Electronic Transactions on Numerical Analysis.
Volume 31, pp. 403-424, 2008.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

ON A MULTILEVEL KRYLOV METHOD FOR THE HELMHOLTZ EQUATION PRECONDITIONED BY SHIFTED LAPLACIAN*

YOGI A. ERLANGGA† AND REINHARD NABBEN‡

Abstract. In Erlangga and Nabben [SIAM J. Sci. Comput., 30 (2008), pp. 1572–1595], a multilevel Krylov method is proposed to solve linear systems with symmetric and nonsymmetric matrices of coefficients. This multilevel method is based on an operator which shifts some small eigenvalues to the largest eigenvalue, leading to a spectrum which is favorable for convergence acceleration of a Krylov subspace method. This shift technique involves a subspace or coarse-grid solve. The multilevel Krylov method is obtained via a recursive application of the shift operator on the coarse-grid system. This method has been applied successfully to 2D convection-diffusion problems for which a standard multigrid method fails to converge.

In this paper, we extend this multilevel Krylov method to indefinite linear systems arising from a discretization of the Helmholtz equation, preconditioned by shifted Laplacian as introduced by Erlangga, Oosterlee and Vuik [SIAM J. Sci. Comput. 27 (2006), pp. 1471–1492]. Within the Krylov iteration and the multilevel steps, for each coarse-grid solve a multigrid iteration is used to approximately invert the shifted Laplacian preconditioner. Hence, a multilevel Krylov-multigrid (MKMG) method results.

Numerical results are given for high wavenumbers and show the effectiveness of the method for solving Helmholtz problems. Not only can the convergence be made almost independent of grid size h, but also linearly dependent on the wavenumber k, with a smaller proportional constant than for the multigrid preconditioned version, presented in the aforementioned paper.

Key words. multilevel Krylov method, GMRES, multigrid, Helmholtz equation, shifted-Laplace preconditioner.

AMS subject classifications. 65F10, 65F50, 65N22, 65N55.

1. Introduction.

Nowadays Krylov subspace methods are the methods of choice for solving large, sparse linear systems of equations

$$Au = b, \qquad A \in \mathbb{C}^{n\times n}. \tag{1.1}$$

If A is Hermitian positive definite, (1.1) is typically solved by the conjugate gradient method (CG). The convergence rate of CG can be bounded in terms of the condition number of A, $\kappa(A)$ [18], which in this case is the ratio of the largest eigenvalue to the smallest one. For an ill-conditioned system, this convergence rate is often too small, so that a preconditioner has to be incorporated.

For general matrices, convergence bounds are somewhat more difficult to establish and do not express a direct connection with the condition number of A. It is, however, a common belief that eigenvalues clustering around a value far from zero improve the convergence. Therefore, without specifically referring to the condition number, it is often sufficient to say that a nonsingular matrix M is a good preconditioner if the eigenvalues of $M^{-1}A$ are more clustered and farther from zero than those of A.

A class of preconditioners which exploits the detail of the spectrum of A can be based on projection methods. These methods accelerate the convergence by removing the components of the residuals corresponding to the smallest eigenvalues during the iteration. One way to achieve this is by deflating a number of the smallest eigenvalues to zero. Nicolaides [16]

*Received October 9, 2007. Accepted June 8, 2009. Published online on September 30, 2009. Recommended by Oliver Ernst.

†TU Berlin, Institut für Mathematik, MA 3-3, Strasse des 17. Juni 136, D-10623 Berlin, Germany. Currently at The University of British Columbia, Department of Earth and Ocean Sciences, 6339 Stores Road, Vancouver, BC V6T 1Z4, Canada (`yerlangga@eos.ubc.ca`).

‡TU Berlin, Institut für Mathematik, MA 3-3, Strasse des 17. Juni 136, D-10623 Berlin, Germany (`nabben@math.tu-berlin.de`).

ETNA
Kent State University
http://etna.math.kent.edu

showed that by adding eigenvectors related to some small eigenvalues, the convergence of CG may be improved. For GMRES, Morgan [14] also shows that by augmenting the Krylov subspace by eigenvectors related to some small eigenvalues, these eigenvectors no longer have components in the residuals, and the convergence bound of GMRES can be made smaller; thus, a faster convergence may be expected; see also a unified discussion on this subject by Eiermann et al. in [3].

A similar approach is proposed in [13], where a matrix resembling deflation of some small eigenvalues is used as a preconditioner. Suppose that the r smallest eigenvalues are to be deflated to zero, and define the deflation matrix as

$$P_D = I - AZE^{-1}Y^T, \qquad E = Y^TAZ, \tag{1.2}$$

where the columns of the full rank matrices $Z, Y \in \mathbb{C}^{n\times r}$ form the basis of the deflation subspaces. The matrix $E \in \mathbb{C}^{r\times r}$ can generally be considered as the *Galerkin* (or more correctly the *Petrov-Galerkin*) matrix associated with A. It can be proved [7, 13, 15] that for a nonsingular A and any full rank Z, Y, the spectrum of P_DA contains r zero eigenvalues. Since the components of the residuals corresponding to the zero eigenvalues do not enter the iteration, the convergence rate is now bounded in terms of the *effective* condition number of P_DA, which for a Hermitian positive definite matrix A and $Y = Z$ is the ratio between the largest eigenvalue and the smallest nonzero eigenvalue of P_DA. Furthermore, it can be shown that a larger r leads to a smaller *effective* condition number [15]. Hence, with a large deflation subspace, convergence can be improved considerably.

A large deflation subspace implies that the matrix E in (1.2) is large. It is then possible that the inversion of E with direct methods becomes impractical, and therefore one has to resort to iterative methods. Related to the computation of E^{-1}, it is shown in [15] that the convergence of CG with P_D deteriorates if E^{-1} is computed inaccurately. We say in this case that P_D is sensitive to an inaccurate computation of E^{-1}. This means that to retain its fast convergence, an iterative method can only be applied to the Galerkin system (i.e., the linear system associated with the Galerkin matrix) with a sufficiently tight termination criterion. Reference [20] discusses this aspect of deflation in detail with extensive numerical tests.

As an alternative to the deflation preconditioner (1.2), another projection-type preconditioner is proposed by the authors in [8]. In this new projection preconditioner, small eigenvalues are shifted *not* towards zero, *but* towards the largest eigenvalue (in magnitude), instead. This leads to eigenvalue clustering in a location far from zero. To discuss this method, we introduce a more general linear system which is equivalent to (1.1), namely

$$\hat{A}\hat{u} = \hat{b}, \tag{1.3}$$

where $\hat{A} = M_1^{-1}AM_2^{-1}$, $\hat{u} = M_2u$, and $\hat{b} = M_1^{-1}b$. Here, M_1 and M_2 are any nonsingular preconditioning matrices. The projection associated with a shift towards the largest eigenvalue of $\hat{A}$ is done via the action of the matrix

$$P_{\hat{N}} = I - \hat{A}Z\hat{E}^{-1}Y^T + \lambda_n Z\hat{E}^{-1}Y^T, \qquad \hat{E} = Y^T\hat{A}Z, \tag{1.4}$$

on the general system (1.3). Here, λ_n is the maximum eigenvalue (in magnitude) of $\hat{A}$. We prefer to use the notation (1.3), because this allows us to consider a more general class of problems involving preconditioners. As discussed in [8], for some problems (e.g., the Poisson and convection-diffusion equation discretized on uniform grids) the preconditioners M_1, M_2 are not actually needed, i.e., it suffices to set $M_1 = M_2 = I$. The role of M_1 and M_2 may become important if, e.g., a nonuniform grid is employed, and in this case the choice $M_1 = I$

ETNA
Kent State University
http://etna.math.kent.edu

and $M_2 = \mathrm{diag}(A)$ is already sufficient. With (1.4), we then solve the left preconditioned system

$$P_{\hat{N}}\hat{A}\hat{u} = P_{\hat{N}}\hat{b}$$

with a Krylov method. Even though its derivation is motivated by projection methods, $P_{\hat{N}}$ is not a projection operator, as $P_{\hat{N}}^2 \neq P_{\hat{N}}$. In this paper we shall call $P_{\hat{N}}$ the shift operator or matrix, instead.

The right preconditioned version of (1.5) can also be defined using the shift matrix

$$Q_{\hat{N}} = I - Z\hat{E}^{-1}Y^T\hat{A} + \lambda_n Z\hat{E}^{-1}Y^T. \tag{1.5}$$

Given (1.5), we then solve the preconditioned system

$$\hat{A}Q_{\hat{N}}\widetilde{u} = \hat{b}, \qquad \hat{u} = Q_{\hat{N}}\widetilde{u}, \tag{1.6}$$

with a Krylov method.

One advantage of (1.4) over (1.2) is that $P_{\hat{N}}$ is insensitive to an inexact inversion of $\hat{E}$. This property allows us to use a large deflation subspace to shift as many small eigenvalues as possible. To obtain an optimal overall computational complexity, the associated Galerkin system is solved by a (inner) Krylov method with a less tight termination criterion. The convergence rate of this inner iteration can be significantly improved if a shift operator similar to (1.4) is also applied to the Galerkin system. The action of this shift operator will require another solve of another Galerkin system, which will be carried out by a Krylov method. If this process is done recursively, a multilevel Krylov method (MK) results. The potential of this multilevel Krylov method is demonstrated in [8].

In this paper we extend the application of the multilevel Krylov method to indefinite linear systems. In particular, we shall focus on the Helmholtz equation. With this application, this paper can be considered as a continuation of our discussion on the multilevel Krylov method, which was presented in [8]. Therefore, for more theoretical results on the method, the readers should consult [8]. Before applying the multilevel Krylov method, the Helmholtz equation is first preconditioned by the shifted Laplacian preconditioner [10]. Since this preconditioner is inverted implicitly by one multigrid iteration, we never have the Galerkin system in an explicit form. We shall demonstrate that with an appropriate approximation to the Galerkin matrix, multigrid-based preconditioners can also be incorporated into the multilevel Krylov framework. We call the resultant method the multilevel Krylov-multigrid (MKMG) method.

In the context of solving the Helmholtz equation, Elman et al. also used Krylov iterations (in their case, GMRES [19]) in a multilevel fashion [4]. However, their approach is basically a multigrid concept specially adapted to the Helmholtz equation. While at the finest and coarsest level, standard smoothers still have good smoothing properties, at the intermediate levels GMRES is employed in place of standard smoothers. Since GMRES does not have a smoothing property, it plays a role in reducing the errors but *not* in smoothing them. A substantial number of GMRES iterations at the intermediate levels, however, is required to achieve a significant reduction of errors.

It is worth mentioning that even though the multilevel Krylov method uses a hierarchy of linear systems similar to multigrid, the way it treats each system and establishes a connection between systems differs from multigrid [9, 8]. In fact, the multilevel Krylov method is not by definition an instance of a multigrid method. With regard to the work in [4], we shall show numerically that the multilevel Krylov method can handle linear systems at the intermediate levels efficiently; i.e., a fast multilevel Krylov convergence can be achieved with only a few Krylov iterations at the intermediate levels.

We organize the paper as follows. In Section 2, we first revisit the Helmholtz equation and our preconditioner of choice, the *shifted Laplace preconditioner*. In Section 3, some relevant theoretical results concerning our multilevel Krylov method are discussed. Some practical implementations are explained in Section 4. Numerical results from 2D Helmholtz problems are presented in Section 5. Finally, in Section 6, we draw some conclusions.

2. The Helmholtz equation and the shifted Laplace preconditioner. The 2D Helmholtz equation for heterogeneous media can be written as

$$\mathcal{A}u := -\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k^2(x,y)\right) u(x,y) = g(x,y), \quad \text{in } \Omega = (0,1)^2, \tag{2.1}$$

where $k(x,y)$ is the wavenumber, and g is the source term. Dirichlet, Neumann, or Sommerfeld (non-reflecting) conditions can be applied at the boundaries $\Gamma \equiv \partial\Omega$; see, e.g., [5]. If a discretization is applied to (2.1) and the boundary conditions, and if the wavenumber is high (as usually encountered in realistic applications), the resultant linear system is large but sparse, and symmetric but indefinite. In most cases, an application of Krylov subspace methods to iteratively solve the linear system results in slow convergence. Standard preconditioners, e.g., ILU-type preconditioners, do not effectively improve the convergence [12].

In [10, 11], for the Helmholtz equation, the shifted Laplacian operator

$$\mathcal{M} := -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - (\alpha - \hat{j}\beta)k^2(x,y), \quad \hat{j} = \sqrt{-1}, \quad \alpha, \beta \in \mathbb{R}, \tag{2.2}$$

is proposed to accelerate the convergence of a Krylov subspace method. The preconditioning matrix M is obtained from discretization of (2.2), with the *same* boundary conditions as for (2.1). The solution u is computed from the (right) preconditioned system

$$AM^{-1}\hat{u} = b, \quad u = M^{-1}\hat{u}, \tag{2.3}$$

where A and M are the Helmholtz and shifted Laplacian matrices respectively.

If (α, β) are well chosen, the eigenvalues of AM^{-1} can be clustered around one. In this paper we shall only consider the pair $(\alpha, \beta) = (1, 0.5)$, which in [10] is shown to lead to an efficient and robust preconditioning operator. Since the convergence of Krylov methods is closely related to the spectrum of the given matrix, we shall give some insight on the spectrum of the preconditioned Helmholtz system (2.3) in the remainder of this section.

The following theorem is a special case of Theorem 3.5 in [22], and holds for the d-dimensional Helmholtz equation.

THEOREM 2.1 ([22]). *Let $A = L + \hat{j}C - K$ and $M = L + \hat{j}C - (\alpha - \beta\hat{j})K$ be the discretization matrices of (2.1) and (2.2), respectively, with L, C, and K the negative Laplacian, the boundary conditions, and the Helmholtz (k^2) term, respectively. Choose $(\alpha, \beta) = (1, 0.5)$.*

(i) For Dirichlet boundary conditions, $C = 0$ and the eigenvalues of $M^{-1}A$ lie on the circle in the complex plane with center $c = (\frac{1}{2}, 0)$ and radius $R = \frac{1}{2}$.

(ii) For Sommerfeld boundary conditions, $C \neq 0$ and the eigenvalues of $M^{-1}A$ are enclosed by the circle with center $c = (\frac{1}{2}, 0)$ and radius $R = \frac{1}{2}$.

Proof. The proof for arbitrary (α, β) can be found in [22]. □

Since $\sigma(M^{-1}A) = \sigma(AM^{-1})$, Theorem 2.1 holds also for AM^{-1}. For the Helmholtz equation with Dirichlet boundary conditions some detailed information about the spectrum, e.g., the largest and smallest eigenvalues, can also be derived. We shall follow the approach used in [11], which was based on a continuous formulation of the problem. The results, however, also hold for the discrete formulation as indicated in [11]. For simplicity, we consider the 1D Helmholtz equation.

ETNA
Kent State University
http://etna.math.kent.edu

At the continuous level, the eigenvalue problem of the preconditioned system can be written as

$$-\left(\frac{d^2}{dx^2} - k^2\right) u = \lambda \left(-\frac{d^2}{dx^2} - (1 - 0.5\hat{j})k^2\right) u, \tag{2.4}$$

with λ the eigenvalue and u now the eigenfunction. By using the ansatz $u = \sin(i\pi x)$, $i \in \mathbb{N}$, from (2.4) we find that

$$\lambda_i = \frac{i^2\pi^2 - k^2}{i^2\pi^2 - (1 - 0.5\hat{j})k^2},$$

with

$$\mathrm{Re}(\lambda_i) = \frac{(i^2\pi^2 - k^2)^2}{(i^2\pi^2 - k^2)^2 + 0.25k^4}, \quad \mathrm{Im}(\lambda_i) = -\frac{0.5(i^2\pi^2 - k^2)k^2}{(i^2\pi^2 - k^2)^2 + 0.25k^4}.$$

From the above relations, observe that $0 < \mathrm{Re}(\lambda_i) < 1$, and therefore

$$\lim_{i\to\infty} \mathrm{Re}(\lambda_i) = \lim_{k\to\infty} \mathrm{Re}(\lambda_i) = 1.$$

The real parts are close to zero if $i^2\pi^2$ are close to k^2. The sign of the imaginary parts depends on the mode i. Also, $\lim_{k\to\infty} \mathrm{Im}(\lambda_i) = 0.5$ and $\lim_{i\to\infty} \mathrm{Im}(\lambda_i) = -0.5$. By eliminating $i^2\pi^2$ in (2.5), we have

$$(\mathrm{Re}(\lambda_i) - 0.5)^2 + \mathrm{Im}(\lambda_i)^2 = 0.25.$$

Thus, λ_i lie on the circle with center $c = (\frac{1}{2}, 0)$ and radius $R = \frac{1}{2}$, as suggested by Theorem 2.1 (i). The largest possible $|\lambda_i|$ is approached as $i \to \infty$, where, in this case, $\mathrm{Re}(\lambda_i) \to 1$ and $\mathrm{Im}(\lambda_i) \to 0$. Thus, $\lim_{i\to\infty} |\lambda_i| = 1$. This result is true for any choice of k.

Suppose now that for some i, $i^2\pi^2 - k^2 = \epsilon$. For $\epsilon \ll k$, $\mathrm{Re}(\lambda_i) = 4\epsilon^2/k^4$ and $\mathrm{Im}(\lambda_i) = -2\epsilon/k^2$, and hence

$$|\lambda_i| = \mathrm{Re}(\lambda_i)^2 + \mathrm{Im}(\lambda_i)^2 = \left(\frac{4\epsilon^2}{k^4}\right)^2 + \left(\frac{2\epsilon}{k^2}\right)^2 \approx \frac{4\epsilon^2}{k^4}.$$

Therefore, while the spectrum of $M^{-1}A$ is more clustered than the spectrum of A, some eigenvalues lie at a distance of order $\mathcal{O}(\epsilon/k^2)$ from zero. Figure 2.1 illustrates this spectral property for a 1D Helmholtz problem with $k = 20$ and 50. Clearly, the largest eigenvalue for both k's is essentially the same and close to one, but the smallest eigenvalue moves towards zero as k increases.

Since small eigenvalues may cause problems to a Krylov method, we discuss in the next section the multilevel Krylov method, used to handle small eigenvalues.

3. Multilevel Krylov method. Consider again the linear system (1.3), where, for our Helmholtz equation, $\hat{A} = AM^{-1}$ and $\hat{b} = b$. Our objective is to shift some small eigenvalues in the spectrum of $\hat{A}$ to a fixed point, such that the new linear system has some more favorable spectrum for convergence acceleration.

As explained in Section 1, one way to achieve this is by using some deflation techniques, in which some small eigenvalues are shifted to zero. Using the multilevel Krylov method, however, we shift these small eigenvalues to the largest eigenvalue, and this shift is done by either (1.4) or (1.5). Note that if we set $\lambda_n = 0$ in (1.4) or (1.5) we recover the deflation preconditioner.

ETNA
Kent State University
http://etna.math.kent.edu

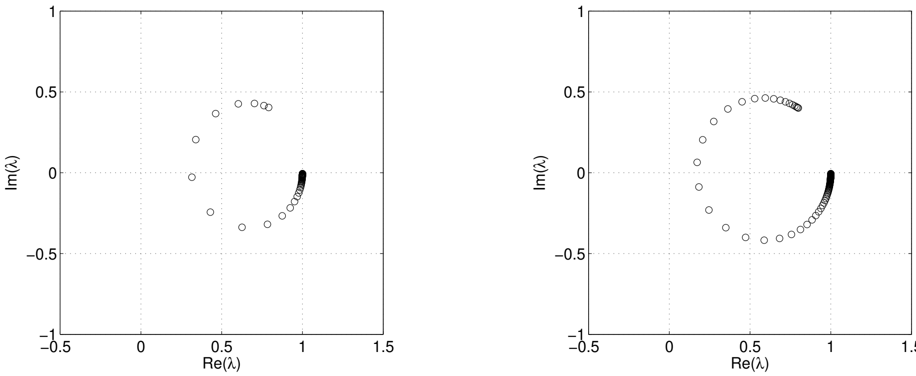


FIGURE 2.1. *Spectrum of a typical 1D Helmholtz problem preconditioned with the shifted Laplacian. The wavenumber k is 20 (left) and 50 (right).*

For (1.5) the following spectral property holds.

THEOREM 3.1. *Suppose that the eigenvalues of $\hat{A}$, $\lambda_1, \dots, \lambda_n \in \sigma(\hat{A}) \subset \mathbb{C}$, are ordered increasingly in magnitude. Let $Z, Y \in \mathbb{C}^{n\times r}$, with $r \ll n$, be full rank matrices*[1] *whose columns are the right and left eigenvectors associated with the r smallest eigenvalues (in magnitude) of $\hat{A}$. Let $Q_{\hat{N}}$ be defined as in* (1.5)*. Then*

$$\sigma(\hat{A}Q_{\hat{N}}) = \{\lambda_n, \dots, \lambda_n, \lambda_{r+1}, \dots, \lambda_n\}.$$

Proof. The proof requires the identity $P_{\hat{D}}\hat{A}Z = 0$, where $P_{\hat{D}} = I - \hat{A}Z\hat{E}^{-1}Y^T$, which is easily verified by a direct computation (see, e.g., [13]), and Theorem 3.5 of [8], which establishes the spectral equivalence $\sigma(P_{\hat{N}}\hat{A}) = \sigma(\hat{A}Q_{\hat{N}})$, with $P_{\hat{N}}$ as in (1.4).

First, for $i = 1, \dots, r$, we have $P_{\hat{N}}\hat{A}Z = P_{\hat{D}}\hat{A}Z + \lambda_n Z\hat{E}^{-1}Y^T\hat{A}Z = \lambda_n$. Next, for $r + 1 \le i \le n$, we have that

$$P_{\hat{N}}\hat{A}z_i = \hat{A}z_i - \hat{A}Z\hat{E}^{-1}Y^T\hat{A}z_i + \lambda_n Z\hat{E}^{-1}Y^T\hat{A}z_i = \lambda_i z_i,$$

due to orthogonality of eigenvectors. Finally, by using Theorem 3.5 of [8], $\sigma(P_{\hat{N}}\hat{A}) = \{\lambda_n, \dots, \lambda_n, \lambda_{r+1}, \dots, \lambda_n\} = \sigma(\hat{A}Q_{\hat{N}})$. ☐

Thus, after applying $Q_{\hat{N}}$ to $\hat{A}$, r eigenvalues are no longer small and have been shifted to λ_n. The smallest eigenvalue (in magnitude) is now λ_{r+1}, and the rest of the spectrum remains untouched. If λ_{r+1} is of the same order of magnitude as λ_n, a Krylov subspace method is expected to converge faster.

The computation of eigenvectors, however, is very expensive for large linear systems. Furthermore, as eigenvectors, Z and Y are dense.

In the following we will consider the deflation and the shift operator under any full rank Z and Y. We start with the deflation operator. Since

$$\hat{A}Q_{\hat{D}}Z = \hat{A}Z - \hat{A}Z\hat{E}^{-1}Y^T\hat{A}Z = \hat{A}Z - \hat{A}Z = 0,$$

we obtain

$$\sigma(\hat{A}Q_{\hat{D}}) := \{0, \dots, 0, \mu_{r+1}, \dots, \mu_n\}.$$

[1] While the theory only requires $r < n$, like in, e.g., multigrid, this condition emphasizes the importance of the sufficiently small deflation subspace to make the overall method practical.

Thus, $\hat{A}Q_{\hat{D}}$ has r zero eigenvalues for arbitrary matrices Z and Y. In contrast to Theorem 3.1, the remaining eigenvalues $\mu_{r+1}, \ldots, \mu_n$ are not, in general, eigenvalues of $\hat{A}$. Thus, some of the eigenvalues of $\hat{A}$ are shifted to zero, some of them are shifted to the μ_i.

The following theorem establishes a spectral relationship between deflation and the shift operator with any full rank Z and Y.

THEOREM 3.2. *Let $Z, Y \in \mathbb{C}^{n\times r}$ be of rank r, $\hat{A}$ be nonsingular, and let $Q_{\hat{D}} = I - Z\hat{E}^{-1}Y^T\hat{A}$. If $Q_{\hat{N}}$ is defined as in (1.5), and Z, Y are such that*

$$\sigma(\hat{A}Q_{\hat{D}}) := \{0, \ldots, 0, \mu_{r+1}, \ldots, \mu_n\},$$

then

$$\sigma(\hat{A}Q_{\hat{N}}) = \{\lambda_n, \ldots, \lambda_n, \mu_{r+1}, \ldots, \mu_n\}.$$

Proof. Combine Theorems 3.4 and 3.5 in [8]. Note, that the columns of Z are the left eigenvectors of $\hat{A}Q_{\hat{D}}$ corresponding to the eigenvalue equal to zero. Then, we obtain

$$\hat{A}Q_{\hat{N}}Z = \lambda_n Z.$$

Theorem 3.5 in [8] gives

$$\sigma(\hat{A}Q_{\hat{N}}) = \sigma(P_{\hat{N}}\hat{A}).$$

Now, if

$$\hat{A}Q_{\hat{D}}x_i = \mu_i x_i,$$

for $r + 1 \leq i \leq n$ and some eigenvectors x_i, we easily obtain

$$P_{\hat{N}}\hat{A}(Q_{\hat{D}}x_i) = \mu_i(Q_{\hat{D}}x_i). \quad \square$$

In the above theorem, $Q_{\hat{D}}$ is the right preconditioning version of the deflation preconditioner. The action of $Q_{\hat{D}}$ on $\hat{A}$ shifts r eigenvalues of $\hat{A}$ to zero. With $Q_{\hat{N}}$, these zero eigenvalues in the spectrum of $\hat{A}Q_{\hat{D}}$ become λ_n in the spectrum of $\hat{A}Q_{\hat{N}}$. Under the arbitrariness of Z and Y, the rest of the eigenvalues is also shifted to μ_i, $i = r+1, \ldots, n$, but these eigenvalues are the same for both $\hat{A}Q_{\hat{D}}$ and $\hat{A}Q_{\hat{N}}$. Their exact values depend on the choice of Z and Y. In particular, $\mu_n \neq \lambda_n$. However, for any μ_n and λ_n, there exists a constant $\omega \in \mathbb{C}$ such that $\mu_n = \omega\lambda_n$. The constant ω is called the *shift scaling factor*. A shift correction can be incorporated in (1.5) by replacing λ_n with $\omega\lambda_n$. With this scaling, the spectrum of $\hat{A}Q_{\hat{D}}$ and $\hat{A}Q_{\hat{N}}$ differ only in the multiple eigenvalue zero and in λ_n. If the convergence is only measured by the ratio of the largest and smallest nonzero eigenvalues, which can be true in the case of symmetric positive definite matrices, a very similar convergence for both methods can be expected.

To construct $Q_{\hat{N}}$, we need two components: the largest eigenvalue λ_n and the rectangular matrices Z and Y.

For λ_n, we note that in general its computation is expensive. As advocated in [8], it is sufficient to use an approximation to λ_n. For example, Gerschgorin's theorem [23] can provide a good approximation to λ_n. For our Helmholtz problems, however, we shall use results in Section 2, i.e., for AM^{-1}, $\mathrm{Re}(\lambda_n(AM^{-1})) = |\lambda_n(AM^{-1})| = 1$. Thus, we set $\lambda_n = 1$ in $Q_{\hat{N}}$.

For Z and Y, we require that these matrices are sparse to avoid excessive memory requirements. Next, we note that $Z : \mathbb{C}^r \mapsto \mathbb{C}^n$, and $Y^T : \mathbb{C}^n \mapsto \mathbb{C}^r$, $r \ll n$, are linear

ETNA
Kent State University
http://etna.math.kent.edu

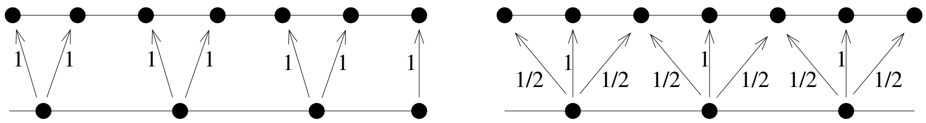


FIGURE 3.1. *Interpolation in 1D: piece-wise interpolation (left) and linear interpolation (right).*

maps similar to prolongation and restriction operators in multigrid. In multigrid, the matrix $\hat{E} = Y^T \hat{A} Z$ is called the Galerkin coarse-grid approximation of $\hat{A}$. Since they are sparse, these multigrid intergrid transfer operators are good candidates for the deflation matrices. In [8], we used the piece-wise constant (zeroth-order) interpolation for Z and set $Y = Z$. This choice is not common in multigrid, but leads to an efficient multilevel Krylov method. Since at the present time we do not have detailed theoretical criteria for the choice of Z and Y, we investigate these two possible options by looking at spectral properties and numerical experiments based on a simple 1D problem. In this case, all eigenvalues can be computed easily and the matrices M and $\hat{E}$ can be inverted exactly. In a 1D finite difference setting, the piece-wise constant interpolation and multigrid prolongation (in this case, linear interpolation) are illustrated in Figure 3.1.

We first consider the spectra of $AM^{-1}Q_{\hat{N}}$, with Z the piece-wise constant interpolation matrix and $Y = Z$. Following the aforementioned discussion, we set $\lambda_n = 1$. Furthermore, we set $\omega = 1$. The spectra are shown in Figure 3.2. Compared to Figure 2.1, Figure 3.2 clearly shows that small eigenvalues near the origin are no longer present. The action of $Q_{\hat{N}}$, however, changes the whole spectrum; i.e., $\lambda_i, i = r+1, \ldots, n$, are also shifted. Nevertheless, this eigenvalue distribution is more favorable for a Krylov method as it is now clustered far from the origin. Figure 3.2 also indicates that increasing the deflation vectors (increasing r) improves the clustering. For $k = 20$ and $r = n/2 = 50$, the eigenvalues of $AM^{-1}Q_{\hat{N}}$ are now clustered compactly around one; cf. Figure 3.2 (c). For $k = 50$, a very similar eigenvalue clustering with $k = 20$ is observed if we set $r = n/2$; in this case, $r = 125$.

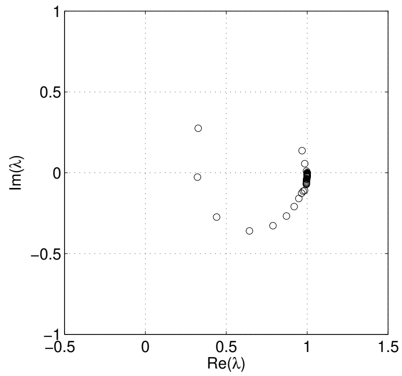
Next, we consider the spectra of $\hat{A}Q_{\hat{N}}$ with Z representing the *linear* interpolation. Similarly, we set $Y = Z$, $\lambda_n = 1$, and $\omega = 1$. The spectra for $k = 20$ and 50 are shown in Figure 3.3 for $r = n/2$. Compared to Figure 3.2 (c) and (d), the spectra are clustered around one as well. Thus, either the piece-wise constant interpolation or the linear interpolation lead to spectrally similar systems, and hence we can expect very similar convergence property for both choices.

To see how the spectral properties translate to the convergence of a Krylov method, we perform numerical experiments based on the 1D Helmholtz problem with constant wavenumber. Again, M and $\hat{E}$ are inverted exactly. We apply GMRES to (1.6) and measure the number of iterations needed to reduce the relative residual by six orders of magnitude. Convergence results are shown in Table 3.1, with $Z \in \mathbb{C}^{n\times r}$ based on either piece-wise constant interpolation or linear interpolation, and with $Y = Z$. In all cases, $r = n/2$, where $n = 1/h$ and h is the mesh size. The mesh size h decreases when the wavenumber k increases, so that the solutions are solved on grids equivalent to 30, 15, and 8 gridpoints per wavelength[2].

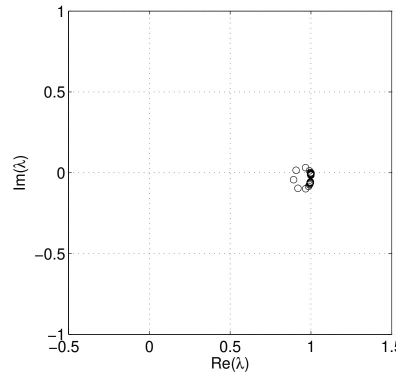
For the case without a "two-level" Krylov step (without $Q_{\hat{N}}$), denoted by "standard", we observe convergence, which depends linearly on the wavenumber k. The convergence becomes less dependent on k if $Q_{\hat{N}}$ is incorporated. In particular, if Z is the linear interpolation

[2] The use of 8 gridpoints per wavelength on the finest grid is, however, too coarse for a second-order finite-difference scheme used in this experiment, as the pollution error becomes dominant, see, e.g., [2, 1]. For a second-order scheme, the rule of thumb is to use at least 12 gridpoints per wavelength. For this reason, this is the only example where 8 gridpoints per wavelength are used.

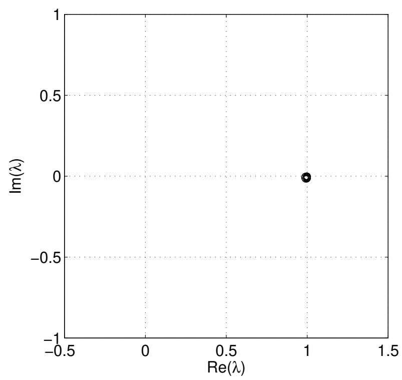
ETNA
Kent State University
http://etna.math.kent.edu



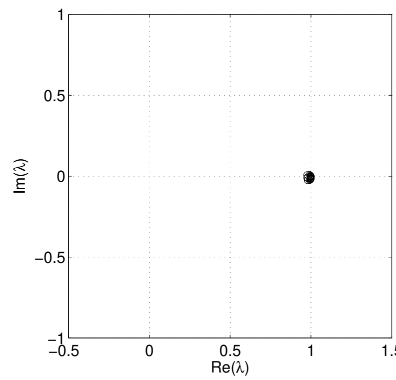
(a) $k = 20$, $r = 5$



(b) $k = 20$, $r = 20$



(c) $k = 20$, $r = n/2 = 50$



(d) $k = 50$, $r = n/2 = 125$

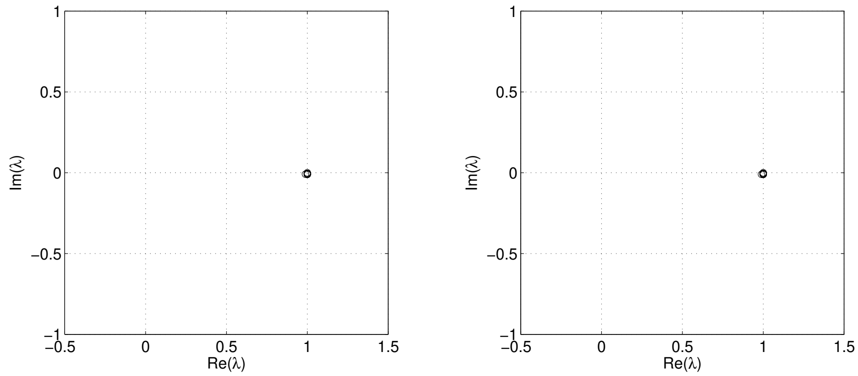
FIGURE 3.2. *Spectra of a preconditioned 1D Helmholtz problem, $k = 20$ and 50. The number of grid points for each k is $n = 100$ and 250, respectively. Z is obtained from the piece-wise constant interpolation.*

matrix, the convergence can be made almost independent of k, unless the grid is too coarse. The convergence deterioration is worse in the case of the piece-wise constant interpolation.

TABLE 3.1
Number of preconditioned GMRES iterations for a 1D Helmholtz problem. Equidistant grids equivalent to 30/15/8 gridpoints per wavelength are used, and $r = n/2$. The relative residual is reduced by six orders of magnitude.

	$k = 20$	$k = 50$	$k = 100$	$k = 200$	$k = 500$
Standard	14/15/15	24/25/26	39/40/42	65/68/78	142/146/157
$Q_{\hat{N}}$, piece-wise constant	4/5/7	4/6/10	5/7/14	6/10/20	7/15/37
$Q_{\hat{N}}$, linear interpolation	3/4/5	3/4/7	3/4/8	3/5/10	3/5/12

4. Multilevel Krylov method with approximate Galerkin systems. In Section 3 we saw that the convergence of GMRES preconditioned by M and $Q_{\hat{N}}$ can be made independent of k, provided that M and $\hat{E}$ are explicitly inverted. In higher dimensions (2D or 3D) this approach is no longer practical. Particular to our preconditioner, the inverse of M is approximately computed by one multigrid iteration. Hence, M^{-1} is not explicitly available.



(a) $k = 20$, $r = n/2 = 50$ (b) $k = 50$, $r = n/2 = 125$

FIGURE 3.3. *Spectra of a preconditioned 1D Helmholtz problem, $k = 20$ and 50. The number of grid points for each k is $n = 100$ and 250, respectively. Z is obtained from the linear interpolation.*

First consider the *two-level* Krylov method. With any full rank $Y, Z \in \mathbb{C}^{n\times r}$, the (right) preconditioning step of a Krylov method can be written as

$$\begin{aligned} w &= M^{-1}Q_{\hat{N}}v = M^{-1}(I - Z\hat{E}^{-1}Y^TAM^{-1} + \omega\lambda_n Z\hat{E}^{-1}Y^T)v \\ &= M^{-1}(v - Z\hat{E}^{-1}Y^Tv'), \end{aligned} \tag{4.1}$$

where

$$v' = (AM^{-1} - \omega\lambda_n I)v \quad \text{and} \quad \hat{E} = Y^T\hat{A}Z. \tag{4.2}$$

In GMRES, the vector v is the Arnoldi vector, which in turn gives v' via (4.2). The vector $v' \in \mathbb{C}^n$ is then restricted to $\mathbb{C}^r$ by Y^T as in (4.1), namely

$$v'_R := Y^Tv'. \tag{4.3}$$

With v'_R, the Galerkin problem in (4.1) now reads

$$v_R := \hat{E}^{-1}v'_R \iff v'_R = \hat{E}v_R. \tag{4.4}$$

It is important to note here that the operator $Q_{\hat{N}}$ remains effective for convergence acceleration under inexact inversion of $\hat{E}$; see [8]. Therefore, a Krylov method can be used to approximately solve (4.4). In general, the accuracy of the solution produced by a Krylov method depends on the termination criteria. For ill-conditioned $\hat{E}$ it is possible that many Krylov iterations are needed for a substantial reduction of residuals/errors. To obtain a large reduction of residuals/errors within a small number of Krylov iterations, shifting similar to (1.5) can also be applied to the Galerkin system. This shift will require solving another but smaller Galerkin system. A recursive application of shifting and iterative Galerkin solution leads to the *multilevel Krylov method*. An algorithm of the multilevel Krylov method is presented in [8].

With respect to the Galerkin solution, one immediate complication arises. Since M^{-1} is only available implicitly (via one multigrid iteration), the Galerkin matrix $\hat{E}$ is not explicitly available. Aside from computational complexity to do inversion, forming $\hat{E}$ explicitly is also not advisable because of M^{-1}, which implies that $\hat{E}$ is dense. To set up a Galerkin system

ETNA
Kent State University
http://etna.math.kent.edu

which is conducive to the multilevel Krylov method, we propose the following approximation. We approximate the inverse M^{-1} by $Z(Y^TMZ)^{-1}Y^T$. This leads to

$$\begin{aligned}\hat{E} &:= Y^T\hat{A}Z = Y^TAM^{-1}Z \\ &\approx Y^TAZ(Y^TMZ)^{-1}Y^TZ = A_HM_H^{-1}B_H =: \hat{A}_H,\end{aligned} \tag{4.5}$$

where the products $A_H := Y^TAZ$, $M_H := Y^TMZ$, and $B_H := Y^TZ$ are the Galerkin matrices associated with A, M, and I respectively.

With the approximation (4.5), the Galerkin system (4.4) can now be written as

$$v_R' = A_HM_H^{-1}B_Hv_R, \tag{4.6}$$

where the solution vector v_R is obtained by using a Krylov subspace method. A fast convergence of a Krylov method for (4.6) can be obtained by applying a projection on (4.6). This immediately defines our multilevel Krylov method.

To construct a multilevel Krylov algorithm, we shall use notations which incorporate level identification. For example, for the two-level Krylov method discussed above, A, M and Z are now denoted by $A^{(1)}$, $M^{(1)}$ and $Z^{(1,2)}$, respectively. With these notations, we have

$$\hat{A}^{(2)} = A^{(2)}M^{(2)^{-1}}B^{(2)},$$

where $A^{(2)} = Y^{(1,2)^T}A^{(1)}Z^{(1,2)}$, $M^{(2)} = Y^{(1,2)^T}M^{(1)}Z^{(1,2)}$, and $B^{(2)} = Y^{(1,2)^T}I^{(1)}Z^{(1,2)}$.

The matrix $\hat{A}^{(2)}$ is the second level ($j = 2$) Galerkin matrix associated with $\hat{A}^{(1)} = A^{(1)}M^{(1)^{-1}}$, etc. If $\hat{A}^{(2)}$ is small enough, the Galerkin system

$$A^{(2)}M^{(2)^{-1}}B^{(2)}v_R^{(2)} = (v_R')^{(2)}$$

can be solved exactly. Otherwise, we shall use a Krylov method to approximately solve it. For the latter, we define the shift operator

$$Q_{\hat{N}}^{(2)} = I - Z^{(2,3)}\hat{A}^{(3)^{-1}}Y^{(2,3)^T}\hat{A}^{(2)} + \omega^{(2)}\lambda_n^{(2)}Z^{(2,3)}\hat{A}^{(3)^{-1}}Y^{(2,3)^T},$$

with $\hat{A}^{(3)} = Y^{(2,3)^T}\hat{A}^{(2)}Z^{(2,3)}$, and solve the linear system

$$A^{(2)}M^{(2)^{-1}}B^{(2)}Q_{\hat{N}}^{(2)}\widetilde{v}_R^{(2)} = (v_R')^{(2)},$$

where $v_R^{(2)} = Q_{\hat{N}}^{(2)}\widetilde{v}_R^{(2)}$, by a Krylov subspace method. In this case, the shift operator $Q_{\hat{N}}$ makes the system better conditioned, improves the convergence on the second level, and hence reduces iterations needed to solve the Galerkin system. The multilevel Krylov method is obtained if the same argument is applied to $\hat{A}^{(3)}$.

Suppose that m levels are used, where at level $m-1$ the associated Galerkin problem is sufficiently small to be solved exactly. The multilevel Krylov method can be written in an algorithm as follows.

Algorithm 1. Multilevel Krylov method with approximate Galerkin matrices

Initialization:

For $j = 1$, set $A^{(1)} := A$, $M^{(1)} := M$, $B^{(1)} := I$, construct $Z^{(1,2)}$, and choose $\lambda_n^{(1)}$ and $\omega^{(1)}$. With this information, $\hat{A}^{(1)} = A^{(1)} M^{(1)^{-1}}$ and $Q_{\hat{N}}^{(1)} = Q_{\hat{N}}$ are in principle determined.

For $j = 2, \ldots, m$, choose $Z^{(j-1,j)}$ and $Y^{(j-1,j)}$, and compute

$$A^{(j)} = Y^{(j-1,j)^T} A^{(j-1)} Z^{(j-1,j)},$$
$$M^{(j)} = Y^{(j-1,j)^T} M^{(j-1)} Z^{(j-1,j)},$$
$$B^{(j)} = Y^{(j-1,j)^T} B^{(j-1)} Z^{(j-1,j)},$$

which define

$$\hat{A}^{(j)} = A^{(j)} M^{(j)^{-1}} B^{(j)}.$$

For $j = 2, \ldots, m-1$, set $\omega^{(j)}$ and $\lambda_n^{(j)}$, and define

$$Q_{\hat{N}}^{(j)} = I - Z^{(j-1,j)} \hat{A}^{(j)^{-1}} Y^{(j-1,j)^T} \left(\hat{A}^{(j-1)} - \omega^{(j)} \lambda_n^{(j)} I\right).$$

Iteration phase:

$j = 1$
Solve $A^{(1)} M^{(1)^{-1}} \widetilde{u}^{(1)} = b$, $u^{(1)} = M^{(1)^{-1}} \widetilde{u}^{(1)}$ with Krylov iterations by computing
    $v_M^{(1)} = M^{(1)^{-1}} v^{(1)}$
    $s^{(1)} = A^{(1)} v_M^{(1)}$
    $t^{(1)} = s^{(1)} - \omega^{(1)} \lambda_n^{(1)} v^{(1)}$
    Restriction: $(v_R')^{(2)} = Y^{(1,2)^T} t^{(1)}$
    If $j = m$
        $v_R^{(m)} = \hat{A}^{(m)^{-1}} (v_R')^{(m)}$
    else
        $j = 2$
        Solve $A^{(2)} M^{(2)^{-1}} B^{(2)} v_R^{(2)} = (v_R')^{(2)}$ with Krylov iterations by computing
            $v_M^{(2)} = M^{(1)^{-1}} B^{(2)} v^{(2)}$
            $s^{(2)} = A^{(2)} v_M^{(2)}$
            $t^{(2)} = s^{(2)} - \omega^{(2)} \lambda_n^{(2)} v^{(2)}$
            Restriction: $(v_R')^{(3)} = Y^{(2,3)^T} t^{(2)}$
            If $j = m$
                $v_R^{(m)} = \hat{A}^{(m))^{-1}} (v_R')^{(m)}$
            else
                $j = 3$
                Solve $A^{(3)} M^{(3)^{-1}} B^{(3)} v_R^{(3)} = (v_R')^{(3)}$
                    $\ldots$
            Interpolation: $v_I^{(2)} = Z^{(2,3)} v_R^{(3)}$
            $q^{(2)} = v^{(2)} - v_I^{(2)}$
            $w^{(2)} = M^{(2)^{-1}} B^{(2)} q^{(2)}$
            $p^{(2)} = A^{(2)} w^{(2)}$
    Interpolation: $v_I^{(1)} = Z^{(1,2)} v_R^{(2)}$
    $q^{(1)} = v^{(1)} - v_I^{(1)}$
    $w^{(1)} = M^{(1)^{-1}} q^{(1)}$
    $p^{(1)} = A^{(1)} w^{(1)}$

REMARK 4.1. In solving the Galerkin problems by a Krylov subspace method, a zero initial guess is always used. With this choice, the initial residual does not have to be computed explicitly because it is equal to the right-hand side vector of the Galerkin system. Hence, we can save one vector multiplication with $A^{(j)}M^{(j)^{-1}}B^{(j)}$.

REMARK 4.2. At every level j, we require an estimate to $\lambda_n^{(j)}$. Our numerical results reveal that with $\omega^{(j)} = 1$, $j = 1, \ldots, n-1$, taking $\lambda_n^{(j)} = 1$ leads to a good method.

5. Multilevel Krylov-multigrid method.

In Algorithm 1, at each level two preconditioner solutions related to $M^{(j)}$ are required to compute $v_M^{(j)}$ and $w^{(j)}$. At the level $j = 1$, this solution is approximately determined by one multigrid iteration. Even though the resultant error reduction factor ρ is not that of the typical text-book multigrid convergence (in this case, $\rho = 0.6$), this choice leads to an effective preconditioner for convergence acceleration of Krylov subspace methods for the Helmholtz equation [12]. Since the size of $M^{(j)}$, $1 < j < m$, may also be large, we shall use one multigrid iteration to approximately compute $M^{(j)^{-1}}$.

A multigrid method consists of a recursive application of presmoothing, restriction, coarse-grid correction, interpolation and defect correction, and postsmoothing. Both pre- and postsmoothing are carried out by basic iterative methods, e.g., damped Jacobi or Gauss-Seidel, which smooth the error. The smooth errors are then restricted to the coarse-grid subspace, where a coarse-grid system is solved to further correct the errors. This correction is then added to the error in the fine-grid subspace, after an interpolation process. For further reading on multigrid, we refer to, e.g, [21]. What is important to us is the multigrid restriction and interpolation process, and the coarse-grid correction step.

Assume that a sequence of fine and coarse grids Ω^j, $j = 1, \ldots, m$, $\Omega^1 \supset \Omega^2 \cdots \supset \Omega^m$ are given. The multigrid transfer operators between two grids Ω^j and Ω^{j+1}, denoted by

$$I_j^{j+1} : \mathcal{G}(\Omega^j) \mapsto \mathcal{G}(\Omega^{j+1}), \qquad I_{j+1}^j : \mathcal{G}(\Omega^{j+1}) \mapsto \mathcal{G}(\Omega^j), \tag{5.1}$$

are associated with the restriction and interpolation (or prolongation) process, respectively, and are given as well. For the Galerkin coarse-grid correction, the coarse-grid system is associated with the Galerkin coarse-grid matrix defined as

$$M_{MG}^{(j)} = I_j^{j+1} M_{MG}^{(j+1)} I_{j+1}^j. \tag{5.2}$$

The processes (5.1) are algebraically the same as what Z and Y^T, respectively, do in the multilevel Krylov method, and (5.2) is similar to E. In multigrid, however, the matrices I_j^{j+1} and I_{j+1}^j should represent a sufficiently accurate interpolation and, respectively, restriction of smooth functions. Since the multilevel Krylov method does not necessarily require this criterion, the matrices I_j^{j+1} and I_{j+1}^j are in general not the same as Z and Y^T, respectively. This implies that, in general, $M^{(j)} \neq M_{MG}^{(j)}$, $j > 1$. But it is not a problem for the multilevel Krylov method to have $Z = I_j^{j+1}$ and $Y^T = I_{j+1}^j$, as the conditions in Theorem 3.2 are met. In this case, $M^{(j)} = M_{MG}^{(j)}$.

We comment on the choice $Z = I_j^{j+1}$ and $Y^T = I_{j+1}^j$. First, as shown for the 1D example in Section 3, with Z based on multigrid linear interpolation the convergence of the *two*-level Krylov method is faster than with the piece-wise constant interpolation. We can expect that this convergence property also holds for the *multi*-level Krylov method. Secondly, since now $M^{(j)} = M_{MG}^{(j)}$, both the multilevel Krylov method and the multigrid steps for the preconditioner solves use the same components. This avoids additional storage for multigrid components. Furthermore, all coarse-grid information used by the multilevel Krylov and multigrid parts are computed only once during the initialization phase of the multilevel Krylov method. This will save the cost of the initialization phase.

ETNA
Kent State University
http://etna.math.kent.edu

A multigrid algorithm for solving, e.g., $v_M^{(j)} = M^{(j)^{-1}} v_B^{(j)}$ in Algorithm 1, with $v_B^{(j)} = B^{(j)} v^{(j)}$, can be written as follows.

Algorithm 2. Multigrid with $(j - m + 1)$ levels

Given $v_{M,\ell}^{(j)}$
Presmoothing: $v_{M,\ell+1/3}^{(j)} = \text{smooth}(M^{(j)}, v_{M,\ell}^{(j)}, v_B^{(j)})$
$r^{(j)} = v_B^{(j)} - M^{(j)} v_{M,\ell+1/3}^{(j)}$
Restriction: $r^{(j+1)} = Y^{(j,j+1)^T} r^{(j)}$
Coarse-grid problem:
 if $j = m$ solve $e^{(m)} = M^{(m)^{-1}} r^{(m)}$
 else
  ...
 endif
Prolongation: $d^{(j)} = Z^{(j,j+1)} e^{(j+1)}$
Defect correction: $v_{M,\ell+2/3}^{(j)} = v_{M,\ell+1/3}^{(j)} + d^{(j)}$
Post-smoothing: $v_{M,\ell+1}^{(j)} = \text{smooth}(M^{(j)}, v_{M,\ell+2/3}^{(j)}, v_B^{(j)})$

Incorporating Algorithm 2 in Algorithm 1, the multilevel Krylov-multigrid method (MKMG) results. Note that in Algorithm 2, the finest multigrid level is always the same as the current level in the multilevel Krylov step. Hence, for the action of $Q_{\hat{N}}^{(j)}$ done at level $j = J < m$, multigrid with $J - m$ grid levels is used to approximate the action of preconditioner $M^{(J)}$.

Figure 5.1 illustrates one MKMG cycle with $m = 5$ levels. The white circles indicate the pre- and postsmoothing process in multigrid applied to M, while the black circles correspond to the multilevel steps. In this figure, the multigrid step is shown with V-cycle, but this can in principle be replaced by other multigrid cycles. At the level j of the multilevel Krylov method, multigrid with $m - j$ levels is called to approximately invert $M^{(j)}$ with the corresponding coarse-grid matrices $M^{(j+1)}, \ldots, M^{(m)}$. Once the multilevel Krylov method reaches the level $j = m - 1$, the Galerkin problem at level $j = m$ is solved exactly.

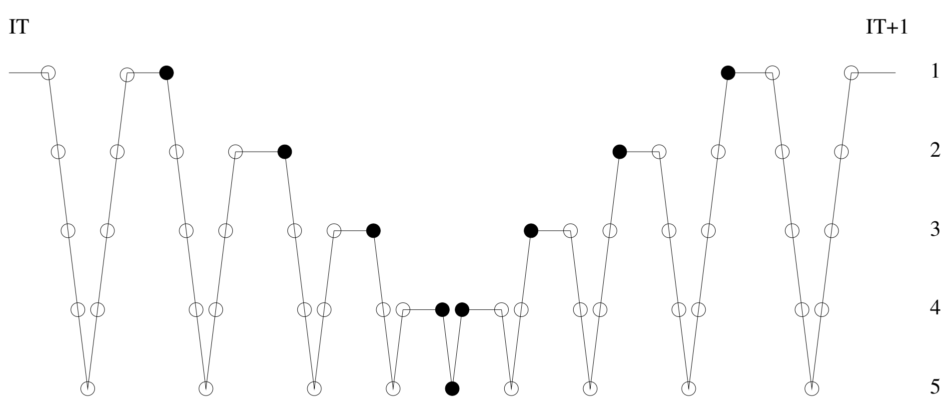


FIGURE 5.1. *Multilevel Krylov-multigrid cycle with $m = 5$. "●": multilevel Krylov step; "○": multigrid step.*

6. Numerical experiments. In this section we present convergence results for the 1D and 2D Helmholtz equation. We compare performance of the multilevel Krylov-multigrid method (denoted by MKMG) with that of Krylov preconditioned by shifted Laplacian (de-

ETNA
Kent State University
http://etna.math.kent.edu

noted by MG). For both methods, we employ one multigrid iteration to invert the shifted Laplacian, with F-cycle and one pre- and postsmoothing. Following [10], Jacobi with underrelaxation ($\omega_R = 0.5$) is used as a smoother. This value was found via the Local Fourier Analysis (LFA), and appeared to be optimal for problems considered there for a wide range of wavenumbers. The coarsest level for both MKMG and MG consists of only one interior grid point.

At each level $j > 1$ of MKMG, GMRES [17] is applied to the preconditioned Galerkin system. Since in this case the preconditioners are not fixed, a flexible version of GMRES, called FGMRES, is employed. For $j = 1$, the finest level, FGMRES is used for MKMG and MG. Convergence for MKMG and MG is declared if the initial relative residual is reduced by six orders of magnitude.

In principle it is not necessary to use the same number of FGMRES iterations at each level. The notation MKMG(6,2,2), for instance, indicates that 6 FGMRES iterations are employed at level $j = 2$, 2 at level $j = 3$ and 2 at level $j = 4, \ldots, m-1$. At level $j = m$ the coarse-grid problem is solved exactly. As observed in [8], it is the accuracy of solving the Galerkin system at the second level which is of importance.

6.1. 1D Helmholtz. In this section, we use the same problem as in Section 3. Convergence results are shown in Tables 6.1–6.3.

Results in Tables 6.1–6.3 suggest that the convergence of MKMG is only mildly dependent on the grid size h. Furthermore, the number of iterations to reach convergence increases only mildly with an increase in the wavenumber k. These results are worse than the ideal situation where the Galerkin system at the second level is solved exactly; cf. Table 3.1. The multilevel Krylov step in MKMG, however, improves the convergence of MG (shown in Table 6.1).

TABLE 6.1
Number of GMRES iterations for 1D Helmholtz problems with constant wave number. g/w stands for "# of grid points per wavelength". Multilevel Krylov method with MKMG(6,2,2). MG is shown in parentheses.

g/w	$k=20$	$k=50$	$k=100$	$k=200$	$k=500$
15	11 (19)	11 (29)	11 (43)	15 (66)	25 (138)
30	9 (18)	11 (28)	12 (42)	14 (68)	22 (136)
60	9 (18)	9 (28)	12 (43)	12 (68)	19 (141)

TABLE 6.2
Number of GMRES iterations for 1D Helmholtz problems with constant wave number. g/w stands for "# of grid points per wavelength". Multilevel Krylov method with MKMG(8,2,2) and MKMG(8,2,1) (in parentheses).

g/w	$k=20$	$k=50$	$k=100$	$k=200$	$k=500$
15	11 (11)	15 (16)	19 (18)	22 (21)	33 (33)
30	10 (10)	13 (13)	13 (13)	15 (15)	20 (20)
60	9 (9)	13 (13)	10 (12)	14 (14)	17 (18)

The significance of the number of iterations at the second level in MKMG can also be seen in Tables 6.1–6.3. While the convergence for MKMG(8,2,2) is slightly better than MKMG(6,2,2), no significant improvement is gained with MKMG(6,4,2) (Table 6.3). We also observe that *one* FGMRES iteration at level $j \geq 4$ is sufficient for fast convergence; see figures in parentheses in Table 6.2.

Our last convergence results for the 1D Helmholtz test problem are associated with the quality of the approximate solution produced by FGMRES at convergence. Here we com-

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 6.3
Number of GMRES iterations for 1D Helmholtz problems with constant wave number. g/w stands for "# of grid points per wavelength". Multilevel Krylov method with MKMG(6,4,2). The ℓ_2 norm of errors are shown in parentheses.

g/w	$k = 20$	$k = 50$	$k = 100$	$k = 200$	$k = 500$
15	11 (2.42E–8)	15 (6.87E–8)	20 (6.68E–8)	23 (1.29E–7)	36 (4.80E–8)
30	10 (6.35E–8)	13 (4.83E–8)	13 (3.39E–8)	14 (1.02E–7)	19 (1.27E–7)
60	9 (1.17E–7)	16 (1.24E–7)	12 (6.78E–8)	16 (1.16E–6)	19 (4.39E–7)

pute the error between the approximate solution of MKMG at convergence and the solution obtained from a sparse direct method. The ℓ_2 norms of the error are shown in parentheses in Table 6.3. For all cases, the ℓ_2 norms of the error fall below 10^{-5}.

6.2. 2D Helmholtz. In this section, 2D Helmholtz problems in a square domain with constant wavenumbers are presented. At the boundaries, the first-order approximation to the Sommerfeld (non-reflecting) condition due to Engquist and Majda [6] is imposed. We consider problems where a source is generated in the middle of the domain.

Following the 1D case, the deflation subspace Z is chosen to be the same as the interpolation matrix in multigrid. For 2D cases, however, care is needed in constructing the interpolation matrix Z. Consider a set of fine grid points defined by

$$\Omega_h := \{(x,y) \mid x = x_{i_x} = i_x h,\ y = y_{i_y} = i_y h,\ i_x = 1,\ldots,N_{x,h},\ i_y = 1,\ldots,N_{y,h}\},$$

associated with the grid points on level $j = 1$. The set of grid points Ω_H corresponding to the coarse-grid level $j = 2$ is determined as follows. We assume that $(x_1, y_1) \in \Omega_H$ coincides with $(x_1, y_1) \in \Omega_h$, as illustrated in Figure 6.1 (left). Starting from this point, the complete set of coarse-grid points is then selected according to the standard multigrid coarsening, i.e., by doubling the mesh size. This results in the coarse grid, for $H = 2h$,

$$\begin{aligned}\Omega_H := \{(x,y) \mid x = x_{i_x} = (2i_x - 1)h,\, y = y_{i_y} = (2i_y - 1)h,& \\ i_x = 1,\ldots,N_{x,H},\ i_y = 1,\ldots,N_{y,H}\}&.\end{aligned}$$

As shown in [12], this coarsening strategy leads to a good multigrid method for the shifted Laplacian preconditioner. Moreover, from a multilevel Krylov method point of view, this coarsening strategy results in larger projection subspaces than if, e.g., $(x_1, y_1) \in \Omega_H$ coincides with $(x_2, y_2) \in \Omega_h$; see Figure 6.1 (right). As shown in Figure 6.1, for example, with 7×7 grid points at the finest level, the latter coarsening approach leads to only 9 deflation vectors, i.e., $r = 9$. In contrast, the earlier approach results in 16 deflation vectors ($r = 16$), which eventually shift 16 small eigenvalues.

Both approaches, however, produce the same number of deflation vectors if an even number of grid points is used in each direction.

Having defined the coarse-grid points according to Figure 6.1 (left), the deflation vectors are determined by using the bilinear interpolation process of coarse-grid value into the fine grid as follows [21], for level 2 to level 1 (see Figure 6.3 (a) for the meaning of the symbols):

$$I_H^h v^{(1)}(x,y) = \begin{cases} v^{(2)}(x,y), & \text{for } \bullet, \\ \frac{1}{2}[v^{(2)}(x,y-h) + v^{(2)}(x,y+h)], & \text{for } \square, \\ \frac{1}{2}[v^{(2)}(x-h,y) + v^{(2)}(x+h,y)], & \text{for } \triangle, \\ \frac{1}{4}[v^{(2)}(x-h,y-h) + v^{(2)}(x-h,y+h) & \\ \quad + v^{(2)}(x+h,y-h) + v^{(2)}(x+h,y+h)], & \text{for } \circ\,. \end{cases}$$

ETNA
Kent State University
http://etna.math.kent.edu

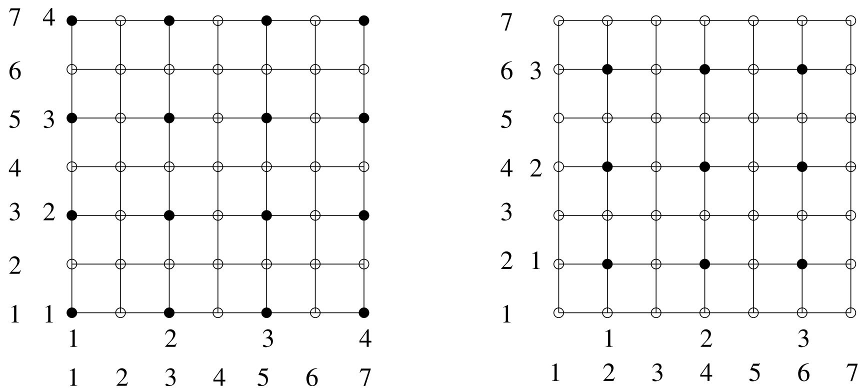


FIGURE 6.1. *Fine (white circles) and coarse (black circles) grid selections in 2D multigrid. Black circles also coincide with the fine grids. Coarsening as depicted in the left figure leads to both better multigrid methods for the shifted Laplacian and larger projection subspaces.*

In some cases, however, such a coarsening may result in the last-indexed coarse-grid points which do not coincide with the last-indexed fine-grid points. This is illustrated in Figure 6.2. There are three possible situations for such coarse-grid points, which are sum-

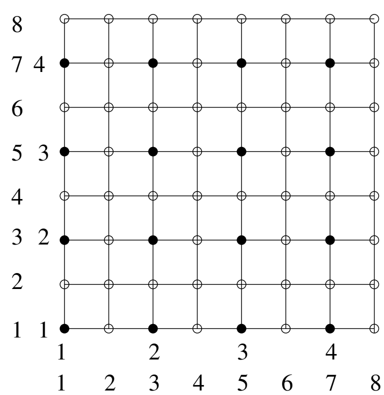


FIGURE 6.2. *Fine (white circles) and coarse (black circles) grid selections in 2D multigrid, where the last indexed gridpoints do not coincide.*

marized in Figure 6.3 (b)–(d). The interpolation associated with $(N_{x,h}h, jh)$, $(ih, N_{y,h}h)$, $(N_{x,h}h, N_{y,h}h) \in \Omega_h$ are given as follows.

- For fine-grid points $(x = N_{x,h}h, y = i_y h)$ (Figure 6.3 (b))

$$I_H^h v^{(1)}(x,y) = \begin{cases} v^{(2)}(x,y), & \text{for } \bullet, \\ \frac{1}{2}[v^{(2)}(x,y-h) + v^{(2)}(x,y+h)], & \text{for } \square, \\ v^{(2)}(x-h,y), & \text{for } \triangle, \\ \frac{1}{2}[v^{(2)}(x-h,y-h) + v^{(2)}(x-h,y+h)], & \text{for } \circ\,. \end{cases}$$

ETNA
Kent State University
http://etna.math.kent.edu

- For fine-grid points $(x = i_x h, y = N_{y,h}h)$ (Figure 6.3 (c))

$$I_H^h v^{(1)}(x,y) = \begin{cases} v^{(2)}(x,y), & \text{for } \bullet, \\ v^{(2)}(x,y-h), & \text{for } \square, \\ \frac{1}{2}[v^{(2)}(x-h,y) + v^{(2)}(x+h,y)], & \text{for } \triangle, \\ \frac{1}{2}[v^{(2)}(x-h,y-h) + v^{(2)}(x+h,y-h)], & \text{for } \circ\,. \end{cases}$$

- For fine-grid points $(x = N_{x,h}h, y = N_{y,h}h)$ (Figure 6.3 (d))

$$I_H^h v^{(1)}(x,y) = \begin{cases} v^{(2)}(x,y), & \text{for } \bullet, \\ v^{(2)}(x,y-h), & \text{for } \square, \\ v^{(2)}(x-h,y), & \text{for } \triangle, \\ v^{(2)}(x-h,y-h), & \text{for } \circ\,. \end{cases}$$

Based on the interpolation matrix I_H^h, we set $Z_{(1,2)} = Z_{(h,H)} = I_H^h$ and $R_h^H = (I_H^h)^T$.

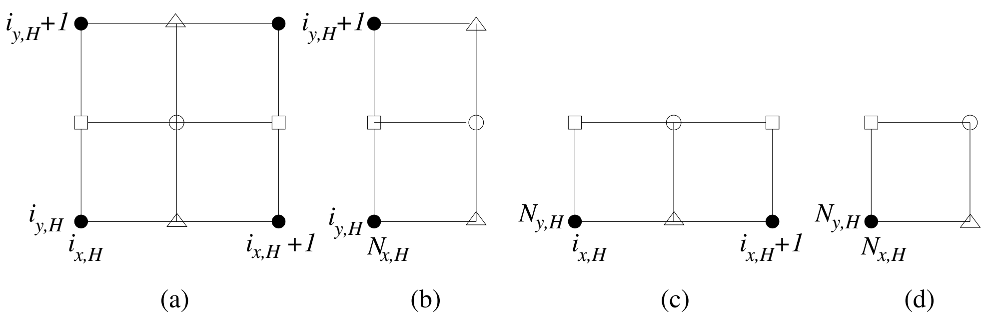


FIGURE 6.3. *Fine (white colored) and coarse (black colored) grid selection indicating the bilinear interpolation in 2D multigrid. Black circles (●) coincide with the fine grids.*

Convergence results are shown in Tables 6.4–6.8 for various wavenumbers. From these tables, for low grid resolutions (e.g., 15 grid points per wavelength) we observe convergence of MKMG which is mildly dependent on the wavenumber k. The convergence becomes less dependent on k if the grid size h is smaller; see also Figures 6.4–6.6 for comparisons with MG.

TABLE 6.4
Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w stands for "# of grid points per wavelength". Multilevel Krylov method with MKMG(4,2,1).

g/w	$k=20$	$k=40$	$k=60$	$k=80$	$k=100$	$k=120$	$k=200$	$k=300$
15	11	14	15	17	20	22	39	64
20	12	13	15	16	18	21	30	45
30	11	12	12	13	13	15	24	39

From Tables 6.4–6.8, it is apparent that MKMG(8,2,1) is the most efficient method, so far, in terms of the number of iterations; it converges faster for all k and h used. If one is more concerned with the number of MKMG iterations to reach convergence, one can use more iterations at the level $j = 3$ (e.g., MKMG(8,3,1), not shown), but this setting does not lead to a further reduction in CPU time.

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 6.5
Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w stands for "# of grid points per wavelength". Multilevel Krylov method with MKMG(5,2,1).

g/w	$k=20$	$k=40$	$k=60$	$k=80$	$k=100$	$k=120$	$k=200$	$k=300$
15	11	14	15	18	19	21	31	52
20	12	13	15	15	16	18	25	37
30	11	12	12	13	13	14	18	28

TABLE 6.6
Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w stands for "# of grid points per wavelength". Multilevel Krylov method with MKMG(6,2,1).

g/w	$k=20$	$k=40$	$k=60$	$k=80$	$k=100$	$k=120$	$k=200$	$k=300$
15	11	14	14	18	18	20	28	47
20	12	13	15	15	16	17	25	36
30	11	12	12	13	13	14	16	25

TABLE 6.7
Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w stands for "#grid points per wavelength". Multilevel Krylov method with MKMG(8,2,1).

g/w	$k=20$	$k=40$	$k=60$	$k=80$	$k=100$	$k=120$	$k=200$	$k=300$
15	11	14	14	17	18	21	27	39
20	12	13	15	14	15	16	20	28
30	11	12	12	12	13	14	15	19

TABLE 6.8
Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w stands fostands for "#grid points per wavelength". Multilevel Krylov method with MKMG(4,3,1).

g/w	$k=20$	$k=40$	$k=60$	$k=80$	$k=100$	$k=120$	$k=200$	$k=300$
15	11	14	15	18	20	22	40	66
20	12	14	15	16	17	20	29	39
30	11	12	12	14	14	15	23	35

In order to gain insight onto the total arithmetic operations needed by MKMG, in Figures 6.4–6.6, we compare CPU time needed by MKMG and MG to reach convergence. We measure the elapsed time on a Pentium 4 machine for the initialization and iteration phase with the MATLAB commands `tic/toc`. Since the `for` loop is used in most parts of the initialization phase, the measured time is too pessimistic.

From Figures 6.4–6.6 we observe that, for low wavenumbers, MG is still faster than any MKMG methods. MKMG only outperforms MG when the wavenumber becomes sufficiently large. For instance, MKMG(8,2,1) is faster than MG for $k > 150$, in terms of number of iterations and CPU time.

For $k = 300$, we were unable to run MG until convergence because of the excessive memory used to keep all Arnoldi vectors. With 30 gridpoints per wavelength, the solution vector alone has 2.25×10^6 complex-valued entries. In this case, restarting GMRES does not help. With full GMRES, we have to terminate the iteration after 86 iterations with the

ETNA
Kent State University
http://etna.math.kent.edu

computed residual only 6.55×10^{-4}, and with about 2.3×10^4 seconds of CPU time. Even though for MKMG the initialization phase also consists of computing coarse-grid information associated with matrices $A^{(j)}$ and $B^{(j)}$, and not only $M^{(j)}$ as in MG, the extra computation does not significantly contribute to the total initialization time, as shown in the lower part of Figures 6.4–6.6 (right). With nearly wavenumber-independent convergence, MKMG requires far less memory than MG for high wavenumbers.

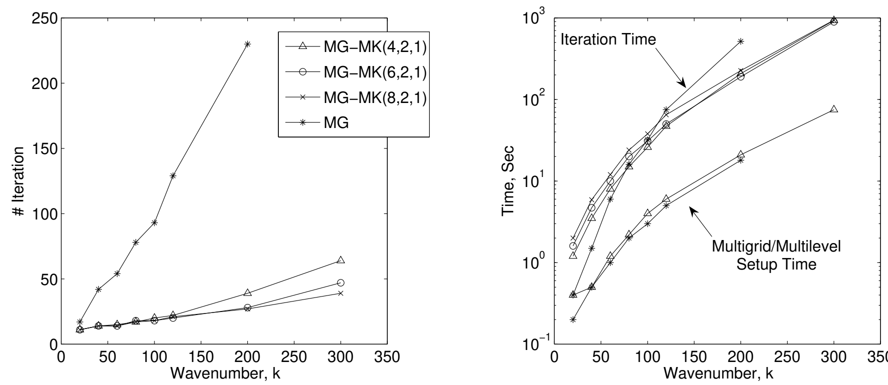


FIGURE 6.4. *Number of iterations and CPU time for GMRES with multigrid applied to the shifted Laplacian preconditioner (MG) and multigrid-multilevel Krylov method (MKMG). 15 grid points per wavelength.*

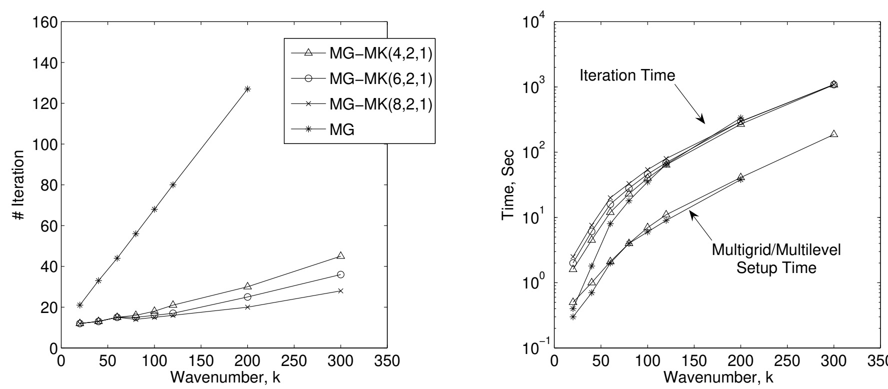


FIGURE 6.5. *Number of iterations and CPU time for GMRES with multigrid applied to the shifted Laplacian preconditioner (MG) and the multigrid-multilevel Krylov method (MKMG). 20 grid points per wavelength.*

7. Conclusions. In this paper, we have discussed a new multilevel Krylov method for solving the 2D Helmholtz equation. This MKMG method is based on a multilevel Krylov method applied to the Helmholtz equation preconditioned by the shifted Laplacian. With this method, small eigenvalues of the original preconditioned system and the associated Galerkin (coarse-grid) systems are shifted to one, leading to favorable spectra for the convergence of Krylov subspace methods. At every level in the MKMG method, a few Krylov iterations are used to solve the projected Galerkin (coarse-grid) preconditioned problems. The preconditioner solves are done by one multigrid iteration, whose maximum level is reduced according to the projection level.

ETNA
Kent State University
http://etna.math.kent.edu

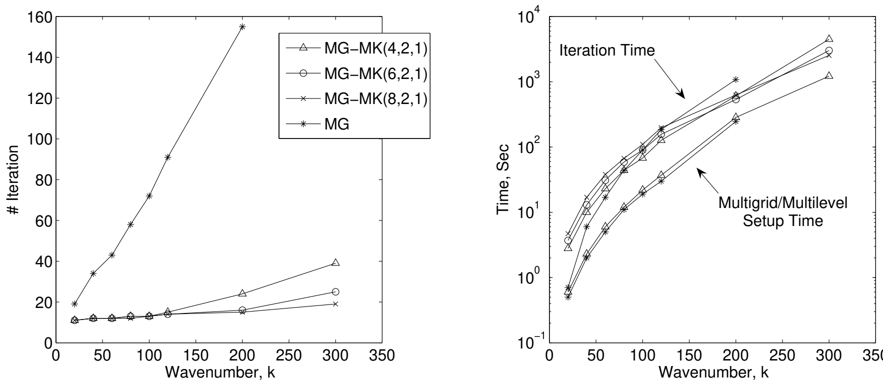


FIGURE 6.6. *Number of iterations and CPU time for GMRES with multigrid applied to the shifted Laplacian preconditioner (MG) and multigrid-multilevel Krylov method (MKMG). 30 grid points per wavelength.*

Numerical experiments have been performed on the 1D and 2D Helmholtz equation with constant wavenumber. The MKMG method leads to only mildly h-dependent and k-dependent convergence. This considerable improvement in the convergence rate leads to a speed up in CPU time when compared to Krylov methods with multigrid-based preconditioner alone.

Finally, this multilevel Krylov method consists of several ingredients: a preconditioner for Krylov iterations, restriction and prolongation operators, an approximation of the maximum eigenvalue, and an approximation to the Galerkin matrix. In this paper, we have chosen a specific choice of all these ingredients, some of which are the same as and have been the integral parts of a multigrid-based preconditioning method for the Helmholtz equation. Nevertheless, other choices or new developments in those methods can be easily implemented in our multilevel Krylov framework to obtain an even faster convergence.

Acknowledgment. We thank two anonymous referees for their constructive comments and remarks, which improve significantly the manuscript, and Gavin Menzel-Jones for his English-related suggestions.

REFERENCES

[1] I. BABUSKA, F. IHLENBURG, T. STROUBOULIS, AND S. K. GANGARAJ, *A posteriori error estimation for finite element solutions of Helmholtz's equation. Part II: Estimation of the pollution error*, Internat. J. Numer. Methods Engrg., 40 (1997), pp. 3883–3900.

[2] A. BAYLISS, C. I. GOLDSTEIN, AND E. TURKEL, *On accuracy conditions for the numerical computation of waves*, J. Comput. Phys., 59 (1985), pp. 396–404.

[3] M. EIERMANN, O. G. ERNST, AND O. SCHNEIDER, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comput. Appl. Math., 123 (2000), pp. 261–292.

[4] H. C. ELMAN, O. G. ERNST, AND D. P. O'LEARY, *A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations*, SIAM J. Sci. Comput., 22 (2001), pp. 1291–1315.

[5] B. ENGQUIST AND A. MAJDA, *Absorbing boundary conditions for the numerical simulation of waves*, Math. Comp., 31 (1977), pp. 629–651.

[6] ———, *Absorbing boundary conditions for the numerical simulation of waves*, Math. Comp., 31 (1977), pp. 629–651.

[7] Y. A. ERLANGGA AND R. NABBEN, *Deflation and balancing preconditioners for Krylov subspace methods applied to nonsymmetric matrices*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 684–699.

[8] ———, *Multilevel projection-based nested Krylov iteration for boundary value problems*, SIAM J. Sci. Comput., 30 (2008), pp. 1572–1595.

ETNA
Kent State University
http://etna.math.kent.edu

[9] ———, *Algebraic multilevel Krylov methods*, SIAM J. Sci. Comput., 31 (2009), pp. 3417–3437.
[10] Y. A. ERLANGGA, C. W. OOSTERLEE, AND C. VUIK, *A novel multigrid-based preconditioner for the heterogeneous Helmholtz equation*, SIAM J. Sci. Comput., 27 (2006), pp. 1471–1492.
[11] Y. A. ERLANGGA, C. VUIK, AND C. W. OOSTERLEE, *On a class of preconditioners for solving the Helmholtz equation*, Appl. Numer. Math., 50 (2004), pp. 409–425.
[12] ———, *Comparison of multigrid and incomplete LU shifted-Laplace preconditioners for the inhomogeneous Helmholtz equation*, Appl. Numer. Math., 56 (2006), pp. 648–666.
[13] J. FRANK AND C. VUIK, *On the construction of deflation-based preconditioners*, SIAM J. Sci. Comput., 23 (2001), pp. 442–462.
[14] R. B. MORGAN, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.
[15] R. NABBEN AND C. VUIK, *A comparison of deflation and the balancing preconditioner*, SIAM J. Sci. Comput., 27 (2006), pp. 1742–1759.
[16] R. A. NICOLAIDES, *Deflation of conjugate gradients with applications to boundary value problems*, SIAM J. Numer. Anal., 24 (1987), pp. 355–365.
[17] Y. SAAD, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.
[18] ———, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003.
[19] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
[20] J. TANG, R. NABBEN, K. VUIK, AND Y. A. ERLANGGA, *Theoretical and numerical comparison of projection methods derived from deflation*, J. Sci. Comput., 39 (2009), pp. 340–370.
[21] U. TROTTENBERG, C. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, New York, 2001.
[22] M. B. VAN GIJZEN, Y. A. ERLANGGA, AND C. VUIK, *Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted Laplacian*, SIAM J. Sci. Comput., 29 (2006), pp. 1942–1952.
[23] R. S. VARGA, *Gerschgorin and His Circles*, Springer, Berlin-Heidelberg, 2004.