Adaptive Multilevel Krylov Methods

René Kehl, Reinhard Nabben,
and Daniel B. Szyld

Report 17-06-24
June 2017

Department of Mathematics
Temple University
Philadelphia, PA 19122

This report is available in the World Wide Web at
`http://www.math.temple.edu/~szyld`

ADAPTIVE MULTILEVEL KRYLOV METHODS*

RENÉ KEHL†, REINHARD NABBEN†, AND DANIEL B. SZYLD‡

Abstract. Inexact (variable) preconditioning of multilevel Krylov methods (MK methods) for the solution of linear systems of equations is considered. MK methods approximate the solution of the local systems on a subspace using a few, but fixed, number of iteration steps of a preconditioned flexible Krylov method. In this paper, using the philosophy of inexact Krylov subspace methods, we use a computable criterion to choose the number of iterations needed on each level to achieve a desired tolerance. This adaptive technique chooses the number of iterations on all levels during the iteration. Numerical experiments for the Poisson and the convection-diffusion equations illustrate the efficiency and robustness of this technique.

Key words. Multilevel Krylov methods, flexible GMRES, inexact Krylov subspace methods, inexact preconditioning

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

1. Introduction.

We consider the iterative solution of linear systems

$$Ax = b \quad A \in \mathbb{R}^{n\times n}, \quad x, b \in \mathbb{R}^n, \tag{1.1}$$

where A is a large, sparse, and possibly nonsymmetric matrix. Krylov subspace methods in combination with multilevel preconditioners are the methods of choice for the approximate solution of (1.1); see, e.g., [5, 24].

Erlangga and Nabben [7, 8, 9] introduced multilevel Krylov methods (MK methods). This type of multilevel method was inspired by the idea of shifting some eigenvalues that are close to zero farther away from zero, e.g., to have value one. This shifting is performed in the spirit of deflation techniques, i.e., one has to solve smaller systems on a subspace, or coarse-grid systems. This in turn leads to the multilevel structure. MK methods, as originally presented, approximate the solution of the subspace systems by performing just a few iterations of a preconditioned flexible Krylov method, such as flexible GMRES [18].

In [7], the MK methods are analyzed, and their potential is demonstrated for the 2D Poisson and the 2D convection-diffusion equations. The latter is an example with a nonsymmetric matrix of coefficients. By using a simple piece-wise constant interpolation to construct the subspace or coarse-grid system, an h-independent convergence for the Poisson equation, and an almost h- and Pe-independent convergence for the convection-diffusion equation are observed, where h is the mesh width and Pe is the Péclet number.

The multilevel Krylov technique was combined with the shifted Laplacian preconditioner [10, 11, 12] to solve high wavenumber 2D Helmholtz equations [8]. There, the linear system is preconditioned by the shifted Laplacian first, and then the shifting of some eigenvalues is performed. Numerical results show that not only the convergence is almost independent of h, but also only mildly dependent on the wavenumber. In

*This version dated 24 June 2017

†TU Berlin, Institut für Mathematik, MA 3-3, Straße des 17. Juni 136, D-10623 Berlin, Germany ({kehl,nabben}@math.tu-berlin.de).

‡Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, Pennsylvania 19122-6094, USA (szyld@temple.edu). Supported in part by the U.S. National Science Foundation under grants DMS-1115520 and DMS-1418882 and the U.S. Department of Energy under grant DE-SC 0016578.

[19, 20] the shift is performed first followed by the shifted Laplacian preconditioner. This approach can be seen as a preconditioned multilevel Krylov method. This combination leads to a powerful method for 3D realistic Helmholtz equations with high wave numbers, further illustrating the potential of MK methods.

The idea of applying Krylov subspace methods on different levels or using them as preconditioners was also considered elsewhere. Notay and Vassilevski [17] propose a recursive Krylov-based multigrid cycle, which uses flexible Krylov methods to solve the coarse-grid systems in a classical (algebraic) multigrid step that includes and depends on smoothing steps. Elman, Ernst, and O'Leary [4] use Krylov subspace methods to solve the coarse-grid system at the intermediate levels of multigrid. In this case, the Krylov subspace method is used as a replacement for the smoother. In [21], flexible Krylov subspace methods that are preconditioned by Krylov subspace methods themselves are studied, but not in the context of a multilevel structure.

In the multilevel Krylov methods described earlier, the number of iterations performed at each level is fixed *a priori*. In other words, one can consider these methods as Krylov subspace methods with inexact preconditioning, in which the level of inexactness is prescribed in an *ad hoc* manner. Inspired by the analysis of inexact Krylov methods in [22, 23], in this paper we consider varying the number of iterations, relaxing the tolerance of each application of the inexact preconditioner, while obtaining the same overall accuracy. We show that for the MK methods the number of iteration steps on the subspace systems can be reduced during the fine level iteration. Thus, we provide a theory that guarantees that with the prescribed level of inexactness, the MK method converges to the desired accuracy.

The numerical results in [7, 8, 19] demonstrate that the number of iterations on the second level is important. Typically, 8 iterations on the second level are used, then 4 on the next level, while only 1 iteration on all other levels. Here we produce a computable criterion to fix the number of iterations needed on each level. This adaptive technique reduces the number of iterations on the second and third level during the iteration. For this adaptive variant of the MK method we thus have a complete convergence theory. Examples for Poisson and convection-diffusion equations illustrate the efficiency and robustness of this technique. An experimental analysis of the dependency of the convergence on the constant used in the method is also provided.

The paper is organized as follows. In section 2, we briefly describe the multilevel Krylov methods. In section 3, we begin with a few comments on general Krylov methods, then we describe inexact Krylov subspace methods and the new results for adaptive multilevel Krylov methods. Numerical results in section 4 illustrate the performance of the proposed adaptive method.

2. Multilevel Krylov Methods. We begin by briefly describing the MK methods; for more details see [6, 9]. For the sake of simplicity, in what follows we mainly discuss the two-level method; the multilevel method is obtained by a recursive application of the two-level method on the corresponding coarse-grid solve. We start with a rectangular matrix $Z \in \mathbb{R}^{n\times r}$ with $r \leq n$, i.e., describing an operator

$$Z : \mathbb{R}^r \mapsto \mathbb{R}^n.$$

In the multilevel language the operator Z is the *prolongation* or *interpolation* operator, which transfers variables from a coarse level to a fine one. The *restriction* operator

$$Z^T : \mathbb{R}^n \mapsto \mathbb{R}^r$$

does the opposite; it transfers some variables from a fine level to a coarse one. The so-called coarse-level (grid) matrix or subspace matrix is then defined by the Galerkin operator

$$E = Z^T AZ \in \mathbb{R}^{r\times r}.$$

Here we assume that E is nonsingular. Note that instead of Z^T any another $r \times n$ matrix Y^T can be used(see [6] for details). We define the operator

$$P_D := I - AZE^{-1}Z^T. \tag{2.1}$$

Note that $P_D^2 = P_D$, thus P_D is the (oblique) projection operator onto the orthogonal complement of the kernel of Z^T along the range of AZ. We then consider the (projected or) preconditioned system

$$P_D Ax = P_D b. \tag{2.2}$$

A two-level Krylov method is obtained if a Krylov subspace method such as GMRES is applied to (2.2), this being one level, while the solution of the linear system with coefficient matrix E is the second level.

The convergence of Krylov subspace methods is driven by the eigenvalues of the coefficient matrix - for (2.2), this $\sigma(P_D A)$ - as well as the corresponding eigenvectors in the nonsymmetric case; see, e.g., [14]. Related to this feature, the following observation is useful to gain some insight on the convergence of the multilevel Krylov method. As shown in [15], for any full rank Z, the spectrum of $P_D A$ is

$$\sigma(P_D A) = \{0, \ldots, 0, \mu_{r+1}, \ldots .\mu_n\}. \tag{2.3}$$

We first note that the values of μ_i, $i = r + 1, \ldots, n$, depend on the specific choice of Z. The operator P_D is known as the deflation operator; see, e.g., [7, 13]. If A is symmetric positive definite and the columns of Z consist of orthogonal eigenvectors of A then the eigenvalues μ_j of the preconditioned operator in (2.3) are eigenvalues of A. Hence, some of the eigenvalues of A are deflated out of the spectrum of A, or in other words they are shifted to zero. A Krylov subspace method that works for singular matrices can be used to solve (2.2). Typically, only the eigenvalues $\mu_{r+1}, \ldots, \mu_n$, are those which enter into the spectral error bounds for such a Krylov subspace method.

In order to obtain the zeros in (2.3), the linear systems corresponding to the product with E^{-1}, i.e., the subspace systems, have to be computed exactly. If they are approximated, or solved inexactly, i.e., if the product with E^{-1} is replaced by one with $E^{-1} + P$, where P is a small perturbation, the zeros in (2.3) are expected to be small numbers. This would slow down the convergence of the Krylov method.

To allow inexact subspace solves one can consider the operator

$$P_N := I - AZE^{-1}Z^T + \lambda ZE^{-1}Z^T, \tag{2.4}$$

where λ can be any real number, e.g., the number one. Note that P_N is no longer a projection. We now have that (in exact arithmetic)

$$\sigma(P_N A) = \{\lambda, \ldots, \lambda, \mu_{r+1}, \ldots .\mu_n\}. \tag{2.5}$$

Thus, some of the eigenvalues are shifted to λ. If the subspace method is solved inexactly, only a small perturbation might be added to the eigenvalues. Of course, P_N can be written as

$$P_N := I - (A - \lambda I)ZE^{-1}Z^T, \tag{2.6}$$

so just one subspace solve is needed, and not two, as it might appear from (2.4).

The concept of multilevel Krylov method is as follows: We use P_N as preconditioner for a flexible Krylov subspace method, and to solve the subspace system with the matrix E we use a few steps of the same flexible Krylov subspace preconditioned by a preconditioner of the same form as P_N in (2.6) (just of smaller dimension). To apply this second level preconditioner a subspace system has to be solved again, which is done in the same manner. Thus, we obtain recursively the multilevel structure. In [7, 8, 19] typically 8 or 4 iterations are used on the second level, 2 iterations on third level and 1 iteration on all other levels. This is denoted by MK(8,2,1) or MK(4,2,1), respectively.

Flexible Krylov methods need right preconditioning. Therefore we define the operator

$$Q_N := I - ZE^{-1}Z^TA + \lambda ZE^{-1}Z^T = I - ZE^{-1}Z^T(A - \lambda I) \tag{2.7}$$

and solve the system

$$AQ_N\tilde{x} = b \quad \text{with} \quad x = Q_N\tilde{x}. \tag{2.8}$$

We note that since

$$P_NA = AQ_N,$$

the spectral properties of P_NA and AQ_N are the same, and the discussion on shifting 0 to $\lambda \in \sigma(AQ_N)$ is thus valid in this case as well.

Finally, we note that the techniques described above can be combined with another (say ordinary) preconditioner in two ways. The first alternative as used in [8, 9], is the following. The matrix A can be preconditioned, so that A is replaced by AM or MA, where M is the preconditioner. Secondly, the shifted systems P_NA or AQ_N can be preconditioned, so that we obtain MP_NA or AQ_NM, respectively. More details about the implementation of MK methods can be found in [9].

3. Inexact Krylov methods. Our goal is to solve the system (1.1) more efficiently by (right) preconditioning with a given nonsingular matrix Q_N as in (2.7), that might not be given explicitly, i.e., we solve approximately (2.8). In the iteration process, when applying Q_N, linear systems with E have to be approximately solved. This is done by a few iteration steps of a preconditioned flexible Krylov subspace method. Thus the preconditioner Q_N varies from step to step and Q_N can not be applied exactly (i.e., it is not given explicitely). In this paper, we choose flexible GMRES (FGMRES) [18] as inner and outer solver of the linear systems in (2.8).

In the rest of this section, we begin with a very brief review on general Krylov subspace methods, followed by a discussion on inexact methods, and its application to the multilevel Krylov method.

3.1. General Krylov subspace methods. Let A be nonsingular and x_0 an initial vector for the solution of (1.1). Krylov subspace methods find at the m-th iteration an approximation x_m in the Krylov subspace $K_m(A, v_1)$, the space spanned by the vectors $v_1, Av_1, \ldots, A^{m-1}v_1$, satisfying some optimality condition, where v_1 is usually chosen as the normalized initial residual $r_0 = b - Ax_0$, i.e., $r_0 = \beta v_1$, $\beta = \|r_0\|$. An orthonormal basis $\{v_1, \ldots, v_m\}$ of this Krylov subspace is usually obtained by the Arnoldi procedure, in which the main operation is the matrix-vector product Av_i.

These basis vectors are collected as columns of the matrix V_m. This gives rise to the the Arnoldi relation

$$AV_m = V_{m+1}H_m, \tag{3.1}$$

where $H_m \in \mathbb{R}^{m+1,m}$ is an upper Hessenberg matrix, and we have assumed that $\|v_k\| = 1$, for all $k \leq m$. For GMRES and its variants, the optimality condition is minimal residual, i.e.,

$$\min_{x \in K_m(A,v_1)} \|r_0 - Ax\| = \min_{y \in \mathbb{R}^m} \|\beta e_1 - H_m y\|, \tag{3.2}$$

where e_1 is the first Euclidean vector.

3.2. Inexact Krylov subspace methods. We review here results from [21, 22, 23] on inexact Krylov subspace methods, which we use in our application to Multilevel Krylov methods.

When the matrix-vector-multiplication Av_i in the Arnoldi procedure is not performed exactly, it can be interpreted as being of the form $(A + G_i)v_i$, where $G_i \in \mathbb{R}^{n \times n}, i = 1, \ldots, m$ can be thought of as pertubations of A. The analogue to (3.1) is an *inexact Arnoldi relation*

$$\begin{aligned} AV_m + [G_1 v_1, \ldots, G_m v_m] &= V_{m+1}H_m, \text{i.e.,} \\ [(A + G_1)v_1, \ldots, (A + G_m)v_m] &= V_{m+1}H_m. \end{aligned} \tag{3.3}$$

With $\mathcal{E}_m = \sum_{k=1}^m G_k v_k v_k^T$, (3.3) can be written in the compact form

$$(A + \mathcal{E}_m)V_m = V_{m+1}H_m.$$

In particular this means that at each iteration the solution given by the Krylov subspace method with inexact matrix-vector products is an element of a perturbed Krylov subspace $K_m(A + \mathcal{E}_m, v_1)$.

One problem that arises during the iteration process is that the residual r_m is no longer an available quantity, since the "true residual" should be calculated with help of the unperturbed matrix A, i.e.,

$$r_m = r_0 - AV_m y_m = r_0 - V_{m+1}H_m y_m + \mathcal{E}_m V_m y_m, \tag{3.4}$$

where y_m is the solution of the minimization problem (3.2), and note that neither $\mathcal{E}_m$ nor AV_m are available. The "computed residual" that is carried out by the inexact method is thus

$$\tilde{r}_m = r_0 - (A + \mathcal{E}_m)V_m y_m = r_0 - V_{m+1}H_m y_m = r_m - \mathcal{E}_m V_m y_m. \tag{3.5}$$

Therefore the difference $\|r_m - \tilde{r}_m\|$ can be bounded as follows

$$\|r_m - \tilde{r}_m\| = \|\mathcal{E}_m V_m y_m\| \leq \sum_{k=1}^m \|G_k\| |y_m^{(k)}|,$$

where $y_m^{(k)}$ is the k-th component of the iterate y_m.

The following results come from [22].

LEMMA 3.1. *Assume that m iterations of the inexact Arnoldi method have been carried out and let y_m be the solution of the GMRES minimization problem* (3.2). *Then, for any $k = 1, \ldots, m$,*

$$|y_m^{(k)}| \leq \frac{1}{\sigma_m(H_m)} \|\tilde{r}_{k-1}\|,$$

where $\sigma_m(H_m)$ denotes the smallest singular value of H_m, and $\tilde{r}_{k-1}$ is as in (3.5).

THEOREM 3.2. *Let $\varepsilon > 0$. Let r_m be the GMRES residual after m iterations of the inexact Arnoldi method, given in* (3.4). *Under the same notation and hypothesis of Lemma 3.1, if for every $k \leq m$,*

$$\|G_k\| \leq \frac{\sigma_m(H_m)}{m} \frac{1}{\|\tilde{r}_{k+1}\|} \varepsilon, \tag{3.6}$$

then $\|r_m - \tilde{r}_m\| \leq \varepsilon$. Moreover, if

$$\|G_k\| \leq \frac{1}{m\kappa(H_m)} \frac{1}{\|\tilde{r}_{k+1}\|} \varepsilon, \tag{3.7}$$

where $\kappa(H_m)$ denote the condition number of H_m, then $\|(V_{m+1}H_m)^T r_m\| \leq \varepsilon$.

Conditions (3.6) and (3.7) are of the form

$$\|G_k\| \leq c_m \cdot \frac{\varepsilon}{\|\tilde{r}_{k-1}\|}$$

While in some cases, $c_m \approx 1$ is a reasonable choice for various problems [1, 2], there are cases, where c_m needs to contain further information on A or H_m. In most of applications, it is not possible to calculate the exact value of c_m since the singular values of H_m at the m-th step are not known beforehand. Nevertheless, there are some estimation strategies one could use, cf. [22, §5, §9]. We discuss this issue further for our application to multilevel Krylov methods, in our numerical experiments in section 4.

3.3. Application to multilevel Krylov methods. We begin by considering the application of a right multilevel preconditioner of the form (2.7), with $\lambda > 0$, $Z \in \mathbb{R}^{n \times r}$, $r \ll n$ and $E = Z^T A Z$, the coarse grid Galerkin matrix. For a given $v_k \in \mathbb{R}^n$, the (approximate) product $Q_N v_k$ proceeds in three steps as follows.

1. Calculate $w_k = Z^T(\lambda I - A)v_k$
2. Solve $Et = w_k$ inexactly to a given tolerance τ, producing the approximation $\tilde{t}_k$, which leads to the coarse grid residual $q_k = w_k - E\tilde{t}_k$ so that $\|q_k\| < \tau$
3. Calculate $\tilde{z}_k = v_k + Z\tilde{t}_k \approx Q_N v_k$

For a two-level method, the system in step 2. is solved with a direct solver, i.e., with $\tau = 0$. In the multilevel case, we set $\tilde{A} = E, \tilde{Z} \in \mathbb{R}^{r \times s}, s \ll r$ and apply a preconditioner of the form of Q_N and perform some iterations with flexible GMRES (FGMRES). One either chooses *a priori* the number of FGMRES iterations, as in [7, 8, 9], or by the appropriate choice of the tolerance τ, as we propose in this paper. We continue until the coarse grid system becomes sufficiently small and solve the coarsest system with a direct solver.

In this section, we develop the criteria on how to choose the tolerance τ for each of the linear systems on the coarse level to guarantee a good fine level performance. This choice will determine the number of FGMRES iterations at each level.

To that end, let $z_k = Q_N v_k$ (the "exact" matrix-vector product), and let $\tilde{z}_k$ be the corresponding approximate matrix-vector product, as per the three steps described above. Let us collect these vectors for $k = 1, \ldots, m$, $\tilde{Z}_m := [\tilde{z}_1, \ldots, \tilde{z}_m]$.

As in the previous subsection, we have the "true" residual

$$r_m = r_0 - AQ_N V_m y_m, \tag{3.8}$$

where y_m comes from the solution of the least squares problem (3.2) in FGMRES. The computed residual is

$$\tilde{r}_m = r_0 - A\tilde{Z}_m y_m. \tag{3.9}$$

Then,

$$r_m - \tilde{r}_m = -A(Q_N V_m - \tilde{Z}_m)y_m = -A\sum_{k=1}^{k} (Q_N v_k - \tilde{z}_k)\, y_m^{(k)}.$$

We now look in detail at the differnce in parenthesis. We have

$$\begin{aligned} Q_N v_k - \tilde{z}_k &= v_k + ZE^{-1}w_k - \tilde{z}_k \\ &= v_k + ZE^{-1}w_k - (v_k + Z\tilde{t}_k) = Z\left(E^{-1}w_k - \tilde{t}_k\right) \\ &= ZE^{-1}\left(w_k - E\tilde{t}_k\right) = ZE^{-1}q_k, \end{aligned}$$

and hence

$$\|r_m - \tilde{r}_m\| \le \|AZE^{-1}\| \sum_{k=1}^{m} \|q_k\| \cdot |y_m^{(k)}|. \tag{3.10}$$

The bound (3.10) gives the desired connection between the allowable tolerance τ for the residual q_k at the coarse level, so that the residual at the fine level can be obtained at the desired accuracy. Using this bound, and the analysis of the form of the components $y_m^{(k)}$ from [22], given in Lemma 3.1, we obtain our main result.

THEOREM 3.3. *Let $\varepsilon > 0$, and assume that m iterations of the inexact Arnoldi method have been carried out for the preconditioned system $AQ_N\bar{x} = b$, $x = Q_N\bar{x}$. Let y_m be the FGMRES minimizer and $\bar{x}_m = x_0 + V_m y_m$ and let $r_m, \tilde{r}_m$ be given as in* (3.8), (3.9). *Let $w_k = Z^T(\lambda I - A)v_k$ and $\tilde{t}_k$ be an approximate solution of $Et = w_k$. If for each iteration $k \le m$ the inner residual $q_k = w_k - E\tilde{t}_k$ satisfies*

$$\|q_k\| \le \frac{\sigma_m(H_m)}{\|AZE^{-1}\|m} \cdot \frac{1}{\|\tilde{r}_{k-1}\|}\varepsilon,$$

it follows that $\|r_m - \tilde{r}_m\| \le \varepsilon$. Moreover, if

$$\|q_k\| \le \frac{1}{\|AZE^{-1}\|m\kappa(H_m)} \frac{1}{\|\tilde{r}_{k-1}\|}\varepsilon,$$

then $\|(V_{m+1}H_m)^T r_m\| \le \varepsilon$.

Theorem 3.3 deals with the residuals of two consecutive levels, such that it can be applied successively in the multigrid context. It can be seen as a global strategy for each (inner or outer) iteration process.

We note that the bound also holds for a residual difference relative to the right hand side, i.e., use $\tilde{\varepsilon} = \|b\| \cdot \varepsilon$ in Theorem 3.3 and obtain the same result.

As discussed in the previous section, these inner bounds may be difficult to calculate exactly. Here the norm of AZE^{-1} is an additional quantity, that has to be estimated. The bound on the inner residual is again of the form

$$\|q_k\| \leq c_m \cdot \frac{\varepsilon}{\|\tilde{r}_{k-1}\|}. \tag{3.11}$$

Due to the proven existence of such a bound, one can use an initial value for c_m and refine that choice if the convergence is too inaccurate, until an appropriate tolerance is achieved.

4. Adaptive MK methods and Numerical examples. The theoretical results of the previous section can be used to devise a strategy to reduce the number of (inner) iterations to solve each coarse grid system in the multilevel method during the outer iteration. In addition to a possible reduction in the number of iterations, this strategy guarantees convergence to tolerance.

Thus, in this section, we investigate first a relaxed adaptive variant of the MK-method which uses the results of Theorem 3.3. Since we want to compare its performance with the basic MK method, in fact, we only allow a maximum number of coarse grid iterations per fine grid-step, say 5. We also investigate a static variant, which is inspired by our theoretical results in the sense that we reduce the number of inner iterations as the method progresses, but we maintain those numbers fixed. This strategy may perform well in practice, but it it is not guaranteed to converge.

We also discuss the possible influence of the estimates of the constant c_m, as defined in Theorem 3.3, on the convergence of the overall method.

In the following, the system

$$AQ\bar{x} = b, \quad Q\bar{x} = x,$$

where Q is the multilevel preconditioner given in (2.7), will be solved iteratively with a flexible GMRES-method.

We compare two basic MK methods with two relaxed MK methods. The first relaxed MK method is the adaptive version that changes the number of coarse grid steps following the strategy given in Theorem 3.3, using (3.11). The second relaxed method is the static version which starts with 5 coarse grid iterations per outer iteration. But after a fixed number of outer iterations the number of coarse grid iterations is reduced to 2 iterations. More specifically, we consider

- **MK(5,5,5)**: Five-level MK method using flexible GMRES. On each coarse grid, there will be 5 iteration steps per one fine grid step.
- **MK(2,2,2)**: Same as the previous case, with only 2 iterations per level.
- **MKRadap**: Five-level MK method starting with 5 coarse grid iterations per fine grid-step and a reduction strategy, that changes the number of coarse grid steps following the strategy given in Theorem 3.3, where the criterion for the inner residual is $\|q_k\| \leq c_m \cdot 10^{-10}/\|\tilde{r}_{k-1}\|$.
- **MKRstat**: Five-level MK method starting with 5 coarse grid iterations per fine grid-step, and after a fixed number of outer iterations (say 10), the method changes to 2 coarse grid iterations per fine grid-step.

For the adaptive method, this means that for the unknown true residual, the following bound holds,

$$\|r_m\| \leq \|\tilde{r}_m\| + \|r_m - \tilde{r}_m\| \leq 2 \cdot 10^{-10} \ .$$

We present various numerical experiments with the two-dimensional Poisson equation, and with the convection-diffusion equation. For the MKRstat version, we choose to change the coarse-grid strategy after 10 (outer) iterations, changing from 5 to 2 coarse grid iterations per fine grid-step.

4.1. 2D Poisson Equation. We start with some experiments with the 2D Poisson equation

$$-\Delta u = f \quad \text{in } \Omega = (0,1)^2$$

with inhomogeneous Dirichlet conditions on the boundary $\partial\Omega$. The discretization with a central difference scheme leads to a symmetric positive definite matrix A. For the multilevel method we choose four coarse levels such that each solution vector on a coarse grid has half the size (rounded down) of the corresponding fine grid solution. The solution on the coarsest grid is carried out with a direct method and the deflation matrix Z result by assembling two neighboring points, i.e., Z has the following form:

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

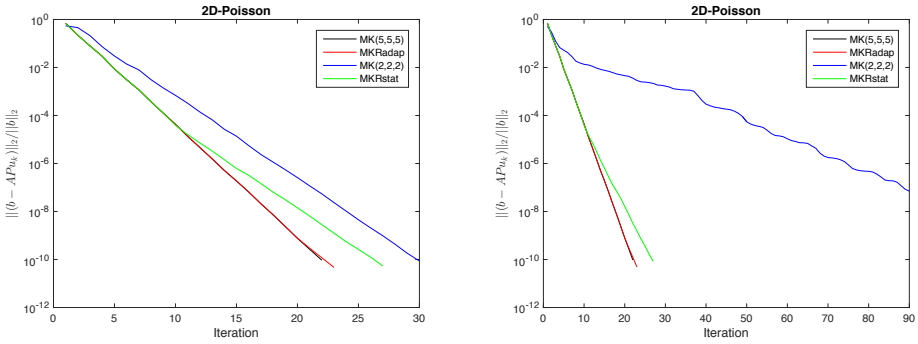
The results of our numerical experiments for this problem for different grid sizes are shown in Figure 4.1. The amount of work as measured by number of direct solves on the coarsest grid, and the number of iterations to converge to a relative residual below $\varepsilon = 10^{-10}$, are given in Table 4.1.

Grid		MK(5,5,5)	MKRadap	MKRstat	MK(2,2,2)
400^2	# direct solves (# It.)	2750(22)	2500(23)[18]	1269(27)	240(30)
	error	$3.9 \cdot 10^{-10}$	$5.4 \cdot 10^{-10}$	$2.9 \cdot 10^{-10}$	$1.4 \cdot 10^{-9}$
600^2	# direct solves (# It.)	2750(22)	2500(23)[18]	1269(27)	- (90)
	error	$5.7 \cdot 10^{-10}$	$7.3 \cdot 10^{-10}$	$7.5 \cdot 10^{-10}$	$1 \cdot 10^{-4}$
800^2	# direct solves (# It.)	2750(22)	2500(23)[18]	1269(27)	- (90)
	error	$7.2 \cdot 10^{-10}$	$1.2 \cdot 10^{-9}$	$7.5 \cdot 10^{-10}$	$2.8 \cdot 10^{-1}$

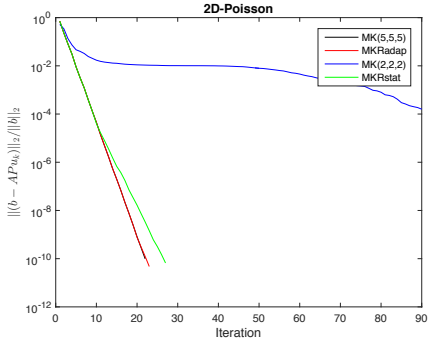
TABLE 4.1

Number of calculated coarsest grid solutions (number of iterations to converge) [iteration number where switching criterion is satisfied], and error $\|x - x_k\|$ of the different MK methods for the 2D Poisson equation.

As it can be observed, for the smallest grid considered, all four versions converge in between 22 and 30 iterations; but for larger grid sizes, the method MK(2,2,2) fails to converge in 90 iterations (set as the maximum allowed). The original version MK(5,5,5) and the adaptive strategy converge essentially in the same number of iterations, while the static version exhibits a certain delay. The computational effort though is varied. The adaptive strategy is about 10% faster than the original MK(5,5,5) method, while the static version is about 50% faster. This is easily explained. While each iteration of the original MK(5,5,5) method uses 5 course iterations per fine grid iteration, the static version reduces this to 2 coarse iterations after



(a) 400^2 grid points (b) 600^2 grid points



(c) 800^2 grid points

FIG. 4.1. *Convergence behavior for the Poisson equation using* 400^2, 600^2, *and* 800^2 *grid points*

10 steps. The adaptive strategy, starts reducing the number of coarse iterations only when the criterion (3.11) is satisfied, which occurs in the 18th iterations (here using $c_m = 10$). Of course, the static strategy works in this case, but this cannot always be guaranteed (as shown in the example in the next section). On the other hand, the adaptive strategy was shown to always converge.

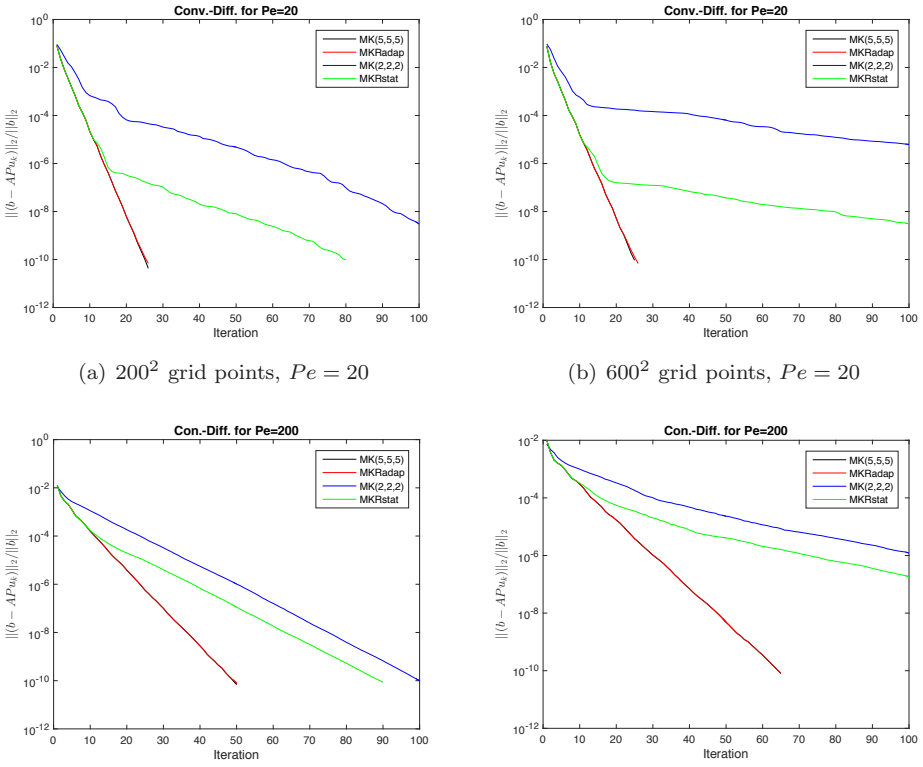
We conclude the section by mentioning that we experimented with larger values of c_m and we were able to observe a further reduction in the computational time, to about 20% of the original method. We explore the influence of the choice of this constant further in the example of the next section.

4.2. The Convection-Diffusion Equation. As a second example we investigate the behavior of the MK methods for the numerical solution of the convection-diffusion equation

$$\frac{\partial u}{\partial y} - \frac{1}{Pe}\Delta u = 0, \quad \text{in } \Omega = (0,1)^2,$$

where Pe is the Péclet number. The boundary conditions are given as follows, $u(x,0) = u(0,y) = 0, u(x,1) = u(1,y) = 1$. The problem is solved in a finite difference context with an upwind scheme as given in [3].

We consider the same four versions of the MK method as described earlier, with the same tolerance $\varepsilon = 10^{-10}$. Here we study the convergence behaviour on two square



(a) 200^2 grid points, $Pe = 20$

(b) 600^2 grid points, $Pe = 20$

(c) 200^2 grid points, $Pe = 200$

(d) 600^2 grid points, $Pe = 200$

FIG. 4.2. *Convergence behavior for the convection-diffusion problem*

grids, namely 200×200 and 600×600 mesh points, for two values of the Péclet number, $Pe = 20$ and $Pe = 200$. Convergence curves for the four methods are reported in Figure 4.2. Similarly to the previous example, one can observe that the adaptive method converges essentially in the same manner as the original MK(5,5,5) version. On the other hand, the convergence delay of the static version is much more marked, leading to non-convergence in some cases. The MK(2,2,2) version fails to converge in 100 iterations. The amount of work as measured by number of direct solves on the coarsest grid, number of iterations, and error at convergence are reported in Table 4.2.

Again, as in the previous example, the adaptive method is faster than the regular MK(5,5,5) version by about 10%, with the value of $c_m = 10$.

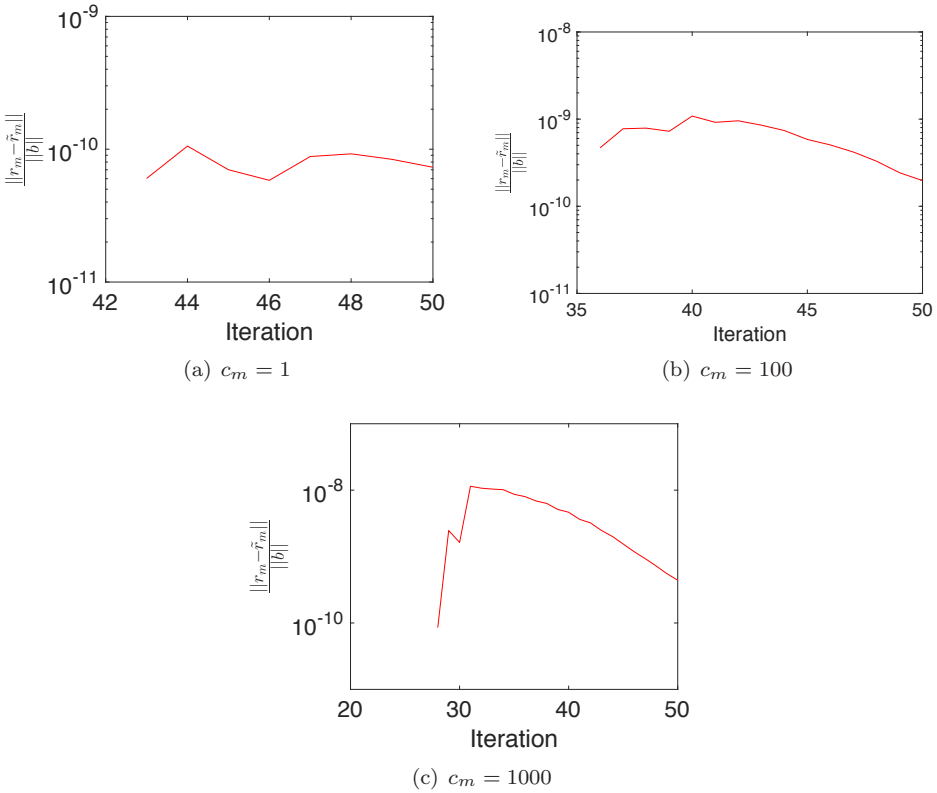
It is a natural question to ask, how would the performance be influenced by the choice of the constant c_m. As we observed earlier, our Theorem 3.3 guarantees that with the appropriate value, our adaptive method converges, and furthermore, its computational effort is expected to be lower than that of the regular MK method. The problem is that in general, we do not have available the quantities defining c_m. As we showed, the value $c_m = 10$ has worked for the problems studied so far. As shown in [21], this value may not always work, and if one does not have information to estimate its value, some experimentation may be needed. We run our examples with larger values of c_m and found that for moderate values, the relative residual is maintained around 10^{-10} as the theory predicts. On the other hand, for very large

Pe = 20		MK(5,5,5)	MKRadap	MKRstat	MK(2,2,2)
200^2	# direct solves (# It.)	3250(26)	2825(26)[21]	1701(81)	−(100)
	error	$5 \cdot 10^{-9}$	$1.2 \cdot 10^{-8}$	$3.2 \cdot 10^{-7}$	$7 \cdot 10^{-6}$
600^2	# direct solves (# It.)	3125(25)	2800(26)[20]	−(100)	- (100)
	error	$4.6 \cdot 10^{-8}$	$9.1 \cdot 10^{-8}$	$1.3 \cdot 10^{-4}$	0.4
Pe = 200					
200^2	# direct solves (# It.)	6250(50)	5725(50)[43]	1773(90)	- (100)
	error	$1.2 \cdot 10^{-8}$	$2 \cdot 10^{-8}$	$2.8 \cdot 10^{-8}$	$2.4 \cdot 10^{-8}$
600^2	# direct solves (# It.)	8175(65)	7750(66)[59]	−(100)	- (100)
	error	$8.4 \cdot 10^{-8}$	$1.9 \cdot 10^{-7}$	$8.3 \cdot 10^{-4}$	$1.1 \cdot 10^{-2}$

TABLE 4.2

Number of calculated coarsest grid solutions (number of iterations to converge) [iteration number where switching criterion is satisfied], and error $\|x - x_k\|$ of the different MK methods for the convection-diffusion problem.

values, the relative residual can grow up to two orders of magnitud larger. This is illustrated in Figure 4.3.



(a) $c_m = 1$ (b) $c_m = 100$

(c) $c_m = 1000$

FIG. 4.3. *Relative residual difference for the convection-diffusion equation with $Pe = 200$ on a 200^2-point grid*

4.3. Further experiments. In order to explore further the possible dependency on the choice of the value of c_m, we ran several experiments with the *dorr* matrices

from the matlab gallery package. These matrices are diagonal dominant, tridiagonal and ill-conditioned. We choose the default value $\alpha = 0.05$, which leads to a condition number of $\approx 3.6 \cdot 10^9$. We ran the adaptive version of the MK method. The experiments show that the convergence behavior does not change when c_m varies within an interval. More precisely, there are intervals such that the convergence behavior of the method does not change if c_m varies within those intervals. On the other hand, from one interval to the other, the behavior does change. Within those intervals, the larger the value of c_m, the faster the convergence, as shown in Table 4.3. It should be noted though, that even though there is a variation of the execution times, they are all within a few percentage points of each other, and they are all about 20% faster than the regular MK method.

c_m	regular MK	1	2.05	2.1	3.15	3.2	3.84	3.845
time	1.217	1.009	0.991	0.995	0.981	0.990	0.983	1.032

TABLE 4.3

CPU time (in seconds) for the solution of a linear system of order $n = 4 \cdot 10^4$ using the adaptive method for different values of c_m.

5. Conclusions. We developed a theory of adaptive multilevel Krylov (AMK) methods. At each level, the number of iterations of the next (coarser) level is determined by the residual at the previous step; using the criterion (3.11). With this strategy, the theory guarantees convergence of the AMK method to tolerance. The constant c_m in (3.11) depends on quantities which are not always available. If one finds that given a value of c_m, the method fails to convergence, the value of c_m should be reduced. We show experimentally that if for a certain value of c_m the method converges, the execution times are not much affeected by increases in this constant.

REFERENCES

[1] A. Bouras and V. Frayssé. Inexact matrix-vector products in Krylov methods for Krylov methods: A relaxation strategy. *SIAM J. Matrix Anal. Appl.*, 26:660–678, 2015.

[2] A. Bouras, V. Frayssé, and L. Giraud. A relaxation strategy for inner-outer linear solvers in domain decomposition methods. Technical Report TR/PA/00/17, CERFACS, Toulouse, France, 2000.

[3] A. N. Brooks and T. J. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Eng.*, 32:199–259, 1982.

[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.*, 22:1291–1315, 2001.

[5] H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite Elements and Fast Iterative Solvers*. Oxford University Press, Oxford, UK, second edition, 2014.

[6] Y. A. Erlangga and R. Nabben. Deflation and balancing preconditioners for Krylov subspace methods applied to nonsymmetric matrices. *SIAM J. Matrix Anal.*, 30:684–699, 2008.

[7] Y. A. Erlangga and R. Nabben. Multilevel projection-based nested Krylov iteration for boundary value problems. *SIAM J. Sci. Comput.*, 30:1572–1595, 2008.

[8] Y. A. Erlangga and R. Nabben. On a multilevel Krylov method for the Helmholtz equation preconditioned by shifted Laplacian. *Elec. Trans. Numer. Anal.*, 31:403–424, 2008.

[9] Y. A. Erlangga and R. Nabben. Algebraic multilevel Krylov methods. *SIAM J. Sci. Comp.*, 31:3417–3437, 2009.

[10] Y. A. Erlangga, C. W. Oosterlee, and C. Vuik. A novel multigrid-based preconditioner for the heterogeneous Helmholtz equation. *SIAM J. Sci. Comput.*, 27:1471–1492, 2006.

[11] 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.

[12] 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.

[13] J. Frank and C. Vuik. On the construction of deflation-based preconditioners. *SIAM J. Sci. Comput.*, 23:442–462, 2001.

[14] J. Liesen and Z. Strakoš. *Krylov subspace methods. Principles and analysis*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013.

[15] R. Nabben and C. Vuik. A comparison of deflation and coarse grid correction applied to porous media flow. *SIAM J. Numer. Anal.*, 42:1631–1647, 2004.

[16] R. A. Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM J. Numer. Anal.*, 24:355–365, 1987.

[17] Y. Notay and P. S. Vassilevski. Recursive Krylov-based multigid cycles. *Numer. Linear Alg. Appl.*, 12:419–451, 2006.

[18] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Sci. Comput.*, 14:461–469, 1993.

[19] A. Sheikh, D. Lahaye, L. G. Ramos, R. Nabben, and C. Vuik. Accelerating the shifted Laplace preconditioner for the Helmholtz equation by multilevel deflation. *J. Comput. Physics*, 322:473-490, 2016.

[20] A. Sheikh, D. Lahaye, and C. Vuik. On the convergence of shifted Laplace preconditioner combined with multilevel deflation. *Numer. Linear Alg. Appl.*, 20:645–662, 2013.

[21] V. Simoncini and D. B. Szyld. Flexible inner-outer Krylov subspace methods. *SIAM J. Nume. Anal.*, 40(6):2219–2239, 2002.

[22] V. Simoncini and D. B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM J. Sci. Comp.*, 25:454–477, 2003.

[23] V. Simoncini and D. B. Szyld. On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods. *SIAM Rev.*, 47:247–272, 2005.

[24] A. Tosseli and O. B. Widlund. *Domain decomposition methods - Algorithms and Theory*, volume 34 of *Series in Computational Mathematics*. Springer, Berlin, Heidelberg, New York, 2005.