

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.* (2007), to appear], a multilevel Krylov method is proposed to solve linear systems with symmetric and nonsymmetric matrix of coefficients. This multilevel method is developed based on shifting (or projecting) some small eigenvalues to the largest eigenvalue, leading to a more favorable spectrum for convergence acceleration of a Krylov subspace method. Such a projection is insensitive with respect to the approximation of the small eigenvalues to be projected, which for a particular choice of deflation subspaces is equivalent to solving a coarse-grid problem analogue to multigrid. Different from multigrid, in the multilevel Krylov method, however, the coarse-grid problem is solved by a Krylov method, whose convergence rate is further accelerated by applying projection to the coarse-grid system. A recursive application of projection and coarse-grid solve by a Krylov iterative method then leads to the multilevel Krylov method. The method has been successfully applied 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]. Since in this case projection must be applied to the preconditioned system AM^{-1} , the coarse-grid matrices are approximated by a product of some low dimension matrices associated with A and M . Within the Krylov iteration and projection step in each coarse-grid solve, a multigrid iteration is used to approximately invert the preconditioner. Hence, a multigrid-multilevel Krylov 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 only mildly independent of wavenumber k .

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

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

1. Introduction. It is well known that the convergence of Krylov subspace methods applied to the linear system

$$Au = b, \quad A \in \mathbb{C}^{N \times N} \quad (1.1)$$

depends to some extent on the spectrum of A . Suppose that A is Hermitian positive definite, convergence of conjugate gradient (CG) methods then depends in particular on the condition number of A , $\kappa(A)$ [17], which in this case is the ratio of the largest eigenvalue to the smallest one. If A is obtained from a discretization of PDEs with elliptic, self-adjoint operator, the convergence is typically characterized by the smallest eigenvalues, which can be of order $O(h^2)$. So, as the grid is refined, the convergence deteriorates.

For general matrices, no convergence results similar to the Hermitian case exist so far. Moreover, for a linear system with matrix of coefficients A having more clustered spectrum a Krylov method will converge faster than for a system with less clustered spectrum. Nevertheless, it is well accepted that eigenvalues close to zero can cause convergence problems.

*The research was supported by the *Deutsche Forschungsgemeinschaft* (DFG), Project Number NA248/2-2

[†]TU Berlin, Institut für Mathematik, MA 3-3, Strasse des 17. Juni 136, D-10623 Berlin, Germany, phone: +49-30-31429290, fax: +49-30-31429621 (erlangga@math.tu-berlin.de)

[‡]TU Berlin, Institut für Mathematik, MA 3-3, Strasse des 17. Juni 136, D-10623 Berlin, Germany, phone: +49-30-31429291, fax: +49-30-31429621 (nabben@math.tu-berlin.de)

A typical remedy for slow convergence of Krylov iterations is by incorporating a preconditioner. The usual criteria for preconditioners are that preconditioning would eventually lead to better clustering of the eigenvalues, and that the preconditioning step must be easy and cheap to carry out.

There exist many ways of preconditioning. Many preconditioners can be classified as preconditioning matrices which approximate in some sense the inverse of A . Incomplete decompositions (e.g., incomplete LU [12]) and approximation inverse, among others, belong to this class. If the preconditioner M is a good approximation to A , we may expect that $M^{-1}A \approx I$, where I is the identity, and therefore $\kappa(M^{-1}A) \approx 1$. For A Hermitian positive definite, CG will then converge in a few number of iterations. This approach does not, however, exploit the detail information of the spectrum of A .

A class of preconditioners which exploits the detail of the spectrum of A can be based on the projection method. Since we know that the smallest eigenvalues are responsible for slow convergence (especially in the initial stage of iterations), the convergence can be accelerated if by any means the components of the residuals corresponding to the smallest eigenvalues can be removed during iterations. One way to achieve this is by deflating a number of smallest eigenvalues to zero. Nicolaides [15] shows that by adding some vectors related to small eigenvalues, the convergence of CG may be improved. For GMRES, Morgan [13] also shows that by augmenting the Krylov subspace by some 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, faster convergence may be expected in this case. See also a unified discussion on this subject by Eiermann *et al.* in [1].

A similar approach is proposed in [9], where a projection matrix resembling deflation of some small eigenvalues is used as preconditioner. Suppose that r smallest eigenvalues are to be deflated to zero. Define the projection

$$P_D = I - AZE^{-1}Y^T, \quad E = Y^T AZ, \quad (1.2)$$

where $Z, Y \in \mathbb{C}^{n \times r}$ are the deflation subspaces and have rank $r \ll n$. The matrix E in this case can generally be considered as the *Galerkin* (or more correctly *Petrov-Galerkin*) matrix associated with A . It can be proved [9, 14, 4] that if A is nonsingular and Z, Y are any rectangular matrices, then the spectrum of $P_D A$ contains r zeros eigenvalues. Furthermore, it can also be shown for a symmetric positive definite matrix A that with larger r , the *effective* condition number becomes smaller [14]. Hence, if one uses a large deflation subspace, convergence can be improved considerably.

Large deflation subspaces, however, rise a negative implication. Since in the preconditioning step one needs to invert the Galerkin matrix E in (1.2), a too large deflation subspace can make this inversion very costly. Related to the computation of E^{-1} , it has been shown in [14] that P_D is sensitive to inaccurate computation of E^{-1} . From the view of the algorithmic implementation, this means that the linear system associated with the Galerkin matrix (called the Galerkin system) must be solved exactly by a sparse direct method, or iteratively up to a sufficiently high accuracy. Reference [18] discusses this aspect of deflation in details with extensive numerical tests.

As an alternative to the deflation (1.2), another projection preconditioner is proposed by the authors in [5]. In this new projection preconditioner, small eigenvalues are shifted *not* towards zero, *but* towards the largest eigenvalue (in magnitude), denoted by λ_n , instead, leading to clustering of eigenvalues in a location far from zero.

Consider the more general linear system

$$\hat{A}\hat{u} = \hat{b}, \quad (1.3)$$

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

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

on the general system (1.3). We prefer to use this notation because this allows us to introduce more general problems involving preconditioners. As discussed in [5] for some problems (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 of $M_1 = I$ and $M_2 = \text{diag}(A)$ is already sufficient.

One clear advantage of (1.4) over (1.2) is that $P_{\hat{N}}$ is insensitive with respect to an inexact solve of the Galerkin system. Such a nice property will eventually allow the use of a large deflation subspace in order to shift as many small eigenvalues as possible, and the use of inner iterations to solve the associated Galerkin system. In principle, one can just employ standard Krylov iterations upto a desired accuracy, which can be set less tightly. The convergence rate of Krylov iterations employed can however be significantly improved if projection similar to (1.4) is also applied to the Galerkin system. Clearly, the action of this projection will require another solve of a Galerkin system, which will be done by Krylov iterations. If this process is done recursively, a ‘‘multilevel projection Krylov’’ method, or simply ‘‘multilevel Krylov’’ method results. The potential of this multilevel Krylov method has been demonstrated in [4].

In the multilevel Krylov method, there is freedom in choosing the deflation subspaces Z and Y . Ideally, the columns of Z should consist of the right eigenvectors associated with the small eigenvalues to be projected. But this is impractical. Alternatively, the matrix Z can for instance be determined from an intergrid process similar to that in multigrid methods. By setting $Y = Z$, it has been shown in [4] that with this choice of Z and Y , the convergence of a Krylov subspace method applied to

$$P_{\hat{N}}\hat{A}\hat{u} = P_{\hat{N}}\hat{b} \quad (1.5)$$

can be made independent of grid size h and physical parameters involved (e.g., the Péclet number in convection-diffusion cases).

In this paper we extend the application of the multilevel Krylov method to indefinite linear systems. In particular, we will focus on the Helmholtz equation. From this context, this paper can be considered as a continuation of our discussion on the multilevel Krylov method presented in [4]. Therefore, for more theoretical results on the method, Reference [4] must always be consulted.

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 of our multilevel Krylov method are discussed. Some practical implementations are explained in Section 4. Convergence results from 2D Helmholtz problems are presented in Section 5. Finally, in Section 6, we draw some conclusions.

For more details about the shifted Laplace preconditioner for the Helmholtz equation, we refer the reader to [7, 6, 20]. In [2], Elman *et al.* used Krylov iterations (in

their case, GMRES) in a multilevel fashion. But, their approach is basically a multi-grid concept specially designed for the Helmholtz equation. While at the finest and coarsest level, standard smoothers still have good smoothing properties, at the intermediate levels GMRES is run in place of smoothers. GMRES, however, does not have smoothing property. Thus, GMRES plays a role in reducing the errors and *not* in smoothing them. A substantial number of GMRES iterations in the intermediate levels, however, is required to do this.

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

$$Au := - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k^2(x, y) \right) u(x, y) = g(x, y), \quad \text{in } \Omega \subset \mathbb{R}^2, \quad (2.1)$$

where $k(x, y) = 2\pi f/c(x, y)$ is the wavenumber related to a given source frequency f and the local speed of sound c . In (2.1), g is the source term. Reflecting, non-reflecting and pressure release conditions can be applied at the boundaries $\Gamma = \partial\Omega$; see, e.g., [3]. If a discretization is applied on (2.1) and the boundary conditions, and the wavenumber is high (as usually encountered in realistic applications) the resultant linear system is large but sparse, and symmetric but indefinite. An application of Krylov subspace methods to iteratively solve the linear system results in slow convergence. Standard preconditioners, e.g. ILU-type preconditioner, do not effectively improve the convergence [8].

In [7],[6], the shifted Laplacian operator

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

has been proposed to accelerate the convergence of a Krylov subspace method. The preconditioner matrix is then obtained from discretization of (2.2), with the *same* boundary conditions as for (2.1). In the end, the (right) preconditioned system

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

has to be solved, where A and M are the associated Helmholtz and preconditioner matrix respectively.

The eigenvalues of AM^{-1} are closely clustered around zero, which positions are determined by the actual choice of α and β . In this paper we will only consider the pair $(\alpha, \beta) = (1, 0.5)$, which in [6] has been 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 will give some insights on the spectrum of the preconditioned Helmholtz system (2.3) throughout the rest of this section. We first discuss Helmholtz problems with Dirichlet or Neumann boundary conditions. For simplicity, we only consider the 1D problem; An extension to higher dimensions is straightforward.

For 1D problems with Dirichlet or Neumann boundary conditions, eigenvalues of the preconditioned problem can be written as

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

with λ the eigenvalue. By using the ansatz $u = \sin(i\pi x)$, $i = 1, \dots, n$, from (2.4) we find that

$$\lambda_i = \frac{i^2\pi^2 - k^2}{i^2\pi^2 - (1 - 0.5\hat{j})k^2}, \quad (2.5)$$

whose real and imaginary part of the eigenvalues are

$$\operatorname{Re}(\lambda_i) = \frac{(i^2\pi^2 - k^2)^2}{(i^2\pi^2 - k^2)^2 + 0.25k^4}, \quad (2.6)$$

$$\operatorname{Im}(\lambda_i) = -\frac{0.5(i^2\pi^2 - k^2)k^2}{(i^2\pi^2 - k^2)^2 + 0.25k^4}. \quad (2.7)$$

Observe that

$$\max_i \operatorname{Re}(\lambda_i) \rightarrow 1, \text{ and } \max_i |\operatorname{Im}(\lambda_i)| \rightarrow 0.5, \quad (2.8)$$

i.e., the eigenvalues are bounded by the box $[0, 1] \times [-0.5, 0.5]$. In particular, it is easily verified that

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

which leads to

$$\operatorname{Re}(\lambda_i)^2 - \operatorname{Re}(\lambda_i) + \operatorname{Im}(\lambda_i)^2 = 0 \implies (\operatorname{Re}(\lambda_i) - 0.5)^2 + \operatorname{Im}(\lambda_i)^2 = 0.25 \quad (2.9)$$

because of the relations (2.6) and (2.7). Thus, the eigenvalues are distributed in the complex plane on a circle with center $c = \frac{1}{2}$ and radius $R = \frac{1}{2}$. Furthermore, relation (2.9) implies that for the maximum real part of the eigenvalues, $\lim_{\operatorname{Re}(\lambda_i) \rightarrow 1} \operatorname{Im}(\lambda_i) = 0$. Hence, the spectral radius and the maximum eigenvalue are bounded above by one.

For more general boundary conditions (e.g., a radiation condition on a part of the boundary), the following theorem can be proved, which is a special consequence of Theorem 3.5 in [20] for $(\alpha, \beta) = (1, 0.5)$.

THEOREM 2.1. [20] *Let $A = L + \hat{j}C - K$ and $M = L + \hat{j}C - (1 - 0.5\hat{j})K$, with L , C and K matrices associated with discretizations of the negative Laplacian, the boundary conditions and the Helmholtz term, respectively. Then the eigenvalues of $M^{-1}A$ are inside or on the circle with center $c = \frac{1}{2}$ and radius $R = \frac{1}{2}$.*

Proof. The proof for arbitrary (α, β) can be found in [20]. \square

Theorem 2.1 is easily verified for the left preconditioning case, and holds also for the right preconditioning case because $\sigma(M^{-1}A) = \sigma(AM^{-1})$. Moreover, the theorem is also valid for higher dimension Helmholtz problems, and for any type of discretizations used since no assumption is made in these regard. In case of non-radiating boundary conditions, $C = 0$ and the eigenvalues lie on the circle $(c, R) = (0.5, 0.5)$.

What is important from the convergence of Krylov subspace methods point of view is that while the spectrum of AM^{-1} is nicely clustered, some eigenvalues at the same time lie close to zero. This is illustrated in Figure 2.1 for 1D case. With an increase in the wavenumber, the number of eigenvalues clustered around zero appears to increase also. While the maximum eigenvalue hardly changes, only the smallest eigenvalues are sensitive to the wavenumber; see discussion in, e.g., [7] for details. Thus, the convergence is mostly determined by the small eigenvalues.

In the subsequent sections we will discuss the use of multilevel Krylov method to handle the small eigenvalues related to preconditioned Helmholtz systems.

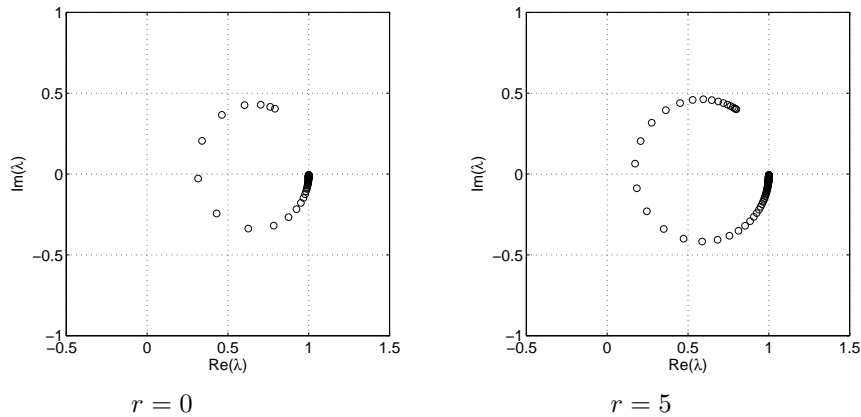


FIG. 2.1. Spectrum of a typical 1D Helmholtz problem preconditioned with the shifted Laplacian. The wavenumber k is 20 (left) and 50 (right)

3. Multilevel Krylov method. The convergence rate of a Krylov subspace method applied to the system (2.3) may be improved by using a projection method. Consider the linear system

$$\hat{A}\hat{u} = \hat{b}, \quad (3.1)$$

where in this case $\hat{A} = AM^{-1}$ and $\hat{b} = b$. Our objective is to project the troublesome eigenvalues in the spectrum to a value such that the projected linear system will have a more favorable spectrum for convergence acceleration.

As explained in Section 1, one way to achieve this is by using some deflation techniques, in which the small eigenvalues are deflated to zero. Deflation, however, is very sensitive to the inexact solve of the Galerkin system. If the Galerkin problem is solved only approximately, the small eigenvalues are not shifted exactly to zero but to some very small values close to zero, leading to even further convergence deterioration.

In the multilevel Krylov method, the projection is done based on shifting small eigenvalues to a value around the largest eigenvalue. This can be achieved by using the algebraic operator

$$Q_{\hat{N}} = I - Z\hat{E}^{-1}Y^T\hat{A} + \lambda_n Z\hat{E}^{-1}Y^T, \quad \hat{E} = Y^T\hat{A}Z, \quad (3.2)$$

where $\lambda_n = \{\lambda_i : \max_i |\lambda_i|\}$, and then solving the projected system

$$\hat{A}Q_{\hat{N}}\tilde{u} = \hat{b}, \quad \hat{u} = Q_{\hat{N}}\tilde{u}. \quad (3.3)$$

Note that if we set $\lambda_n = 0$ in (3.2) we recover the deflation preconditioner. Furthermore, $Q_{\hat{N}}$ is the right preconditioner version of (1.4).

For (3.3) the following spectral property holds.

THEOREM 3.1. *Let $Z, Y \in \mathbb{C}^{n \times r}$ ($r \ll n$) be the rectangular matrices whose columns are the right and left eigenvectors associated with the smallest r eigenvalues (in magnitude) of \hat{A} . Suppose that $\lambda_1, \dots, \lambda_n \in \sigma(\hat{A}) \subset \mathbb{C}$ are increasingly ordered based on their magnitude. Then*

$$\sigma(\hat{A}Q_{\hat{N}}) = \{\lambda_n, \dots, \lambda_n, \lambda_{r+1}, \dots, \lambda_n\}. \quad (3.4)$$

Proof. The proof requires the identity $P_D AZ = 0$, which is easily verified by direct computation (see, e.g., [9]), and Theorem 3.5 of [4] which states that $\sigma(P_{\hat{N}}\hat{A}) = \sigma(\hat{A}Q_{\hat{N}})$.

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 $r+1 \leq i \leq n$, 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 [4], $\sigma(P_{\hat{N}}\hat{A}) = \{\lambda_n, \dots, \lambda_n, \lambda_{r+1}, \dots, \lambda_n\} = \sigma(\hat{A}Q_{\hat{N}})$. \square

Thus, after projection r eigenvalues are no longer small and have been shifted to λ_n . The smallest eigenvalue (in magnitude) now is λ_{r+1} . The rest of the spectrum is remained untouched. If λ_{r+1} is of the same order of magnitude with λ_n , convergence of a Krylov subspace method can be accelerated considerably.

The computation of eigenvectors however is very expensive for large linear systems. Hence, Z and Y can not be easily constructed in this way. In practice, we use some r arbitrary sparse vectors satisfying condition that $Z, Y \in \mathbb{C}^{n \times r}$ are full rank. With arbitrary full rank Z and Y , Theorem 3.1 is however no longer completely true. While it can be verified that the smallest r eigenvalues are shifted to λ_n , nothing can be said about the rest of the spectrum. In comparison with the action of the (right) deflation preconditioner $Q_{\hat{D}} := I - Z\hat{E}^{-1}Y^T\hat{A}$, however, the action of $Q_{\hat{N}}$ on \hat{A} leads to a similar spectrum, as stated below.

THEOREM 3.2. *Let $Z, Y \in \mathbb{C}^{n \times r}$ be of rank r . Let \hat{A} be nonsingular. If*

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

then

$$\sigma(\hat{A}Q_{\hat{N}}) = \{\lambda_n, \dots, \lambda_n, \mu_{r+1}, \dots, \mu_n\}. \quad (3.5)$$

Proof. Combine Theorems 3.4 and 3.5 in [4]. \square

In the above theorem, in general $\mu_i \neq \lambda_i$, $i = 1, \dots, n$. Hence, generally speaking, $\mu_n \neq \lambda_n$. Thus, the r small eigenvalues are not shifted to μ_n , but to λ_n . If we simply assume that $\mu_n - \lambda_n \neq 0$, we can always find a constant $\omega \in \mathbb{C}$ such that $\mu_n = \omega\lambda_n$. The constant ω is called the *shift scaling factor*. With this scaling factor, a shift correction can be introduced to (3.2) by replacing λ_n in (3.2) by $\omega\lambda_n$.

To employ $Q_{\hat{N}}$, we need two components: (i) the value λ_n and (ii) the rectangular matrices Z and Y .

Following the discussion in [4], it is not necessary to use the exact value of λ_n in $Q_{\hat{N}}$. In fact, we have to avoid this because computing λ_n from \hat{A} is expensive. For our purpose it is sufficient to use only an approximation to it. An approximation to $\lambda_n(\hat{A})$ can be obtained, e.g., via Gerschgorin's theorem [21]. In many cases, this would lead to a very good estimate. For our Helmholtz problems, however, we will use information we have had in Section 2. For AM^{-1} we have that $|\lambda_n| \rightarrow 1$. Thus, setting $\lambda_n = 1$ will then be the appropriate choice.

Secondly, we need to determine the matrices Z and Y . Following the above discussion we will choose vectors for Z and Y , which are not the eigenvectors of AM^{-1} .

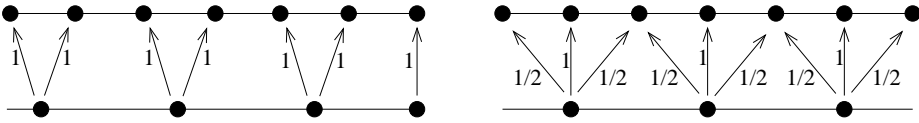


FIG. 3.1. *Interpolation in 1D: piece-wise interpolation (left) and linear interpolation (right)*

We note that $Z : \mathbb{C}^r \mapsto \mathbb{C}^n$, and $Y^T : \mathbb{C}^n \mapsto \mathbb{C}^r$, $r \ll n$. In this case, we can consider Y^T as a *restriction* operator into the subspace \mathbb{C}^r , and Z a *prolongation* operator into the subspace \mathbb{C}^n . Similar notations are used in multigrid methods, and in this context, the matrix Z represents an interpolation process of smooth errors on the coarse grid into the fine. In multigrid, $\hat{E} = Y^T \hat{A} Z$ is then called the Galerkin coarse-grid approximation of \hat{A} . In the multilevel Krylov method, since Z is not necessarily an interpolation process related to the intergrid transfer, we simply call $\hat{E} = Y^T \hat{A} Z$ the ‘‘Galerkin matrix’’.

As advocated in [9] within the deflation preconditioner context, one can choose a sparse matrix arising from an intergrid transfer similar to that in the multigrid context. For example, one can use *piece-wise* constant interpolation from coarse grid into fine grid to determine Z , as illustrated in Figure 3.1 (left) for 1D grid. In this case, the matrix Z can be seen as an interpolation matrix where coarse grid is obtained from agglomeration of some points on the fine grid. Alternatively, we can as well employ higher order interpolation strategies like the *linear* interpolation in 1D or the bilinear interpolation in 2D; see Figure 3.1 (right). Finally, we set $Y = Z$. Such constructions lead to the matrices Z and Y which are sparse.

For illustration, we first consider spectra of $\hat{A} = AM^{-1}$ preconditioned by $Q_{\hat{N}}$, with Z the piece-wise constant interpolation matrix. The spectra are shown in Figure 3.2 for the same 1D Helmholtz problem shown previously in Section 2. In this case, $\lambda_n = 1$, $\omega = 1$, and $Y = Z$. With this choice of Z , as seen in the figure, small eigenvalues nearby the origin are no longer present (cf. Figure 2.1). The action of $Q_{\hat{N}}$ however changes the whole spectrum; i.e. λ_i , $i = r + 1, \dots, n$ are also shifted somewhat. Nevertheless, the projected spectrum is much favorable for Krylov iterations as it is now clustered rather far from the origin. In Figure 3.2(a)–(c), we keep $k = 20$ and vary r , the number of smallest eigenvalues to be shifted. As clearly seen, the larger r is (and hence, the larger the deflation subspace Z is), the more the spectrum is clustered around one. Surprisingly, if $r = n/2$, the spectrum hardly depends on the wavenumber k . This is shown in Figure 3.2(c) and Figure 3.2(d), where the case $k = 20$ and $k = 50$ are considered with $n = 100$ and $n = 250$, respectively. With $r = n/2$, the spectra are now strongly clustered around one. With this very clustered spectrum, we can expect a very fast convergence of a Krylov method. These results are generalized to higher dimension Helmholtz problems.

Next, we consider the spectra of $\hat{A}Q_{\hat{N}}$ if the *linear* interpolation is used for constructing Z . These are shown in Figure 3.3 for $r = n/2$, for $k = 20$ and 50. Compared to Figure 3.2(c) and Figure 3.2(d), the spectra are clustered around one as well. But now, the spectrum for $k = 20$ can not be distinguished from the spectrum for $k = 50$. In this case, we may expect a convergence rate of a Krylov method, which is independent of the wavenumber k .

To see the potential of incorporating a projection method on the preconditioned Helmholtz linear system, we present numerical results from a 1D Helmholtz problem with constant wave number k . For 1D problems, the preconditioner matrix M is easily

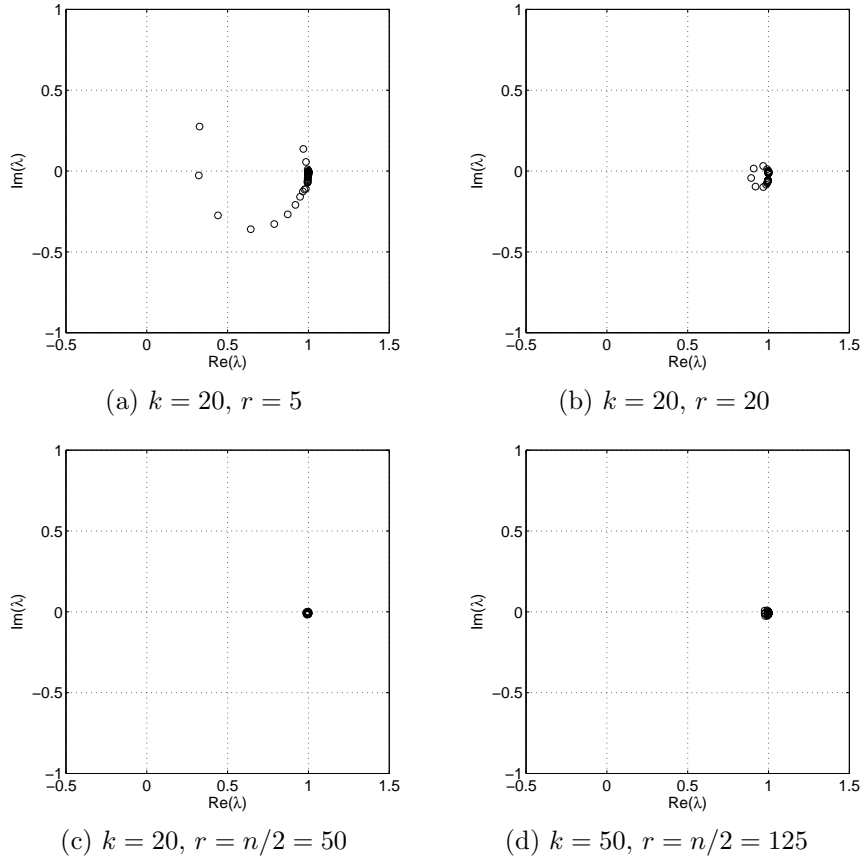


FIG. 3.2. Spectra of a preconditioned 1D Helmholtz problem, $k = 20$ and 50 . The number of grid points for each k is $n = 100$ and 250 , respectively. Z is obtained from the piece-wise constant interpolation

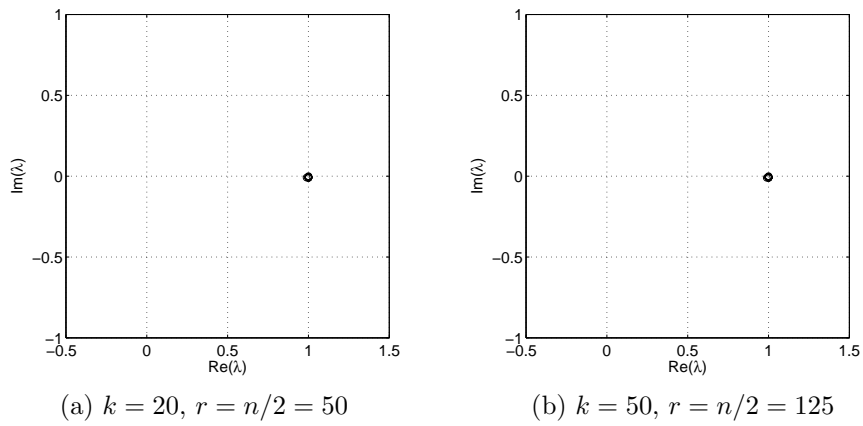


FIG. 3.3. Spectra of a preconditioned 1D Helmholtz problem, $k = 20$ and 50 . The number of grid points for each k is $n = 100$ and 250 , respectively. Z is obtained from the linear interpolation

inverted, and henceforth, the matrix \hat{E} can be explicitly computed. We compare the convergence results for two different strategies for constructing the deflation subspace Z explained above, with $r = n/2$. We note that constructing Z with the piece-wise constant interpolation strategy will eventually lead to slightly sparser Z than with the linear interpolation. The results are shown in Table 3.1. In this table, “standard” means that the preconditioned GMRES is run without projection. Recall that M is the shifted Laplacian preconditioner.

For “standard” iteration, we observe the typical convergence behavior of Krylov methods preconditioned by the shifted Laplacian: the number of iterations to reach convergence increases linearly with respect to k . If the projection operator $Q_{\hat{N}}$ is incorporated and if linear interpolation is used for constructing Z , the convergence becomes independent of the wavenumber k . With piece-wise constant interpolation, the convergence is *almost* independent of k .

TABLE 3.1

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

	$k :$				
	20	50	100	200	500
Standard	14	24	39	65	142
$Q_{\hat{N}}$ with piece-wise constant	4	4	5	6	7
$Q_{\hat{N}}$ with linear interpolation	3	3	3	3	3

4. Multilevel Krylov method with approximate Galerkin systems.

We first note here that the effectiveness of the projection $Q_{\hat{N}}$ is insensitive to the inexact solve of the Galerkin system; see [4]. Therefore, instead of solving the Galerkin system exactly, we can use a Krylov method to approximately solve it. In general, the accuracy of the Galerkin solve depends on how small the residual/error can be reduced by the Krylov method. One can allow a quiet number of Krylov iterations for solving the Galerkin system. But, if the Galerkin system associated with \hat{A} is not better conditioned than the original system \hat{A} , too many Krylov iterations on the Galerkin system would not make the overall performance of the method viable. The convergence of a (inner) Krylov method can be accelerated by applying projection similar to (3.2) on the Galerkin system. A recursive application of projection and iterative Galerkin solve leads to the *multilevel Krylov method*. An algorithm of the multilevel Krylov method is presented in [4].

With respect to the Galerkin solve, one immediate complication arises. Since M^{-1} in $\hat{A} = AM^{-1}$ can not be easily determined by an explicit computation, we can not directly employ a Krylov method to approximately invert $\hat{E} = Y^T AM^{-1}Z$. The matrix M^{-1} is in general dense, and so is the Galerkin matrix \hat{E} . Storing both M^{-1} and \hat{E} will require a considerable amount of memory.

In this section, we discuss a way to approximate \hat{E} in a form which is favorable for an application of a Krylov method. The way we handle the action of preconditioner M will be discussed in Section 5.

First consider the *two-level* Krylov method. With $Q_{\hat{N}}$ included and with any full ranked $Y, Z \in \mathbb{C}^{n \times r}$, the (right) preconditioning step of a Krylov method can be

written as

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

where

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

For example in GMRES, the vector v is the Arnoldi vector computed at the previous iteration. The vector $v' \in \mathbb{C}^n$ of (4.2) is then restricted to \mathbb{C}^r by Y^T in (4.1), namely

$$v'_R := Y^Tv'. \quad (4.3)$$

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

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

Our concern now is on the way $\hat{E}^{-1}v'_R =: v_R$ is invoked. Since $\dim \hat{E} = r \ll N$, it is of our interest to somehow approximate \hat{E} with a product of some matrices of dimension r . If we succeed in doing this, the work of solving the Galerkin system (4.4) will constitute of work for solving smaller systems of size $r \times r$. Furthermore, by restricting the approximation to only a product of some matrices, a Krylov method can then be applied to the Galerkin system (4.4) in a more natural way.

The approach we propose below can not be algebraically justified, but fortunately leads to a practical approximation to \hat{E} . Furthermore, this approach is easily extended to our multilevel setting.

For the two-level Krylov method, the Galerkin matrix \hat{E} is approximated as follows:

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

where the products $A_H := Y^TAZ$ and $M_H := Y^TMZ$ and $B_H := Y^TZ$ are the Galerkin matrices associated with A , M , and I respectively. Due to the approximation (4.5), the Galerkin system (4.4) can now be written as

$$v'_R = A_HM_H^{-1}B_Hv_R, \quad (4.6)$$

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.

In order to construct a multilevel Krylov algorithm, we need to use notations which incorporate level identification. For example, for our 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

$$\begin{aligned} A^{(2)} &= Y^{(1,2)T}A^{(1)}Z^{(1,2)}, \\ M^{(2)} &= Y^{(1,2)T}M^{(1)}Z^{(1,2)}, \\ B^{(2)} &= Y^{(1,2)T}I^{(1)}Z^{(1,2)}. \end{aligned}$$

The matrix $\hat{A}^{(2)}$ is the second level ($j = 2$) Galerkin matrix associated with $\hat{A}^{(1)} = A^{(1)}M^{(1)^{-1}}$, etc. In order to solve the Galerkin system related to $\hat{A}^{(2)}$, we define the projection

$$Q_{\hat{N}}^{(2)} = I - Z^{(2,3)}\hat{A}^{(3)^{-1}}Y^{(2,3)^T}\hat{A}^{(2)} + \omega^{(2)}\lambda_n^{(2)}Z^{(2,3)}\hat{A}^{(3)^{-1}}Y^{(2,3)^T}, \quad (4.7)$$

with

$$\hat{A}^{(3)} = Y^{(2,3)^T}\hat{A}^{(2)}Z^{(2,3)}, \quad (4.8)$$

and solve the projected linear system

$$A^{(2)}M^{(2)^{-1}}B^{(2)}Q_{\hat{N}}^{(2)}\tilde{v}_R^{(2)} = (v'_R)^{(2)}, \quad \text{where } v_R^{(2)} = Q_{\hat{N}}^{(2)}\tilde{v}_R^{(2)}, \quad (4.9)$$

by a Krylov subspace method. If $\hat{A}^{(2)}$ is small enough, $\hat{A}^{(3)}$ in (4.7) can be explicitly computed and inverted. On the other hand, if it is still too large to become practical for exact inversion, we can again consider a similar approximation to (4.5). But in this case we would have

$$\hat{A}^{(3)} = A^{(3)}M^{(3)^{-1}}B^{(3)}, \quad (4.10)$$

where $B^{(3)} := Y^{(2,3)^T}B^{(2)}Z^{(2,3)}$. The corresponding Galerkin system at level $j = 3$ is then defined by

$$\hat{A}^{(3)}v_R^{(3)} = A^{(3)}M^{(3)^{-1}}B^{(3)}v_R^{(3)} = (v'_R)^{(3)}. \quad (4.11)$$

Defining a similar projection to (4.7), the approximate solution of (4.11) is obtained from solving the projected linear system similar to (4.9), but now with the superscript index 3.

Suppose that m levels of projection 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 $j = 2, \dots, m$, choose $Z^{(j-1,j)}$ and $Y^{(j-1,j)}$, and compute

$$A^{(j)} = Y^{(j-1,j)^T}A^{(j-1)}Z^{(j-1,j)}, \quad (4.12)$$

$$M^{(j)} = Y^{(j-1,j)^T}M^{(j-1)}Z^{(j-1,j)}, \quad (4.13)$$

$$B^{(j)} = Y^{(j-1,j)^T}B^{(j-1)}Z^{(j-1,j)}, \quad (4.14)$$

which in principle define

$$\hat{A}^{(j)} = A^{(j)}M^{(j)^{-1}}B^{(j)}. \quad (4.15)$$

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

$$Q_{\hat{N}}^{(j)} = I - Z^{(j-1,j)}\hat{A}^{(j)^{-1}}Y^{(j-1,j)^T} \left(\hat{A}^{(j-1)} - \omega^{(j)}\lambda_n^{(j)}I \right). \quad (4.16)$$

Iteration phase:

$j = 1$

Solve $A^{(1)}M^{(1)-1}\tilde{u}^{(1)} = b$, $u^{(1)} = M^{(1)-1}\tilde{u}^{(1)}$ with Krylov iterations by computing:

$$v_M^{(1)} = M^{(1)-1}v^{(1)};$$

$$s^{(1)} = A^{(1)}v_M^{(1)};$$

$$t^{(1)} = s^{(1)} - \omega^{(1)}\lambda_n^{(1)}v^{(1)};$$

$$\text{Restriction: } (v'_R)^{(2)} = Y^{(1,2)T}t^{(1)};$$

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:

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

$$s^{(2)} = A^{(2)}v_M^{(2)};$$

$$t^{(2)} = s^{(2)} - \omega^{(2)}\lambda_n^{(2)}v^{(2)};$$

$$\text{Restriction: } (v'_R)^{(3)} = Y^{(2,3)T}t^{(2)};$$

If $j = m$

$$v_R^{(m)} = \hat{A}^{(m)-1}(v'_R)^{(m)};$$

else

$j = 3$

$$\text{Solve } A^{(3)}M^{(3)-1}B^{(3)}v_R^{(3)} = (v'_R)^{(3)}$$

...

$$\text{Interpolation: } v_I^{(2)} = Z^{(2,3)}v_R^{(3)};$$

$$q^{(2)} = v^{(2)} - v_I^{(2)};$$

$$w^{(2)} = M^{(2)-1}B^{(2)}q^{(2)};$$

$$p^{(2)} = A^{(2)}w^{(2)};$$

$$\text{Interpolation: } v_I^{(1)} = Z^{(1,2)}v_R^{(2)};$$

$$q^{(1)} = v^{(1)} - v_I^{(1)};$$

$$w^{(1)} = M^{(1)-1}q^{(1)};$$

$$p^{(1)} = A^{(1)}w^{(1)};$$

REMARK 4.1. *In solving the Galerkin problems by a Krylov subspace method, zero initial guess is always used. With this choice, the initial residual needs not 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 $\omega^{(j)} = 1$, $j = 1, \dots, n-1$, taking $\lambda_n^{(j)} = 1$ leads to a good method.*

5. Multigrid-multilevel Krylov method. In Algorithm 1, at every level two preconditioner solves related to M are required to compute $v_M^{(j)}$ and $w^{(j)}$. For these solves, one can, e.g., first construct an incomplete ILU factorization of M and then do forward-backward substitutions. For the Helmholtz equation, our experiments show, however, that this approach does not lead to an efficient multilevel Krylov method, and the convergence of the multilevel Krylov method quickly deteriorates with an increase in the wavenumber. To compensate the convergence deterioration one needs to increase fill-in in the ILU factors, which is often limited by memory

requirement. Moreover, at every level an ILU factors associated with $M^{(j)}$ needs be computed. This computation is in general costly and may make the initialization phase becoming expensive.

An efficient way of solving the preconditioning steps can be done by using a multigrid method. The use of multigrid to approximate the shifted Laplacian preconditioner has been presented in [8], which consists of F-cycle with one pre- and postsmoothing with underrelaxed Jacobi as the smoother. Even though the resultant error reduction factor is not that of the typical text-book multigrid convergence ($\rho = 0.6$), this choice leads to an effective preconditioner for convergence acceleration of Krylov subspace methods for the Helmholtz equation. We will use this multigrid setting for the preconditioner solves in our multilevel Krylov method explained in Algorithm 1.

Assume that a sequence of fine and coarse grids Ω^j , $j = 1, \dots, m$ are given. Two transfer operators between two grids Ω^j and Ω^{j+1} , denoted by

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

are associated with the restriction and interpolation (or prolongation) process, respectively, and are as well given. A multigrid algorithm related to, 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: $r^{(j+1)} = I_j^{j+1} r^{(j)}$;

Coarse-grid problem:

if $j = m$ solve $e^{(m)} = M^{(m)-1} r^{(m)}$;

else

...

endif

Prolongation: $d^{(j)} = I_{j+1}^j e^{(j+1)}$;

Defect correction: $v_{M,\ell+2/3}^{(j)} = v_{M,\ell+1/3}^{(j)} + d^{(j)}$;

Post-smoothing: $v_{M,\ell+1}^{(j)} = \text{smooth}(M^{(j)}, v_{M,\ell+2/3}^{(j)}, v_B^{(j)})$;

Note that in Algorithm 2, the finest multigrid level is always the same as the current level in the multilevel Krylov step. Hence, if a particular projection is done at level $j = J < m$, then multigrid with $J - m$ grid levels is used to approximate the action of preconditioner $M^{(J)}$.

In multigrid, the restriction operator I_j^{j+1} and the prolongation operator I_{j+1}^j are chosen such that smooth errors can be well represented on every grid. This is typically done by an interpolation of smooth errors on the coarse grid into the fine grid. In the multilevel Krylov method, beside the use of eigenvectors associated with the first r smallest eigenvalues, there is so far no criterion on how to construct the deflation subspace. However, numerical results in Table 3.1 suggest that the deflation subspace based on linear interpolation leads to better convergence of GMRES than the piece-wise constant interpolation. Thus, an interpolation matrix used in multigrid is one of possible candidates for the deflation subspace Z .

From an implementation point of view, the use of linear interpolation as the basis for the deflation subspace provides another advantage. Since in this case $Z^{(j-1,j)} = I_j^{1-1}$, the matrices $M^{(j)}$ used in the multilevel Krylov and the multigrid step are now the same. (For the latter, this matrix is rather called the (Galerkin) coarse-grid matrix.) Hence, we can use both $A^{(j)}$ and $M^{(j)}$ in the multilevel Krylov and multigrid algorithm. These matrices are computed only once in the multilevel Krylov method initialization phase. Since in the multilevel Krylov method, we also need $B^{(j)}$ to determine the Galerkin system associated with \hat{A} , this initialization phase also computes these matrices.

Clearly, the overall initialization phase is at least three times more expensive than the multigrid initialization phase. The total storage needed is also as three times expensive as that of multigrid.

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

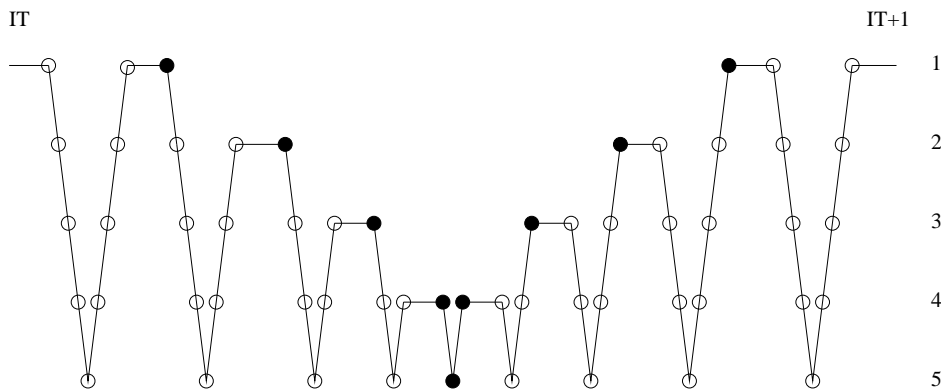


FIG. 5.1. Multigrid-multilevel Krylov cycle with $m = 5$. “•”: multilevel Krylov step; “○”: multigrid step

6. Numerical experiments. In this section we present numerical examples from the 1D and 2D Helmholtz equation. The convergence of multigrid-multilevel Krylov method described in the previous section is compared with the convergence obtained by Krylov method preconditioned by a multigrid iteration acting on the shifted Laplacian as discussed in [8]. For simplicity, we use abbreviations MG-MK for the multigrid-multilevel Krylov method and MG for the second method in the sequel.

6.1. 1D Helmholtz. For the sake of completeness, in this section we employ the multilevel projection method to solve the 1D Helmholtz problem used to generate Table 3.1. Following [8], for multigrid preconditioner F-cycle multigrid is used with one pre- and postsmoothing. The smoother is point-Jacobi smoother with underrelaxation $\omega_{JAC} = 0.5$. We consider problems with various wavenumbers $k = 20, 50, \dots, 500$. The coarsest level for both multilevel Krylov and multigrid method consists of only

one interior grid point.

At every level of the multilevel Krylov method, flexible GMRES (FGMRES) [16] is used to solve the preconditioned Galerkin system. The use of FGMRES is important because preconditioners used at each level are not constant due to (inner) FGMRES iterations in solving the corresponding Galerkin system. In principle it is not necessary to use the same number of FGMRES iterations at each level. As observed in [4], it is the accuracy of solving the Galerkin system at the second level which is of importance. Convergence results are shown in Tables 6.1–6.3. The notation MG-MK(6,2,2), e.g., means that 6 FGMRES iterations are employed at level $j = 2$, 2 at level $j = 3$ and 2 at level $j = 4, \dots, m - 1$. At level $j = m$ the coarse-grid problem is solved exactly.

Results in Tables 6.1–6.3 suggest that the convergence is in general almost independent of the grid size h . The number of iterations to reach convergence (i.e., if the residual is reduced by 6 orders of magnitude) only mildly increases with an increase in the wavenumber k . Compared to the results in Table 3.1, these results are worse than the ideal situation where the coarse-grid problem at the second level is solved exactly. These convergence results, however, are very much improved compared to the results if no projection method is incorporated (i.e., MG).

TABLE 6.1

Number of GMRES iterations for 1D Helmholtz problems with constant wave number. g/w means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(6,2,2). MG is shown between parentheses

g/w	k :				
	20	50	100	200	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 means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(8,2,2) and MG-MK(8,2,1) (between parentheses)

g/w	k :				
	20	50	100	200	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)

TABLE 6.3

Number of GMRES iterations for 1D Helmholtz problems with constant wave number. g/w means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(6,4,2). The ℓ_2 norm of errors are shown between parentheses

g/w	k :				
	20	50	100	200	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)

The significance of the number of iterations at the second level to the convergence of the multigrid-multilevel Krylov method can be seen also in Tables 6.1–6.3. While with MG-MK(8,2,2) the convergence is slightly improved as compared to MG-MK(6,2,2), a significant improvement is hardly gained with MG-MK(6,4,2) (Table 6.3). We also observe that the use of only *one* FGMRES iteration at level $j \geq 4$ is sufficient (see figures between parentheses in Table 6.2).

The last results from our 1D Helmholtz test problem is related to the quality of the approximate solution produced by FGMRES at convergence. In order to do this, we compute the error between the approximate solution of MG-MK at convergence and the solution obtained from a sparse direct method. The typical ℓ_2 norm of the error is shown between parentheses in Table 6.3. For all cases, the ℓ_2 norm of the error falls below 10^{-5} .

6.2. 2D Helmholtz. Two-dimensional Helmholtz problems in a square domain with constant wavenumbers are presented in this section. At the boundaries, the first order radiation conditions are imposed. We consider problems where a source is generated in the middle of the domain. The discretized Helmholtz equation is right preconditioned by the discrete shifted Laplacian. The resultant preconditioned linear system is then solved by using the multigrid-multilevel Krylov method.

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

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

associated with the grid points at level $j = 1$. The set of grid points Ω_H corresponding to the coarse-grid level $j = 2$ is determined as follows. We set $(x_1, y_1) \in \Omega_H$ to coincide with $(x_1, y_1) \in \Omega_h$. This is 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 $H = 2h$:

$$\Omega_H := \{(x, y) | x = x_{i_x} = (2i_x - 1)h, y = y_{i_y} = (2i_y - 1)h, i_x = 1, \dots, N_{x,H}, i_y = 1, \dots, N_{y,H}\}.$$

It has been shown in [8] that this coarsening leads to a good multigrid method for the shifted Laplacian preconditioner. Moreover, with respect to the multilevel Krylov method, this coarsening strategy will also lead to larger projection subspaces compared to, e.g., if $(x_1, y_1) \in \Omega_H$ coincides with $(x_2, y_2) \in \Omega_h$ (Figure 6.1 (right)). As shown in this figure, for example, starting with 7×7 grid points at the finest level, the second coarsening approach leads to only 9 projection vectors, while the former approach results in 16 projection vectors. Only for even number of grid points in each direction that both approaches lead to the same number of projection vectors. This eventually leads to possible projection of 16 small eigenvalues.

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 [19], for level 2 to level 1 (see Figure 6.3(a) for the meaning

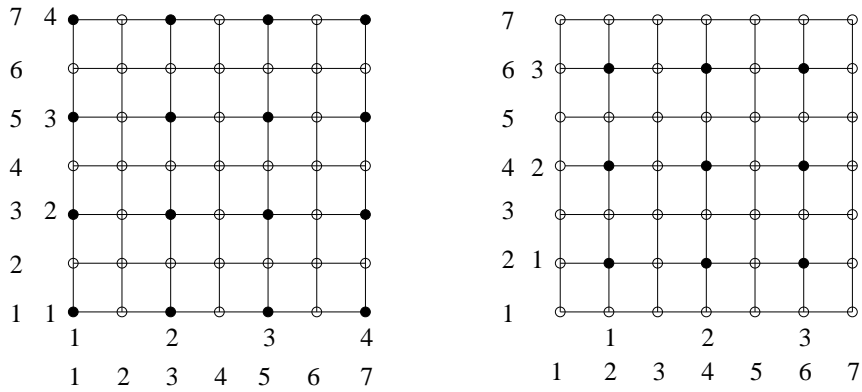


FIG. 6.1. Fine (white circles) and coarse (black circles) grid selections in 2D multigrid. Black circles also coincide with the fine grids. Coarsening as depicted in the left figure leads to both better multigrid methods for the shifted Laplacian and larger projection subspaces

of the symbols):

$$I_H^h v^{(1)}(x, y) = \begin{cases} v^{(2)}(x, y) & \text{for } \bullet \\ \frac{1}{2}[v^{(2)}(x, y-h) + v^{(2)}(x, y+h)] & \text{for } \square \\ \frac{1}{2}[v^{(2)}(x-h, y) + v^{(2)}(x+h, y)] & \text{for } \triangle \\ \frac{1}{4}[v^{(2)}(x-h, y-h) + v^{(2)}(x-h, y+h) \\ + v^{(2)}(x+h, y-h) + v^{(2)}(x+h, y+h)] & \text{for } \circ. \end{cases} \quad (6.1)$$

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

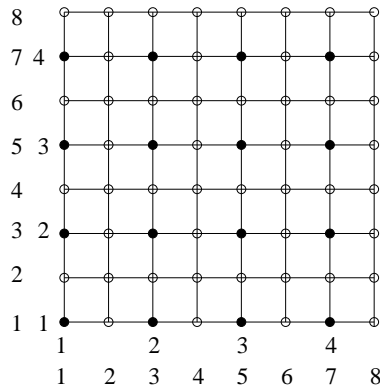


FIG. 6.2. Fine (white circles) and coarse (black circles) grid selections in 2D multigrid. Black circles also coincide with the fine grids. Coarsening as depicted in the left figure leads to both better multigrid methods for the shifted Laplacian and larger projection subspaces

points, which are summarized 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))

$$I_H^h v^{(1)}(x, y) = \begin{cases} v^{(2)}(x, y) & \text{for } \bullet \\ \frac{1}{2}[v^{(2)}(x, y-h) + v^{(2)}(x, y+h)] & \text{for } \square \\ v^{(2)}(x-h, y) & \text{for } \triangle \\ \frac{1}{2}[v^{(2)}(x-h, y-h) + v^{(2)}(x-h, y+h)] & \text{for } \circ. \end{cases} \quad (6.2)$$

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

$$I_H^h v^{(1)}(x, y) = \begin{cases} v^{(2)}(x, y) & \text{for } \bullet \\ v^{(2)}(x, y-h) & \text{for } \square \\ \frac{1}{2}[v^{(2)}(x-h, y) + v^{(2)}(x+h, y)] & \text{for } \triangle \\ \frac{1}{2}[v^{(2)}(x-h, y-h) + v^{(2)}(x+h, y-h)] & \text{for } \circ. \end{cases} \quad (6.3)$$

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

$$I_H^h v^{(1)}(x, y) = \begin{cases} v^{(2)}(x, y) & \text{for } \bullet \\ v^{(2)}(x, y-h) & \text{for } \square \\ v^{(2)}(x-h, y) & \text{for } \triangle \\ v^{(2)}(x-h, y-h) & \text{for } \circ. \end{cases} \quad (6.4)$$

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

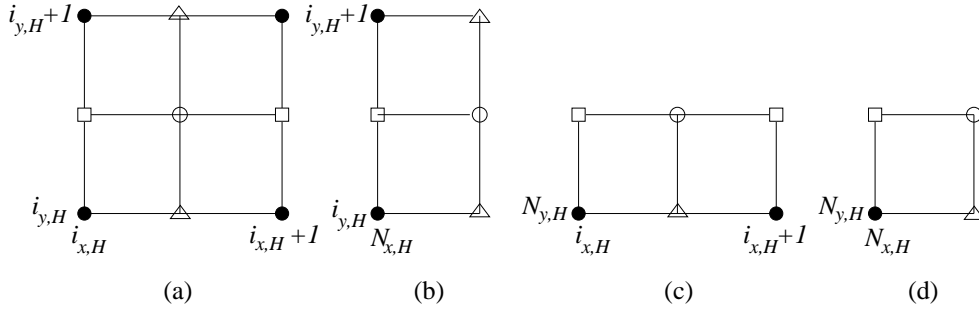


FIG. 6.3. Fine (white colored) and coarse (black colored) grid selection indicating the bilinear interpolation in 2D multigrid. Black circles (\bullet) coincide with the fine grids

Convergence results are shown in Tables 6.4–6.8 for various wavenumbers and different numbers of FGMRES iterations in the projection steps of the multilevel Krylov method. Different numbers of grid points per wavelength are used to resolve the solutions. For multigrid steps, F-cycle with one point-Jacobi pre- and postsmoother is used. The relaxation factor is $\omega_{JAC} = 0.5$. This is an optimal multigrid components that is used in [8]. Following the observation in the previous section, we use only one FGMRES iteration for projection steps at level $j \geq 4$. Convergence of the multilevel Krylov method is assumed if the relative residual of FGMRES at level $j = 1$ has been decreased by six orders of magnitude.

From Tables 6.4–6.8, we observe that for low grid resolutions (e.g., 15 grid points per wavelength) the convergence is mildly dependent of the wavenumber k , with only

small proportionality constant. The convergence is likely improved if more grid points are used. The positive effect of a grid refinement becomes more pronounced in high wavenumber cases. (This is actually somewhat desirable since for a second order discretization scheme, the approximation error is proportional to k^3h^2 ; see, e.g., [10] and [11].) As shown in the tables, with higher grid resolutions, the convergence of the multigrid-multilevel Krylov (MG-MK) method is practically independent of the wavenumber k ; see also Figures 6.4–6.6 for comparison with MG. (Note that in [8], Bi-CGSTAB is used, instead of GMRES.)

TABLE 6.4

Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(4,2,1)

g/w	k:							
	20	40	60	80	100	120	200	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

TABLE 6.5

Number of GMRES iterations for 2D Helmholtz problems with constant wave number. g/w means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(5,2,1)

g/w	k:							
	20	40	60	80	100	120	200	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 means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(6,2,1)

g/w	k:							
	20	40	60	80	100	120	200	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 means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(8,2,1)

g/w	k:							
	20	40	60	80	100	120	200	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 means “#grid points per wavelength”. Multilevel Krylov method with MG-MK(4,3,1)

g/w	k:							
	20	40	60	80	100	120	200	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

Comparing different numbers of Krylov iterations for coarse-grid problems, from Tables 6.4–6.8 it appears that MG-MK(8,2,1) lead to a more efficient method in terms of number of iterations; it converges faster for all wavenumbers and grid resolutions used. If one is more concerned with the number of iterations to reach convergence, one can use many iterations at the level $j = 3$. To add more iterations at the level $j > 3$ does not seem to be effective enough to improve the convergence.

From the first sight, the multigrid-multilevel Krylov method defined in Algorithm 1 and 2 combined seem to require excessive arithmetics. In order to clearly see the convergence acceleration of the method with respect to the total arithmetic operations, we compare the MG-MK with MG (i.e., FGMRES with multigrid-based preconditioner without multilevel projection, the original method of [8]) in terms of CPU time. For CPU time, we measure the elapsed time during the FGMRES iteration on a Pentium 4 machine, and as well the initialization phase. As both methods are implemented in MATLAB, we used the command `tic ... toc` to measure the elapsed time. The results are plotted in Figures 6.4–6.6. In terms of CPU time, MG converges faster for low wavenumbers, but MG-MK outperforms MG for large wavenumbers. Also, the gain in the number of iterations for MG-MK(8,2,1) pays off for high wavenumber cases, where it now turns to be also efficient in terms of CPU time. For wavenumber $k = 300$, we could not run MG until convergence because of memory limitation. To give an illustration, for $k = 300$, with 30 gridpoints per wavelength, the linear system to be solved involves $2.25 \cdot 10^6$ complex-valued unknowns, which limits the memory allocation for storing Arnoldi vectors. Restarting GMRES in this case does not help. With full GMRES, we terminate the iteration after 86 iterations. At this stage, the computed residual is only $6.55 \cdot 10^{-4}$, and CPU time spent is already about 23,000 second. Furthermore, even though in the MG-MK algorithm the initialization phase consists also of computing coarse-grid informations associated with matrices $A^{(j)}$ and $B^{(j)}$, and not only $M^{(j)}$ as in case of MG, this extra computation does not significantly contribute to the total initialization time, as shown in the lower part of Figures 6.4–6.6 (right).

Also of importance is the linear increase of the number of iterations observed in MG, which amounts to considerably large memory to store the GMRES vectors. With nearly wavenumber-independent convergence, such a memory requirement is less stringent in the case of the multigrid-multilevel Krylov method.

7. Conclusions. In this paper, we have discussed an application of the multilevel Krylov method for solving the 2D Helmholtz equation, which is based on multilevel projection 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, a

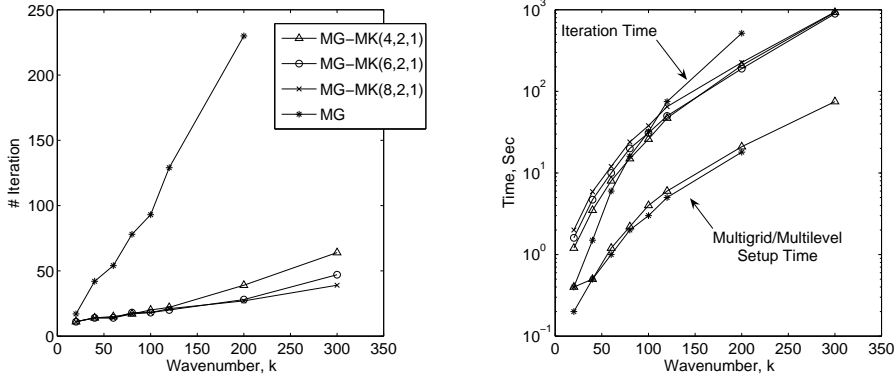


FIG. 6.4. Number of iterations and CPU time for GMRES with multigrid applied to the shifted Laplacian preconditioner (MG) and multigrid-multilevel Krylov method (MG-MK). 15 grid points per wavelength

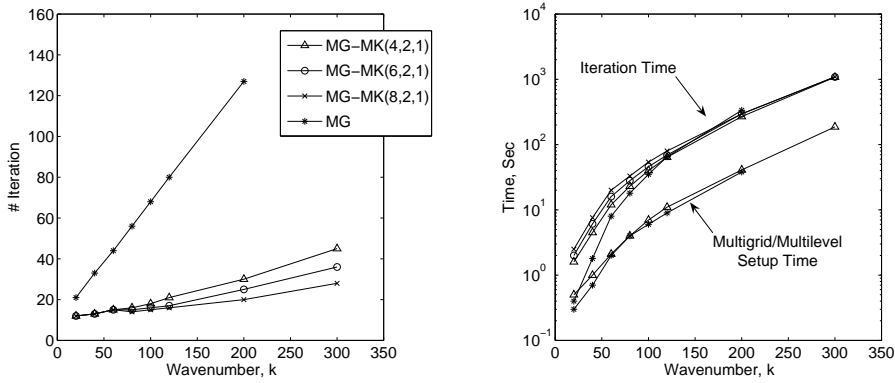


FIG. 6.5. Number of iterations and CPU time for GMRES with multigrid applied to the shifted Laplacian preconditioner (MG) and the multigrid-multilevel Krylov method (MG-MK). 20 grid points per wavelength

few Krylov iterations are used to solve the projected Galerkin (coarse-grid) problems, which consist also of preconditioner solves. The preconditioner solves are done by one multigrid iteration, whose maximum level is reduced according the projection level. In the implementation, the projection matrix is set equal to the interpolation matrix in multigrid.

Numerical experiments have been performed on the 1D and 2D Helmholtz equation with constant wavenumber. With this method, we observed h -independent and k -independent convergence. This considerable improvement in the convergence rate leads to a speed up in CPU time as well, as compared to Krylov methods with multigrid-based preconditioner alone.

Finally, this multilevel Krylov method consists of several ingredients, like a preconditioner for Krylov iterations, restriction and prolongation operators, 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

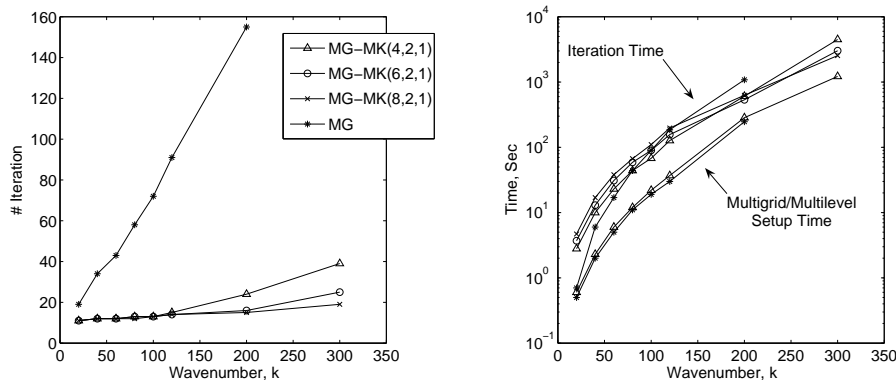


FIG. 6.6. Number of iterations and CPU time for GMRES with multigrid applied to the shifted Laplacian preconditioner (MG) and multigrid-multilevel Krylov method (MG-MK). 30 grid points per wavelength

method for the Helmholtz equation. Nevertheless, other choices or new developments in those methods can be easily implemented as well in our multilevel Krylov framework (to obtain an even faster convergence).

REFERENCES

- [1] M. Eiermann, O. G. Ernst, and O. Schneider. Analysis of acceleration strategies for restarted minimal residual methods. *J. Comput. Appl. Math.*, 123:261–292, 2000.
- [2] 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.*, 22:1291–1315, 2001.
- [3] B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Math. Comput.*, 31:629–651, 1977.
- [4] Y. A. Erlangga and R. Nabben. Deflation and balancing preconditioners for Krylov subspace methods applied to nonsymmetric matrices. *SIAM J. Matrix Anal. Appl.* (submitted), 2007.
- [5] Y. A. Erlangga and R. Nabben. Multilevel projection-based nested Krylov iteration for boundary value problems. *SIAM J. Sci. Comput.*, to appear.
- [6] Y. A. Erlangga, C. W. Oosterlee, and C. Vuik. A novel multigrid-based preconditioner for the heterogeneous Helmholtz equation. *SIAM J. Sci. Comput.*, 27:1471–1492, 2006.
- [7] Y. A. Erlangga, C. Vuik, and C. W. Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Appl. Numer. Math.*, 50:409–425, 2004.
- [8] Y. A. Erlangga, C. Vuik, and C. W. Oosterlee. Comparison of multigrid and incomplete LU shifted-Laplace preconditioners for the inhomogeneous Helmholtz equation. *Appl. Numer. Math.*, 56:648–666, 2006.
- [9] J. Frank and C. Vuik. On the construction of deflation-based preconditioners. *SIAM J. Sci. Comput.*, 23:442–462, 2001.
- [10] F. Ihlenburg and I. Babuska. Dispersion analysis and error estimation of Galerkin finite element methods for the Helmholtz equation. *Int. J. Numer. Methods Engrg.*, 38:3745–3774, 1995.
- [11] F. Ihlenburg and I. Babuska. Finite element solution of the Helmholtz equation with high wave number. Part I: The h -version of the FEM. *Comput. Math. Appl.*, 30(9):9–37, 1995.
- [12] J. A. Meijerink and H. A. van der Vorst. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. *Math. Comp.*, 31 (137):148–162, 1977.
- [13] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Anal. Appl.*, 16:1154–1171, 1995.
- [14] R. Nabben and C. Vuik. A comparison of deflation and the balancing preconditioner. *SIAM J. Sci. Comput.*, 27:1742–1759, 2006.
- [15] R. A. Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM J. Numer. Anal.*, 24:355–365, 1987.

- [16] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Sci. Comput.*, 14:461–469, 1993.
- [17] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, 2003.
- [18] J. Tang, R. Nabben, C. Vuik, and Y. A. Erlangga. Theoretical and numerical comparison of various projection methods derived from deflation, domain decomposition and multigrid methods. Technical Report 07-04, Delft Institute of Applied Mathematics, available under: ta.twi.tudelft.nl, submitted, 2007.
- [19] U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, New York, 2001.
- [20] M. B. van Gijzen, Y. A. Erlangga, and C. Vuik. Spectral analysis of the shifted Laplace preconditioner. *SIAM J. Sci. Comput.*, to appear, 2007.
- [21] R. S. Varga. *Gerschgorin and His Circles*. Springer-Verlag, Berlin-Heidelberg, 2004.