of the two grid method will not be better than $1/3$.

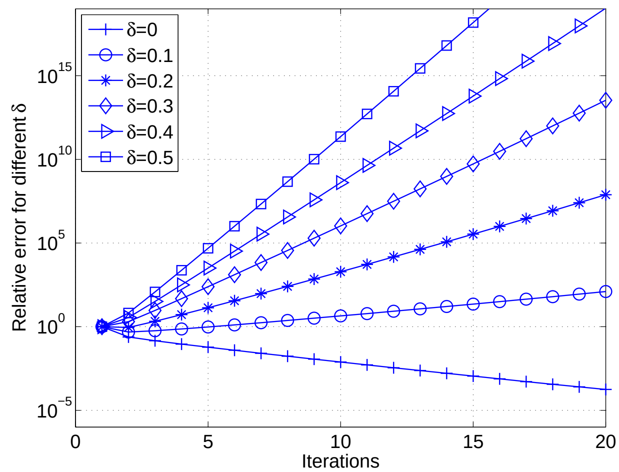

Fig. 2 Relative error versus iteration index for $C = 0.45$ and various values of $\delta = 0.2$

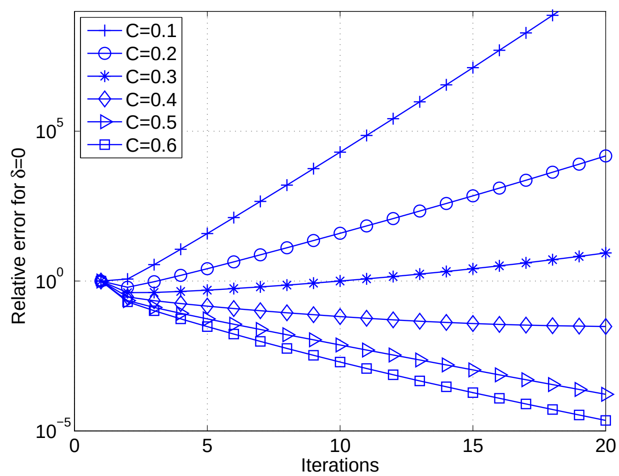

Fig. 3 Relative error versus iteration index for $\delta = 0$ and various values of C

5 Conclusions

We have analyzed for the shifted Helmholtz operator how large a shift of the form $\varepsilon = Ck^{2-\delta}$ has to be to obtain a uniformly convergent two grid method. We have proved for a one dimensional model problem that uniform convergence in the wavenumber k is not possible if $\delta > 0$. For $\delta = 0$, we have shown that if the constant $C > C^* = 0.3850$, then uniform convergence in the wavenumber k can be achieved. Our results are for the particular case of a one dimensional problem with a second order finite difference discretization, using a Galerkin coarse grid correction with full weighting and a Jacobi smoother with particular relaxation parameter. Using a different relaxation parameter, for example $\omega = 2/3$, leads to slightly worse results in this case, e.g. C^* becomes approximately 0.75 instead of 0.3850. Our analysis can be generalized, for example to higher dimensions, or other discretizations.

There is therefore indeed a big gap in the requirements for using the shifted Laplacian as a preconditioner when solving discretized Helmholtz problems: for multigrid to invert the preconditioner efficiently, the shift needs to be $O(k^2)$, but to prove that the preconditioner is effective, the shift needed to be at most $O(k)$, see [9], where numerical experiments also indicate that this estimate is sharp. Any compromise with the shift, i.e. using a shift of $O(k^\alpha)$ with $\alpha \in (1,2)$, will therefore lead to a preconditioner which is outside the requirements one would like to impose.

References

1. Airaksinen, T., Heikkola, E., Pennanen, A., Toivanen, J.: An algebraic multigrid based shifted-laplacian preconditioner for the Helmholtz equation. Journal of Computational Physics **226**(1), 1196–1210 (2007)
2. Bayliss, A., Goldstein, C., Turkel, E.: An iterative method for the Helmholtz equation. J. Comput. Phys. **49**, 443–457 (1983)
3. Brandt, A., Livne, O.E.: Multigrid Techniques, 1984 Guide with Applications to Fluid Dynamics, Revised Edition. No. 67 in Classics in Applied Mathematics. SIAM (2011)
4. Cools, S., Vanroose, W.: Local Fourier analysis of the complex shifted Laplacian preconditioner for Helmholtz problems. Numerical Linear Algebra with Applications **19**(2), 232–252 (2013)
5. Erlangga, Y.: Advances in iterative methods and preconditioners for the Helmholtz equation. Archives Comput. Methods in Engin. **15**, 37–66 (2008)
6. Erlangga, Y., Vuik, C., Oosterlee, C.: On a class of preconditioners for solving the Helmholtz equation. Applied Numerical Mathematics **50**, 409–425 (2004)
7. Ernst, O., Gander, M.: Why it is difficult to solve Helmholtz problems with classical iterative methods. In: I. Graham, T. Hou, O. Lakkis, R. Scheichl (eds.) Numerical Analysis of Multiscale Problems, pp. 325–363. Springer Verlag (2012)
8. Gander, M., Ernst, O.: Multigrid methods for Helmholtz problems: A convergent scheme in 1d using standard components. In: M.M. I. Graham U. Langer, M. Sini (eds.) Direct and Inverse Problems in Wave Propagation and Applications, pp. 135–186. De Gruyter (2013)
9. Gander, M., Graham, I.G., Spence, E.A.: How should one choose the shift for the shifted laplacian to be a good preconditioner for the Helmholtz equation? Numerische Mathematik (2014). Submitted
10. Gijzen, M.V., Erlangga, Y., Vuik, C.: Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted Laplacian. SIAM J. Sci. Comput. **29**(5), 1942–1958 (2007)
11. Hackbusch, W.: Multi-Grid Methods and Applications. Springer-Verlag (1985)
12. Trottenberg, U., Oosterlee, C.C.W., Schüller, A.: Multigrid. Academic Press (2001)