

Some Experiences with Multilevel Krylov Methods

Y.A. Erlangga

Abstract This note discusses convergence behaviors of multilevel Krylov methods for some simple problems, mainly focusing on the possible choice of transfer operators. This study is part of the search for an optimal multilevel Krylov method.

1 Introduction

In [3], Erlangga and Nabben proposed a multilevel method for solving the linear system

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}, \quad (1)$$

which mimics a multigrid process, but, instead of a smoother, with a Krylov method used at each level. Since a (optimal) Krylov method reduces, not smoothes, errors in some norm, the underlying concept of this Krylov-based multilevel method is different from that of multigrid. Algorithm 4 shows the main body of the two-level version of a multilevel Krylov (MK) method, based on (flexible) GMRES [7].

Notice that a multigrid component is in play: the course-grid solve, which involves a coarse-grid matrix A_c . In multigrid, this solve is associated with the reduction of slow varying components of errors. The fine and coarse subspace are connected by two transfer operators I_H^h and I_h^H , which prolongate and restrict, respectively, some quantities in the iteration process. In multigrid, these quantities are related to errors made by an approximate solution, obtained after a few smoothing steps or coarse-grid solves. In MK, this quantity is associated with vectors, which build the approximation subspace.

Y.A. Erlangga (✉)
Mathematics Department, Alfaisal University, At-Takhassusi Road (South), P.O. Box 10927,
Riyadh, 11533, Saudi Arabia
e-mail: yerlangga@alfaisal.edu; yogiae@gmail.com

Algorithm 4 Two-level Krylov

Set an initial guess of solution x_0 ;
 Compute the residual $r_0 = b - Ax_0$; Set $v_1 = r_0/\|r_0\|$;
 For $j = 1, \dots$, until convergence
 $p = Av_j$;
 $q = I_h^H(p - \lambda_n v_j)$;
 Solve for s : $A_c s = q$;
 $z_j = v_j - I_h^H s$;
 $w = Az_j$;
 Gram-Schmidt orthogonalization;
 End
 Extract the approximate solution from $Z = [z_1 \dots z_j]$.

From an abstract point of view, if the second-level problem is solved exactly, Algorithm 4 is the consequence of applying GMRES on the preconditioned linear system

$$AQ_N \tilde{x} = b, \quad x = Q_N \tilde{x}, \tag{2}$$

with $Q_N = I - I_H^h A_c^{-1} I_h^H A + \lambda_n I_H^h A_c^{-1} I_h^H$ and $\lambda_n = \max\{|\lambda_i|\}_{1 \leq i \leq n}$, with λ_i the eigenvalue of A . Q_N is called the shift operator, due to the following spectral property [3].

Theorem 1. *Let A be symmetric positive definite and $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$, $0 < \lambda_i \leq \lambda_j$, for $i < j$. If $I_H^h = [z_1 \dots z_m]$, with $Az_i = \lambda_i z_i$, $m < n$, $I_H^H = (I_H^h)^T$, and $A_c = (I_H^h)^T A I_H^h$ (Galerkin coarse grid), then $\sigma(AQ_N) = \{\lambda_{m+1}, \dots, \lambda_n\}$, with λ_n having multiplicity of at least $m + 1$.*

Thus, under the assumptions set in Theorem 1, Q_N somehow shifts m smallest eigenvalues of A to λ_n , without changing the rest, making the spectrum more clustered. Obviously, the system (2) is more favorable for a Krylov method than the original system (1).

The first two terms in Q_N form the (right) deflation operator [6], denoted by Q_D . The effect of Q_N and Q_D on A are spectrally equivalent in the following sense [3].

Theorem 2. *Let A be symmetric positive definite. Let $Q_D = I - I_H^h A_c^{-1} I_H^H A$.*

$$\sigma(AQ_D) := \{0, \mu_{m+1}, \dots, \mu_n\} \iff \{\mu_{m+1}, \dots, \mu_n, \lambda_n\} =: \sigma(AQ_N).$$

Furthermore, if $P_D = Q_D^T$ and $P_N = I - A I_H^h A_c^{-1} I_h^H + \lambda_n I_H^h A_c^{-1} I_h^H$, then $\sigma(P_D A) = \sigma(AQ_D)$ and $\sigma(P_N A) = \sigma(AQ_N)$, and the same equivalence holds.

Theorem 2 can be easily extended to a more general class of matrices A .

In this short note, we present observed convergence behaviors of this method, based on some relatively simple problems. We shall base the presentation on the two level Krylov method (Algorithm 4), which represents the best performance possibly attained in terms of numbers of iterations, for a given multilevel Krylov setup.

Throughout this note, $\sigma(\cdot)$, $\lambda(\cdot)$, $\kappa(\cdot)$, $\mathcal{R}(\cdot)$, and $\mathcal{N}(\cdot)$ denote respectively the spectrum, eigenvalue, condition number, range, and null space of the argument.

2 Spectral Properties and Observed Convergence

In this section, we shall consider a hypothetical problem: a diagonal matrix $A = \text{diag}(1, 2, \dots, n)$. In this case, $\sigma(A) = \{1, 2, \dots, n\}$, the eigenvectors $v_i = \alpha_i \mathbf{e}_i$, $\alpha_i \neq 0$, and for any $b = [b_i]$, the linear system (1) has the solution $x = [x_i]_{1 \leq i \leq n} = [b_i/a_{ii}]$.

2.1 Eigenvectors

Let $I_H^h = [\mathbf{e}_i]_{1 \leq i \leq m}$, $I_H^h = (I_H^h)^T$, and set $A_c = I_H^h A I_H^h$ (Galerkin-type coarse-grid matrix). In this case, according to Theorem 2, $\sigma(P_N A) = \{m + 1, \dots, n\}$, with $\lambda_n = n$ having a multiplicity $m + 1$. Furthermore, $\sigma(P_D A) = \{0, m + 1, \dots, n\}$. As $P_D A$ is symmetric positive semi-definite, $\kappa_{\text{eff}}(P_D A) = n/(m + 1) < n = \kappa(A)$.

Let $x_0 = 0$, and consider the left-preconditioned version of the two-level Krylov method: $P_N A x = P_N b$. The Krylov subspace after the k -th iteration is

$$\mathcal{K}(P_N A, r_0) = \text{span}\{r_0, P_N A r_0, \dots, (P_N A)^{k-1} r_0\}.$$

Straight-forward computations lead to the following results:

$$r_0 = \lambda_n D_m^{-1} b + D_{n-m}^0 b, \quad (P_N A)^i r_0 = \lambda_n^i D_m^{-1} b + D_{n-m}^{i-1} b, \quad i = 1, \dots, k - 1,$$

with $D_m^{-1} = \text{diag}(A_m^{-1} \ 0)$ and $D_{n-m} = \text{diag}(0 \ A_{n-m})$. Thus,

$$\begin{aligned} x_{k,N} &\in \text{span}\{\lambda_n D_m^{-1} b + D_{n-m}^0 b, \lambda_n D_m^{-1} b + D_{n-m} b, \dots, \lambda_n^{k-1} D_m^{-1} b + D_{n-m}^{k-1} b\} \\ &\subseteq \text{span}\{D_m^{-1} b, D_{n-m}^0 b, D_{n-m} b, \dots, D_{n-m}^{k-1} b\}. \end{aligned}$$

Consider the (deflated) CG iteration applied to $P_D A x = P_D b$. In this case, CG generates a sequence of approximate solutions, $\{\tilde{x}_{k,D}\}$, to this singular linear system such that

$$\tilde{x}_{k,D} \in \text{span}\{D_{n-m}^0 b, D_{n-m} b, \dots, D_{n-m}^{k-1} b\}.$$

The solution of $Ax = b$ is then constructed as follows: $x = (I - P_D^T)x + P_D^T x = I_H^h A_c^{-1} I_H^h b + P_D^T x$, implying a sequence $x_{k,D} = I_H^h A_c^{-1} I_H^h b + P_D^T \tilde{x}_{k,D}$, and

$$\begin{aligned} x_{k,D} &\in I_H^h A_c^{-1} I_H^h b + \text{span}\{P_D^T D_{n-m}^0 b, P_D^T D_{n-m} b, \dots, P_D^T D_{n-m}^{k-1} b\} \\ &= D_{n-m}^{-1} b + \text{span}\{D_{n-m}^0 b, D_{n-m} b, \dots, D_{n-m}^{k-1} b\}. \end{aligned}$$

Fig. 1 Convergence history of CG with deflation P_D (solid) and shift P_N (dotted)

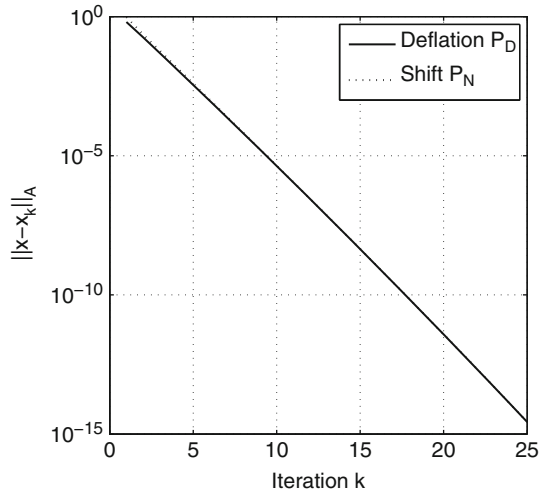
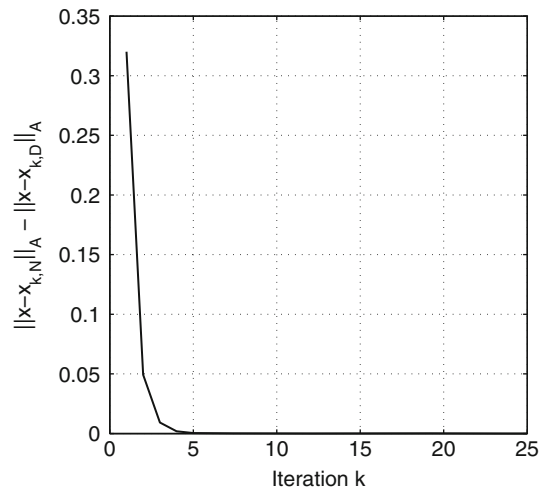


Fig. 2 Difference in the A-norm: $\|x - x_{k,N}\|_A - \|x - x_{k,D}\|_A$



So, $x_{k,N}$ and $x_{k,D}$ are members of the same subspace. From [5], for every $x_k \in \text{span}\{D_{n-m}^{-1}b, D_{n-m}^0b, D_{n-m}b, \dots, D_{n-m}^{k-1}b\} \equiv \mathcal{L}$,

$$\|x - x_{k,D}\|_A^2 \leq \|x - x_k\|_A^2.$$

Since $x_{k,N} \in \mathcal{L}$,

$$\|x - x_{k,D}\|_A^2 \leq \|x - x_{k,N}\|_A^2.$$

This result suggests the superiority of deflation technique to the shift operator, shown in Figs. 1 and 2 for the hypothetical problem.

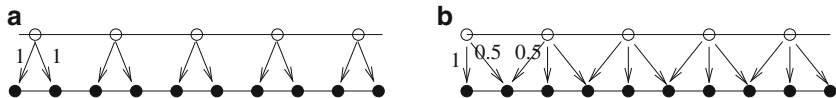


Fig. 3 One dimensional (a) aggregation and (b) linear interpolation, with \circ and \bullet indicating the fine and course nodes, respectively

A similar result holds also for (right preconditioned) GMRES, reading

$$\|b - Ax_{k,D}\|_2 \leq \|b - Ax_{k,N}\|_2.$$

The proof is skipped, and will be presented in another paper.

2.2 Interpolation Matrices

Consider any sparse, full rank interpolation matrix $I_H^h \in \mathbb{R}^{n \times m_1}$ and let $\mathcal{I}_H^h \in \mathbb{R}^{n \times m_2}$ be another full rank matrix. The following theorem is proved in [5]:

Theorem 3. Let $A_{c,1} = (I_H^h)^T A I_H^h$ and $A_{c,2} = (\mathcal{I}_H^h)^T A \mathcal{I}_H^h$. Let $P_{D,1} = I - A I_H^h A_{c,1}^{-1} (I_H^h)^T$, and similarly for $P_{D,2}$. If $\mathcal{R}(I_H^h) \subseteq \mathcal{R}(\mathcal{I}_H^h)$, then

1. $\lambda_n(P_{D,1}A) \geq \lambda_n(P_{D,2}A)$, and
2. $\lambda_{m_1+1}(P_{D,1}A) \leq \lambda_{m_2+1}(P_{D,2}A)$.

The classical examples of interpolation matrices are those associated with aggregation and linear interpolation in multigrid. They are illustrated for 1D in Fig. 3. Let I_H^h and \mathcal{I}_H^h be matrices, associated with the “standard” aggregation and augmented aggregation, respectively. For the latter, we augment I_H^h by a column vector $(1 \ 0 \ \dots \ 0 \ 1)^T$. So, $m_2 = m_1 + 1$, I_H^h and \mathcal{I}_H^h are full rank, and $I_H^h \subseteq \mathcal{I}_H^h$. Thus, Theorem 3 holds. But, $P_N A$ (or $A Q_N$) is no longer symmetric, even if A is a diagonal matrix; CG certainly breaks down in this case, and GMRES has to be employed. Convergence of two-level Krylov is shown in Fig. 4. The figure suggests the faster convergence of the method with augmented aggregation. For deflation, this performance is predicted by Theorem 3 and the well-known convergence bound of CG (due to $\kappa_{\text{eff}}(P_{D,\mathcal{I}}A) \leq \kappa_{\text{eff}}(P_{D,I}A)$). For P_N , the GMRES convergence bound of [7] is, however, not useful for extracting detailed information about the behavior of the method (see the residuals at the initial stage of iterations). However, better clustering affects the overall convergence.

In Fig. 4, we show also the convergence based on the linear interpolation. The associated interpolation matrix (denoted by \mathcal{I}_H^h) is set such that it is of the same rank as the aggregation matrix (denoted by I_H^h). In this case, however, the inclusion condition of Theorem 3 does not hold. Use of linear interpolation clearly leads to a better convergence. This behavior is unfortunately not generally the rule, as we shall see in some examples in the subsequent sections.

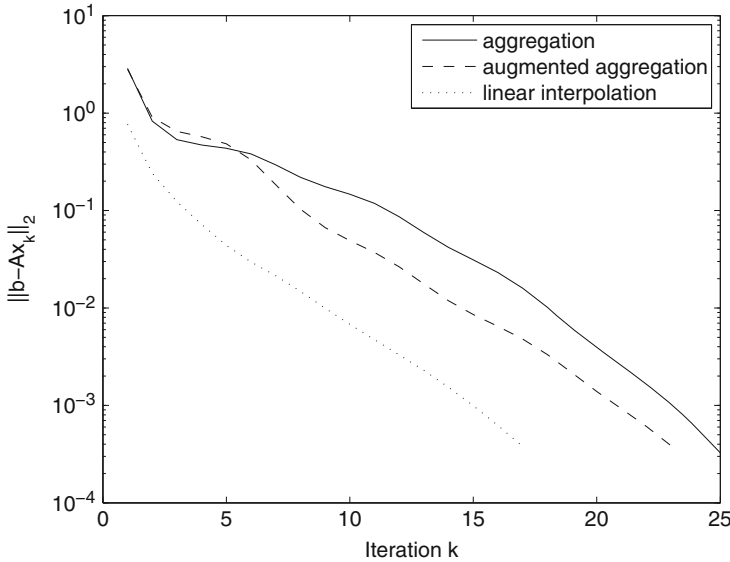


Fig. 4 GMRES convergence history, based on “standard” aggregation, augmented aggregation, and linear interpolation

3 Does Eigenvector-Based Transfer Operator Lead to a Better Convergence?

An almost general wisdom in deflation is to use eigenvectors or some (accurate) approximations to accelerate the convergence. The only reason one in practice avoids using them is because of the computational cost of computing even some of them. We shall address this issue in this section.

We consider a 1D Poisson problem in $[0, 1]$:

$$-u''(x) = f(x), \quad u(0) = u(1) = 0, \tag{3}$$

discretized by the second-order central difference. The eigenvalues and eigenvectors of the associated finite difference matrix A are

$$\lambda_k = 4 \sin^2 \left(\frac{\theta_k}{2} \right), \quad v_k = (\sin \theta_k, \sin(2\theta_k), \dots, \sin(n\theta_k))^T, \quad k = 1, \dots, n,$$

with $\theta_k = k\pi/(n + 1)$, with n the number of interior grid points (and hence, the size of A). In Fig. 5, we compare the performance of two-level MK method based on eigenvectors, (augmented) aggregation, and linear interpolation for $n = 200$ and $m = n/2 = 100$. Interestingly, methods based on aggregation perform the best; they converge to the machine accuracy in two iterations. The spectrum of AQ_N in this case consists of just two eigenvalues, i.e., $\sigma(AQ_N) = \{2, 3.999\}$, while with eigenvectors, $\sigma(AQ_N) = \{100, 101, \dots, 200\}$.

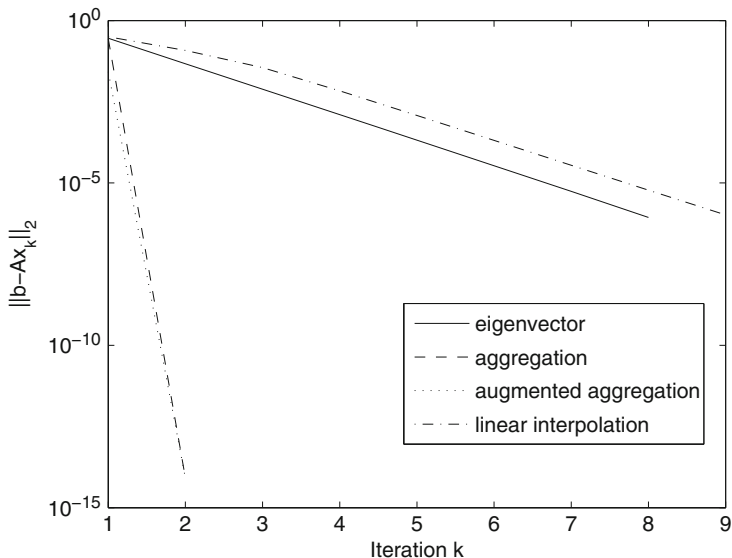


Fig. 5 GMRES convergence history for a 1D Poisson problem, based on eigenvectors, aggregation, and linear interpolation

Figure 6 shows a similar comparison, based on the 2D version of (3), discretized on a uniform finite difference mesh. In this case, eigenvector-based two-level method converges faster than the other scenarios, but aggregation-based techniques remain promising, and the (bi)linear interpolation performs the worst. This kind of performance is not one that we typically expect from multigrid (bilinear interpolation works well, and aggregation does not converge).

4 Singular but Consistent Systems

The last example is based on the 2D diffusion equation with Neumann conditions at the boundaries, set such that the resultant linear system is consistent. The matrix A is now singular, and is of rank $n - 1$. If A_c is nonsingular, it can be shown that $\mathcal{N}(P_N A) = \mathcal{N}(A Q_N) = \mathcal{N}((P_N A)^T)$ [2]. In this case, according to Theorem 2.4 of [1], GMRES is guaranteed to converge to the least-squares solution of (2). An invertible coarse-grid matrix A_c can be obtained by modifying a_{nn} such that the sum of the last row is nonzero.

Figure 7 shows the convergence history for different choices of l_H^h , for constant density. In this case, the (augmented) aggregation strategy outperforms both eigenvector- and bilinear interpolation-based approach. Convergence of the problem with one bubble is shown in Fig. 8. Use of eigenvectors leads to the fastest convergence, even though it is not as fast as the convergence seen from the previous examples.

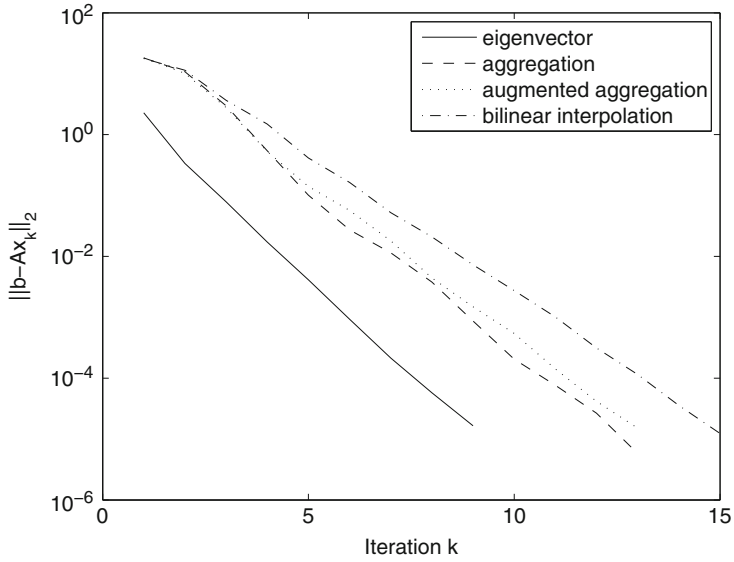


Fig. 6 GMRES convergence history for a 2D Poisson problem, based on eigenvectors, aggregation, and bilinear interpolation

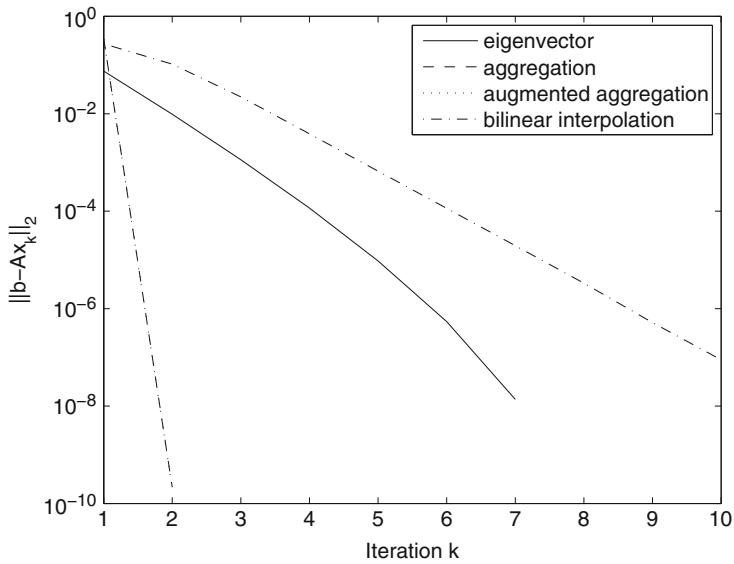


Fig. 7 GMRES convergence history for a 2D diffusion problem with uniform density

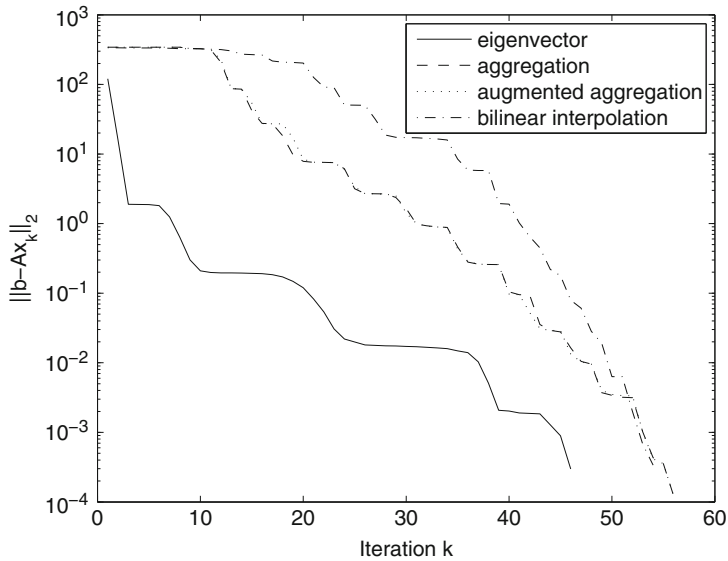


Fig. 8 GMRES convergence history for a 2D diffusion problem with one bubble and density ratio 10^3

5 Final Remarks

The potential of multilevel Krylov methods has been demonstrated in a number of papers; see also, e.g., [4]. This short note sheds some additional glimpse about this potential. Some examples suggest that eigenvector-based approach does not necessarily turn out to be the best way of setting up a multilevel Krylov method. Therefore, one can also argue that any approximation to eigenvectors may also lead to a non-optimal method. On the other hand, some examples show that it may be possible to get an extremely fast converging method using a rather simple transfer operator. What is crucial in this respect seems to lie in the choice of the transfer operator I_H^h . Unfortunately, at this point, no guideline is available in this direction.

References

1. Peter N Brown and Homer F Walker. GMRES on (nearly) singular systems. *SIAM J. Matrix Anal. Appl.*, 18:37–51, 1997.
2. Yogi A Erlangga. Multilevel krylov for singular systems. *manuscript*.
3. Yogi A Erlangga and Reinhard Nabben. Multilevel projection-based nested Krylov iteration for boundary value problems. *SIAM J. Sci. Comput.*, 30:1572–1595, 2008.
4. Yogi A Erlangga and Reinhard Nabben. Algebraic multilevel Krylov methods. *SIAM J. Sci. Comput.*, 31:3417–3437, 2009.

5. 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.
6. R A Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM J. Numer. Anal.*, 24:355–365, 1987.
7. Y Saad and M H Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.