

Augmented GMRES-type methods

James Baglama¹ and Lothar Reichel^{2,*,†}

¹*Department of Mathematics, University of Rhode Island, Kingston, RI 02881, U.S.A.*

²*Department of Mathematical Sciences, Kent State University, Kent, OH 44242, U.S.A.*

SUMMARY

GMRES is a popular iterative method for the solution of large linear systems of equations with a square non-symmetric matrix. The method generates a Krylov subspace in which an approximate solution is determined. We present modifications of the GMRES and the closely related RRGMR methods that allow augmentation of the Krylov subspaces generated by these methods by a user-supplied subspace. We choose this subspace to enable the representation of certain known non-smooth features of the desired solution, such as jumps, or to make it possible to represent certain smooth functions, such as constants or linear functions. The latter choice of augmenting subspace appears to be new. Applications to the solution of both well-posed and ill-posed problems are presented. Copyright © 2006 John Wiley & Sons, Ltd.

Received 20 August 2005; Revised 20 December 2005; Accepted 17 January 2006

KEY WORDS: GMRES; RRGMR; augmented Krylov subspace; linear systems

1. INTRODUCTION

The present paper is concerned with the iterative solution of large linear systems of equations with a non-symmetric matrix

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

The matrix A is assumed to be non-singular unless explicitly stated otherwise. GMRES [1, 2] is one of the most popular iterative solution methods for such systems. Let the initial approximate solution be $\mathbf{x}_0 = \mathbf{0}$. Then the j th iterate, \mathbf{x}_j , determined by GMRES is characterized by

$$\|\mathbf{Ax}_j - \mathbf{b}'\| = \min_{\mathbf{x} \in \mathcal{K}_j(A, \mathbf{b}')} \|\mathbf{Ax} - \mathbf{b}'\|, \quad \mathbf{x}_j \in \mathcal{K}_j(A, \mathbf{b}') \quad (2)$$

*Correspondence to: Lothar Reichel, Department of Mathematical Sciences, Kent State University, Kent, OH 44242, U.S.A.

†E-mail: reichel@math.kent.edu, URL: <http://www.math.kent.edu/~reichel>

Contract/grant sponsor: URI Research Proposal Development Grant Program

Contract/grant sponsor: NSF Grant; contract/grant numbers: DMS-0311786, DMS-0107858

where

$$\mathcal{K}_j(A, \mathbf{b}') = \text{span}\{\mathbf{b}', A\mathbf{b}', \dots, A^{j-1}\mathbf{b}'\} \quad (3)$$

is a Krylov subspace. Here and throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm. We assume that j is sufficiently small so that $\dim \mathcal{K}_j(A, \mathbf{b}') = j$.

Let \mathbf{x}' denote the solution of (1), let $d_1, d_2 \in \mathbb{R}$, $d_1 \neq 0$, $\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, and introduce the vectors

$$\mathbf{x}'' = d_1\mathbf{x}' + d_2\mathbf{w}, \quad \mathbf{b}'' = d_1\mathbf{b}' + d_2A\mathbf{w} \quad (4)$$

Then

$$A\mathbf{x}'' = \mathbf{b}'' \quad (5)$$

The vectors \mathbf{x}'' and \mathbf{b}'' are related to \mathbf{x}' and \mathbf{b}' , respectively, by simple linear transformations; however, the application of j steps of GMRES with initial approximate solution $\mathbf{x}_0 = \mathbf{0}$ to the two systems (1) and (5) can give approximate solutions of significantly different quality. Transformations of the form (4) may arise when changing units in a mathematical model. For instance, if \mathbf{x}' and \mathbf{x}'' represent temperatures in degrees Fahrenheit and Kelvin, respectively, then $d_1 = 5/9$, $d_2 = 255.37$, and $\mathbf{w} = [1, 1, \dots, 1]^T$ in (4).

The lack of invariance of GMRES under the simple transformations (4) can be addressed in several ways. For instance, we may carry out an initial transformation of the form (4) of the variables with the constants d_1 and d_2 chosen to yield high accuracy of the approximate solutions determined by GMRES or we may augment the Krylov subspaces (3) generated by GMRES by $\text{span}\{\mathbf{w}\}$. This paper pursues the latter approach and describes how to augment the Krylov subspaces determined by GMRES by an arbitrary linear space of low dimension. We remark that differently from discussions on augmented iterative methods available in the literature, we do not seek to augment by a space close to an invariant subspace of A .

The range restricted GMRES (RRGMRES) method differs from GMRES only in that the minimization problem (2) is replaced by

$$\|A\mathbf{x}_j - \mathbf{b}'\| = \min_{\mathbf{x} \in \mathcal{K}_j(A, A\mathbf{b}')} \|A\mathbf{x} - \mathbf{b}'\|, \quad \mathbf{x}_j \in \mathcal{K}_j(A, A\mathbf{b}') \quad (6)$$

where $\mathcal{K}_j(A, A\mathbf{b}') = A\mathcal{K}_j(A, \mathbf{b}')$. RRGMRRES was introduced in [3] and the method was shown in [4] to often give higher accuracy than GMRES when applied to the solution of linear discrete ill-posed problems with a contaminated right-hand side; see also [5] for a recent discussion of the method.

Section 2 describes implementations of the augmented GMRES and RRGMRRES methods based on the Arnoldi process and modified Gram–Schmidt orthogonalization, and discusses a few properties of these methods. Section 3 is concerned with the application of augmented GMRES and RRGMRRES to the solution of large linear discrete ill-posed problems with a right-hand side that is contaminated by an error. Computed examples are presented in Section 4, and concluding comments can be found in Section 5.

We remark that Chapman and Saad [6] and Morgan [7, 8] proposed augmentation of Krylov subspaces generated by restarted GMRES by spaces spanned by certain eigenvectors or Ritz vectors. Morgan [9] recently presented an elegant implementation of the latter approach. Convergence properties of Krylov subspace methods augmented by spaces close to invariant subspaces are discussed by Saad [10], who also considers block methods. This paper proposes augmentation

by low-dimensional subspaces, determined by vectors that represent smooth functions, such as constants, linear functions or quadratics, or by vectors that are able to represent known non-smooth features of the solution, such as jumps. We would like the augmenting subspace to contain a rough approximation of the desired solution, such that the difference between this approximation and the desired solution can be approximated well by an element of the Krylov subspace generated.

An augmented CGLS method is described by Calvetti *et al.* [11], who show that augmentation of the solution subspaces generated by standard CGLS by the linear space spanned by the right-hand side can give significantly improved accuracy in the computed approximate solution of linear discrete ill-posed problems with a right-hand side that is contaminated by an error. However, augmentation of the solution subspaces determined by RRGMRES by this subspace is not meaningful, since the method so obtained is GMRES; instead of augmenting RRGMRES by the linear space spanned by the right-hand side one should apply GMRES. We found that GMRES and RRGMRES can yield higher accuracy when the solution subspaces generated by these methods are augmented by spaces that allow the representation of simple functions, such as constants and linear functions. Augmentation of CGLS by such spaces is also likely to be beneficial, but this will not be pursued in the present paper. Yet another approach to augmentation methods, based on decomposing the solution, is presented in [12].

2. AUGMENTED GMRES AND RRGMRES

This section describes implementations of the augmented GMRES and RRGMRES methods used in the computed examples in Section 4. The augmented methods are obtained by modifying the standard implementations of GMRES and RRGMRES based on the Arnoldi process. Let the columns of the matrix $W \in \mathbb{R}^{n \times p}$ form a basis of the space \mathcal{W} . The j th iterate, \mathbf{x}_j , determined by GMRES augmented by \mathcal{W} with initial approximate solution $\mathbf{x}_0 = \mathbf{0}$ satisfies

$$\|\mathbf{A}\mathbf{x}_j - \mathbf{b}'\| = \min_{\mathbf{x} \in \mathcal{K}_j(\mathbf{A}, \mathbf{b}') \cup \mathcal{W}} \|\mathbf{A}\mathbf{x} - \mathbf{b}'\|, \quad \mathbf{x}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{b}') \cup \mathcal{W} \quad (7)$$

and, analogously, the j th iterate determined by augmented RRGMRES satisfies

$$\|\mathbf{A}\mathbf{x}_j - \mathbf{b}'\| = \min_{\mathbf{x} \in \mathcal{K}_j(\mathbf{A}, \mathbf{A}\mathbf{b}') \cup \mathcal{W}} \|\mathbf{A}\mathbf{x} - \mathbf{b}'\|, \quad \mathbf{x}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{A}\mathbf{b}') \cup \mathcal{W} \quad (8)$$

Augmentation by \mathcal{W} is carried out by first computing the QR-factorization

$$\mathbf{A}\mathbf{W} = \mathbf{V}_p \mathbf{R} \quad (9)$$

where $\mathbf{V}_p \in \mathbb{R}^{n \times p}$ has orthonormal columns and $\mathbf{R} \in \mathbb{R}^{p \times p}$ is upper triangular. We then append columns to \mathbf{V}_p , which are determined by a modified Arnoldi process. This process yields orthonormal vectors that also are orthogonal to the columns of \mathbf{V}_p . The initial vector for the modified Arnoldi process is chosen to be $(\mathbf{I} - \mathbf{V}_p \mathbf{V}_p^T) \mathbf{b}'$ for augmented GMRES and $(\mathbf{I} - \mathbf{V}_p \mathbf{V}_p^T) \mathbf{A}\mathbf{b}'$ for augmented RRGMRES. The generated vectors are appended to the matrix \mathbf{V}_p as they become available. After j steps of this modified Arnoldi process, we obtain the modified Arnoldi decomposition

$$\mathbf{A}[\mathbf{W} \mathbf{V}_{p+1:p+j}] = \mathbf{V}_{p+j+1} \mathbf{H} \quad (10)$$

where $V_{p+j+1} = [V_p \ V_{p+1:p+j+1}] \in \mathbb{R}^{n \times (p+j+1)}$ has orthonormal columns, and the first column of the trailing $n \times (j+1)$ submatrix $V_{p+1:p+j+1}$ of V_{p+j+1} is $(I - V_p V_p^T) \mathbf{b}' / \|(I - V_p V_p^T) \mathbf{b}'\|$ for augmented GMRES and $(I - V_p V_p^T) A \mathbf{b}' / \|(I - V_p V_p^T) A \mathbf{b}'\|$ for augmented RRGMRES. The remaining columns of $V_{p+1:p+j+1}$ are determined by the j steps of the modified Arnoldi process. The leading principal $p \times p$ submatrix of the upper Hessenberg matrix $H \in \mathbb{R}^{(p+j+1) \times (p+j)}$ is the upper triangular matrix R in the QR-factorization (9). The entries of the trailing j columns of H are determined by the modified Arnoldi process; see Algorithm 2.1 below for details.

Substituting (10) into (7) and (8) shows that the iterate \mathbf{x}_j determined by augmented GMRES or RRGMRES can be expressed as

$$\mathbf{x}_j = [W V_{p+1:p+j}] \mathbf{y}_j \quad (11)$$

where \mathbf{y}_j solves the least-squares problem

$$\min_{\mathbf{y} \in \mathbb{R}^{p+j}} \|V_{p+1:p+j}^T \mathbf{b}' - H \mathbf{y}\| \quad (12)$$

We remark that Saad [10] describes a related but different augmented Arnoldi process. In the following algorithm the entries of H are denoted by h_{jk} . Entries of H that are not explicitly defined are assumed to be zero.

Algorithm 2.1 (Augmented GMRES and RRGMRES)

Input: $A \in \mathbb{R}^{n \times n}$, $\mathbf{b}' \in \mathbb{R}^n$, $W \in \mathbb{R}^{n \times p}$, j ;
Output: Computed approximate solution \mathbf{x}_j ;

1. $\mathbf{v}_{p+1} := \mathbf{b}'$;
2. if RRGMRES then $\mathbf{v}_{p+1} := A \mathbf{b}'$; endif;
3. if $p > 0$ then
4. Compute QR-factorization $AW = V_p H$, where $V_p := [\mathbf{v}_1, \dots, \mathbf{v}_p]$;
5. for $i = 1, 2$ do $\mathbf{v}_{p+1} := \mathbf{v}_{p+1} - V_p (V_p^T \mathbf{v}_{p+1})$; endfor;
6. endif;
7. $\mathbf{v}_{p+1} := \frac{\mathbf{v}_{p+1}}{\|\mathbf{v}_{p+1}\|}$ and $V_{p+1} := [V_p \ \mathbf{v}_{p+1}]$;
8. for $k = p+1, \dots, p+j$ do
9. $\mathbf{v}_{k+1} := A \mathbf{v}_k$;
10. for $i = 1, \dots, k$ do
11. $h_{ik} := \mathbf{v}_i^T \mathbf{v}_{k+1}$; $\mathbf{v}_{k+1} := \mathbf{v}_{k+1} - h_{ik} \mathbf{v}_i$;
12. endfor;
13. for $i = 1, \dots, k$ do (reorthogonalization)
14. $\alpha := \mathbf{v}_i^T \mathbf{v}_{k+1}$; $\mathbf{v}_{k+1} := \mathbf{v}_{k+1} - \alpha \mathbf{v}_i$; $h_{ik} := h_{ik} + \alpha$;
15. endfor;
16. $h_{k+1,k} := \|\mathbf{v}_{k+1}\|$;
17. $\mathbf{v}_{k+1} := \frac{\mathbf{v}_{k+1}}{h_{k+1,k}}$ and $V_{k+1} := [V_k \ \mathbf{v}_{k+1}]$;
18. endfor;
19. Compute the solution \mathbf{y}_j of the least-squares problem (12) and determine \mathbf{x}_j from (11);

The computation of \mathbf{x}_j by GMRES augmented by $\mathcal{W} = \text{range } W$ using Algorithm 2.1 requires the evaluation of $p+j$ matrix–vector products with A ; when RRGMRES is used instead of

GMRES one additional matrix–vector product with A has to be evaluated. Lines 5 and 13–15 of the algorithm implement reorthogonalization because this can reduce the number of iterations required to determine an approximate solution of desired accuracy when the matrix A is very ill-conditioned, which is the case for the linear systems of equations considered in Section 3.

Algorithm 2.1 is used for the numerical examples of Section 4; however, we remark that a black-box implementation of our method has to allow the k -loop (lines 8–18) to be exited as soon as the range of $[WV_{p+1:p+k}]$ contains a vector \mathbf{x} , such that the associated residual error $\|A\mathbf{x} - \mathbf{b}'\|$ is sufficiently small. The norm of this residual error can be evaluated inexpensively just as in standard GMRES and RRGMRRES. A black-box algorithm also has to be able to handle break-down, i.e. the situation when $\mathbf{v}_{p+1} = \mathbf{0}$ in line 7 or $\mathbf{v}_{k+1} = \mathbf{0}$ in line 17 of Algorithm 2.1; see Theorem 2.2 below for further details on break-down.

The input parameter j in Algorithm 2.1 generally is chosen fairly small, say $j \leq 30$, in order to limit the requirement of computer memory for the storage of the matrices W and V_{j+p+1} . If a larger value of j is required in order to determine an approximate solution of (1) of sufficient accuracy, the algorithm can be restarted instead of increasing the value of j , similarly to standard GMRES and RRGMRRES. We first show that augmented GMRES is invariant under transformation of the solution by W .

Theorem 2.1

Assume that $A \in \mathbb{R}^{n \times n}$ is non-singular, and let $W \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \in \mathbb{R}^p$. Let \mathbf{x}' solve (1). Then $\mathbf{x}'' = \mathbf{x}' + W\mathbf{y}$ solves (5) with $\mathbf{b}'' = \mathbf{b}' + AW\mathbf{y}$. Apply j steps of GMRES augmented by $\mathcal{W} = \text{range } W$ with initial approximate solution $\mathbf{x}_0 = \mathbf{0}$ to (1) and (5), and denote the computed iterates by \mathbf{x}'_j and \mathbf{x}''_j , respectively. Then

$$\mathbf{x}' - \mathbf{x}'_j = \mathbf{x}'' - \mathbf{x}''_j \tag{13}$$

Proof

The iterates can be expressed as $\mathbf{x}'_j = [WV_{p+1:p+j}]\mathbf{y}'_j$ and $\mathbf{x}''_j = [WV_{p+1:p+j}]\mathbf{y}''_j$, where \mathbf{y}'_j and \mathbf{y}''_j solve

$$\min_{\mathbf{y} \in \mathbb{R}^{p+j}} \|A[WV_{p+1:p+j}]\mathbf{y} - \mathbf{b}'\| \quad \text{and} \quad \min_{\mathbf{y} \in \mathbb{R}^{p+j}} \|A[WV_{p+1:p+j}]\mathbf{y} - \mathbf{b}''\|$$

respectively. Therefore

$$A\mathbf{x}'_j - \mathbf{b}' = A\mathbf{x}''_j - \mathbf{b}''$$

which shows (13). □

We remark that Theorem 2.1 remains valid if GMRES is replaced by RRGMRRES. This replacement changes the matrix $V_{p+1:p+j}$ in the above proof; see Algorithm 2.1 for how this matrix has to be modified.

Break-down of standard GMRES implies that the solution of the linear system of equations (1) lives in the Krylov subspace already generated. The following theorem shows the analogous result for augmented GMRES.

Theorem 2.2

Let $A \in \mathbb{R}^{n \times n}$ be non-singular and assume that Algorithm 2.1 for augmented GMRES breaks down, i.e. $\mathbf{v}_{k+1} = \mathbf{0}$ for some integer k with $p \leq k \leq p + j \leq n$. If the matrix $[WV_{p+1:k}] \in \mathbb{R}^{n \times k}$

determined by the algorithm is of full rank, then its range contains the solution of the linear system of equations (1).

Proof

Algorithm 2.1 can break down at lines 7 and 17. We consider the former break-down first. Then $k = p$, $\mathbf{v}_{p+1} = \mathbf{0}$, and the matrix $[WV_{p+1:k}]$ simplifies to W . The fact that both A and W are of full rank shows that the upper triangular matrix R in the QR-factorization (9) is non-singular. It follows from $\mathbf{b}' \in \text{range } V_p$ and the factorization (9) that $WR^{-1}V_p^T\mathbf{b}'$ solves (1).

We turn to the break-down at line 17. Then $\mathbf{v}_{k+1} = \mathbf{0}$ for some $k > p$, and the Arnoldi decomposition (10) can be expressed as

$$A[WV_{p+1:k}] = V_k H \tag{14}$$

where $H \in \mathbb{R}^{k \times k}$ is of upper Hessenberg form. The entries of H are the elements h_{ij} determined so far by the algorithm, except for $h_{k+1,k}$. By assumption the matrix on the left-hand side of (14) is of full rank and, therefore, so is H . The break-down signals that $\mathbf{b}' \in \text{range } V_k$ and we can use (14) to express the solution of (1) as $[WV_{p+1:k}]H^{-1}V_k^T\mathbf{b}'$. \square

The following theorem shows that just as for standard GMRES, break-down for standard RRGMRES is benign, i.e. at break-down the solution of (1) lives in the Krylov subspace already generated.

Theorem 2.3

Assume that $A \in \mathbb{R}^{n \times n}$ is non-singular. If Algorithm 2.1 breaks down for standard RRGMRES, then the Krylov subspace available at break-down contains the solution of the linear system of equations (1).

Proof

Since A is non-singular, the linear system of equations

$$A\mathbf{x} = A\mathbf{b}' \tag{15}$$

has the unique solution $\mathbf{x} = \mathbf{b}'$. We may assume that $\mathbf{b}' \neq \mathbf{0}$. Then, for standard RRGMRES, Algorithm 2.1 only can break down at line 17, in which case $\mathbf{v}_{k+1} = \mathbf{0}$ and

$$AV_k = V_k H \tag{16}$$

where $H \in \mathbb{R}^{k \times k}$ is of upper Hessenberg form and non-singular. Since $A\mathbf{b}' \in \text{range } V_k$, it follows from (16) that the solution of (15) can be expressed as $\mathbf{x} = V_k H^{-1}V_k^T A\mathbf{b}'$. This implies $\mathbf{b}' \in \text{range } V_k$ and it now follows from (16) that the solution of (1) can be expressed as $V_k H^{-1}V_k^T\mathbf{b}'$. \square

Break-down for augmented RRGMRES does not imply that the solution lives in the union of the Krylov subspace generated so far and \mathcal{W} . If $\mathbf{v}_{p+1} = \mathbf{0}$ at line 7 of Algorithm 2.1, then $A\mathbf{b}' \in \text{range } V_p$. Using (9), we obtain

$$\min_{\mathbf{y} \in \mathbb{R}^p} \|AW\mathbf{y} - \mathbf{b}'\| = \min_{\mathbf{y} \in \mathbb{R}^p} \|V_p R\mathbf{y} - \mathbf{b}'\| = \|(I - V_p V_p^T)\mathbf{b}'\| \tag{17}$$

where the minimum is achieved for $\tilde{\mathbf{y}} = R^{-1}V_p^T\mathbf{b}'$. The vector $\tilde{\mathbf{x}} = W\tilde{\mathbf{y}}$ is an acceptable approximate solution of (1) if the right-hand side of (17) is sufficiently small. If this is not the case, then we let

$\mathbf{v}_{p+1} := A\mathbf{v}_{p+1}$ and resume the computations at line 5 of Algorithm 2.1. If necessary, we repeat the multiplication of \mathbf{v}_{p+1} by A until break-down does not occur at line 7. Break-down at line 17 can be handled analogously.

The following corollary shows that if we choose the columns of W to be orthogonal to \mathbf{b}' , then break-down for augmented RRGMRRES is benign, i.e. at break-down the solution of (1) lives in the solution subspace already generated. The proof is analogous to that of Theorem 2.3.

Corollary 2.4

Let $A \in \mathbb{R}^{n \times n}$ be non-singular, let W be of full rank, and such that $W^T \mathbf{b}' = \mathbf{0}$. Assume that Algorithm 2.1 for augmented RRGMRRES breaks down. Then the solution of the linear system (1) is in the union of $\mathcal{W} = \text{range } W$ and the available Krylov subspace at break-down.

Proof

Assume first that Algorithm 2.1 breaks down at line 7. Then $A\mathbf{b}' \in \text{range } V_p$, and it follows from the non-singularity of A that $\mathbf{b}' \in \mathcal{W}$. This contradiction shows that break-down cannot occur at line 7. Now assume that Algorithm 2.1 breaks down at line 17. Then $\mathbf{v}_{k+1} = \mathbf{0}$ and we have decomposition (14) with $H \in \mathbb{R}^{k \times k}$ non-singular. Similarly as in the proof of Theorem 2.3, consider the linear system (15). Since $A\mathbf{b}' \in \text{range } V_k$, it follows that $\mathbf{x} = [WV_{p+1:k}]H^{-1}V_k^T A\mathbf{b}'$. Thus, \mathbf{x} , and therefore \mathbf{b}' , are in the range of $[WV_{p+1:k}]$. But the range of W is orthogonal to \mathbf{b}' , and therefore $\mathbf{b}' \in \text{range } V_{p+1:k} \subset \text{range } V_k$. It follows that the solution of (1) can be expressed as $[WV_{p+1:k}]H^{-1}V_k^T \mathbf{b}'$. \square

Break-down is a rare event. We therefore in the numerical experiments with augmented RRGMRRES reported in Section 4 do not require the range of W to be orthogonal to \mathbf{b}' .

3. LINEAR DISCRETE ILL-POSED PROBLEMS

A matrix $A \in \mathbb{R}^{n \times n}$ is said to be of ill-determined rank if it has many singular values of different orders of magnitude close to the origin. Such matrices are severely ill-conditioned and may be singular. They arise from the discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. Linear systems of equations (1) with a matrix of ill-determined rank are often referred to as linear discrete ill-posed problems. A nice introduction to numerical methods for the solution of such problems is provided by Hansen [13]. We will in this section assume that (1) is a linear discrete ill-posed problem.

The right-hand side \mathbf{b}' in linear discrete ill-posed problems that arise in applications typically is contaminated by an error $\mathbf{e} \in \mathbb{R}^n$, which may stem from measurement or discretization errors. Let $\hat{\mathbf{b}} \in \mathbb{R}^n$ denote the unknown error-free vector associated with \mathbf{b}' , i.e.

$$\mathbf{b}' = \hat{\mathbf{b}} + \mathbf{e} \tag{18}$$

and assume that the linear system of equations with the unknown error-free right-hand side

$$A\mathbf{x} = \hat{\mathbf{b}} \tag{19}$$

is consistent. The available linear system (1) is not required to be consistent.

We would like to determine a solution $\hat{\mathbf{x}}$ of (19). If A is singular, then we may be interested in computing the least-squares solution of minimal Euclidean norm. Since the right-hand side

$\hat{\mathbf{b}}$ is not available, we seek to determine an approximation of $\hat{\mathbf{x}}$ by computing an approximate solution of the available linear system of equations (1). A popular approach to determining an approximation of $\hat{\mathbf{x}}$ for large-scale linear discrete ill-posed problems is to apply a few, say j , steps of an iterative method, such as GMRES or RRGMR, to (1); see, e.g. [4, 14, 15] for discussions of these iterative methods, and [11, 13, 16, 17] for discussions of several other iterative methods applied to the solution of linear discrete ill-posed problems. As in Section 1, we denote the j th iterate determined by any one of these iterative methods by \mathbf{x}_j . Since $\hat{\mathbf{x}}$ is assumed not to satisfy (1), there typically is a fairly small non-negative integer $j_* \ll n$, such that

$$\|\mathbf{x}_{j_*} - \hat{\mathbf{x}}\| = \min_{j \geq 0} \|\mathbf{x}_j - \hat{\mathbf{x}}\| \tag{20}$$

see, e.g. the computed examples in Section 4 for illustrations. It is therefore important that the low-dimensional subspaces which contain iterates \mathbf{x}_j with j small allow representation of pertinent features of $\hat{\mathbf{x}}$. These features may be jumps, spikes, or just linear increase. This suggests that if low-dimensional Krylov subspaces determined by GMRES or RRGMR do not allow representation of desirable features, then these spaces should be augmented by subspaces that do. The following section illustrates that augmentation can enhance the quality of computed approximations of $\hat{\mathbf{x}}$.

4. NUMERICAL EXAMPLES

This section presents a few computed examples which illustrate that augmentation of Krylov subspaces may reduce the computational effort for the solution of well-conditioned problems. We also show examples in which augmented Krylov subspace iterative methods applied to the solution of linear discrete ill-posed problems with a contaminated right-hand side yield better approximations of the desired solution $\hat{\mathbf{x}}$ of the associated unavailable linear discrete ill-posed problem (19) with error-free right-hand side than standard Krylov subspace methods.

For large linear systems of equations, the dominating computational work when applying GMRES or RRGMR, with or without augmentation, is the evaluation of matrix–vector products with the matrix A . We therefore report the number of matrix–vector product evaluations as a measure of the computational effort required. All computations were carried out in Matlab with machine epsilon 2×10^{-16} . Throughout this section we let

$$W_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}, \quad W_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix}$$

and define the linear spaces

$$\mathcal{W}_k = \text{range } W_k, \quad 1 \leq k \leq 3 \tag{21}$$

We comment on the choice of these spaces in some of the example below. Here, we only note that augmentation by the space \mathcal{W}_1 is suitable for the temperature conversion example of Section 1.

This augmentation secures that the computed iterates, up to a scaling factor, are independent of the temperature units chosen; cf. Theorem 2.1.

In all computed examples, we use the initial approximate solution $\mathbf{x}_0 = \mathbf{0}$. When solving linear systems of equations (1) with an error-free right-hand side, it is desirable to choose the augmented subspace \mathcal{W} so that the component of the solution of (1) not in \mathcal{W} can be approximated well by an element in a low-dimensional Krylov subspace $\mathcal{K}_j(A, \mathbf{b}')$. When instead solving linear discrete ill-posed problems with a contaminated right-hand side, we would like the subspace \mathcal{W} be such that the component of the desired solution $\hat{\mathbf{x}}$ of the unavailable linear system (19) not in \mathcal{W} can be approximated well by an element in a low-dimensional Krylov subspace $\mathcal{K}_j(A, \mathbf{b}')$. The following examples illustrate the performance of augmented GMRES and RRGMRES for several choices of augmenting subspaces.

Example 4.1

Let $A \in \mathbb{R}^{500 \times 500}$ be a non-symmetric Toeplitz matrix with first column $[1, 1/2, 1/3, \dots, 1/500]^T$ and first row $[1, 1/4, 1/9, \dots, 1/500^2]$. This matrix is well-conditioned; it has condition number $\kappa(A) = 1.2 \times 10^1$, where $\kappa(A) = \|A\| \|A^{-1}\|$. Define the solution $\mathbf{x}' = [x'_1, x'_2, \dots, x'_{500}]^T$ with entries $x'_j = \exp(-(j - 1)/500)$, and let $\mathbf{b}' = A\mathbf{x}'$.

Figure 1 displays the Euclidean norm of the error in the computed approximate solutions of (1) obtained by standard and augmented GMRES *versus* the number of matrix–vector product evaluations with the matrix A . For instance, when augmenting by \mathcal{W}_2 , the computation of the approximate solution $\mathbf{x}_8 \in \mathcal{K}_8(A, \mathbf{b}') \cup \mathcal{W}_2$ requires the evaluation of 10 matrix–vector products with A , and illustrates that augmentation can reduce the error significantly for a fixed number

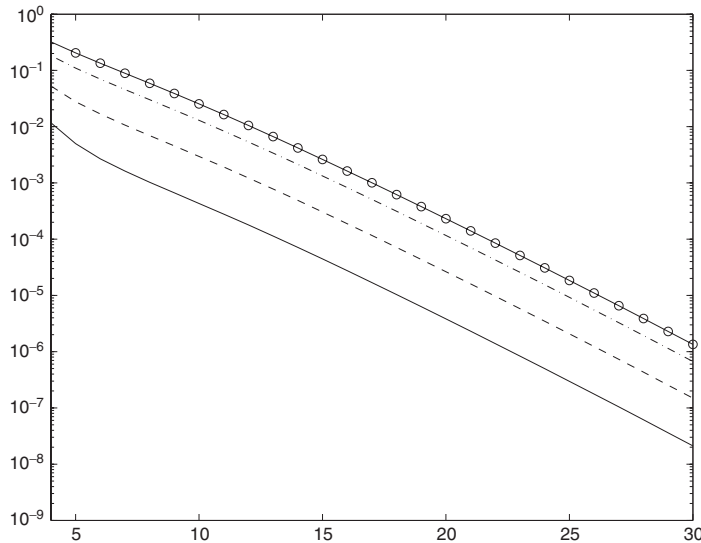


Figure 1. Example 4.1. Euclidean norm of the error in the computed iterates *versus* the number of matrix–vector product evaluations for standard GMRES (solid curve with circles), GMRES augmented by \mathcal{W}_1 (dash-dotted curve), GMRES augmented by \mathcal{W}_2 (dashed curve), and GMRES augmented by \mathcal{W}_3 (solid curve).

Table I. Example 4.2. Comparison of standard and augmented GMRES.

Augmentation	j	$\ \mathbf{x}' - \mathbf{x}_j\ $	Mat.-vec. prod.
—	10	5.0×10^{-2}	10
\mathcal{W}_Δ	9	4.2×10^{-2}	10
\mathcal{W}_3	7	9.3×10^{-2}	10
$\mathcal{W}_\Delta \cup \mathcal{W}_3$	6	8.5×10^{-4}	10

of matrix–vector product evaluations. Figure 1 indicates that the augmentation does not affect the asymptotic rate of convergence.

Example 4.2

This example illustrates that it can be expedient to augment GMRES by a space that allows the representation of a jump discontinuity of the solution. Let A be the same matrix as in Example 4.1 and let the entries of the solution $\mathbf{x}' = [x'_1, x'_2, \dots, x'_{500}]^T$ be given by

$$x'_k = \begin{cases} \exp(-(k - 1)/500), & 1 \leq k \leq 250 \\ \exp(-(k - 1)/500) + 1, & 251 \leq k \leq 500 \end{cases}$$

Thus, \mathbf{x}' represents the discretization of the sum of an exponential function and a step function. Define $\mathbf{b}' = A\mathbf{x}'$. We would like to solve the linear system of equations (1).

Let $\mathbf{w} = [w_1, w_2, \dots, w_{500}]^T$ be the discretization of a piece-wise constant function with a jump where the solution \mathbf{x}' is discontinuous, specifically we let

$$w_k = \begin{cases} 0, & 1 \leq k \leq 250 \\ 1, & 251 \leq k \leq 500 \end{cases} \tag{22}$$

and define $\mathcal{W}_\Delta = \text{range } \mathbf{w}$.

Table I summarizes the performance of standard and augmented GMRES, and compares the errors in computed approximate solutions that all require the evaluation of 10 matrix–vector products with A . The table illustrates that significant improvement in accuracy can be achieved if vectors in the augmented space can represent known non-smooth features of the solution.

Augmented GMRES requires the approximate location of the discontinuity to be known in order to deliver accurate approximate solutions. We are presently investigating the possibility of combining augmented GMRES with schemes that locate discontinuities. A scheme for locating discontinuities of small to medium-sized problems has been described by Hansen and Mosegaard [18].

The following examples are concerned with the approximate solution of linear discrete ill-posed problems of the form (1) with a right-hand side (18) that has been contaminated by an error \mathbf{e} of relative norm ε , i.e.

$$\frac{\|\mathbf{e}\|}{\|\hat{\mathbf{b}}\|} = \varepsilon \tag{23}$$

The entries of \mathbf{e} are normally distributed with zero mean.

Table II. Example 4.3. Comparison of standard and augmented RRGMRES applied to the solution of (24).

Augmentation	$\ \hat{\mathbf{x}} - \mathbf{x}_{j^*}\ $	j^*	Mat.-vec. prod.
—	6.3×10^{-1}	4	5
\mathcal{W}_1	5.0×10^{-2}	2	4

Example 4.3

Consider the Fredholm integral equation of the first kind,

$$\int_0^\pi \exp(s \cos(t))x(t) dt = b(s), \quad 0 \leq s \leq \frac{\pi}{2} \tag{24}$$

with

$$b(s) = 2 \frac{\sinh(s)}{s} + 100 \int_0^\pi \exp(s \cos(t)) dt$$

A related integral equation is discussed by Baart [19]. We modified the Matlab function `baart` from the program package Regularization Tools by Hansen [20], which discretizes the integral equation considered by Baart [19], to yield a discretization of (24). Specifically, we determined a non-symmetric numerically singular matrix $A \in \mathbb{R}^{1000 \times 1000}$ and a scaled discrete approximation $\hat{\mathbf{x}}$ of the solution

$$x(t) = 100 + \sin(t) \tag{25}$$

of (24). Let $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$ and define the right-hand side of (1) by (18) with $\varepsilon = 1 \times 10^{-5}$ in (23).

For this and the following examples the standard and augmented RRGMRES methods yield approximations of $\hat{\mathbf{x}}$ with a smaller error than the corresponding GMRES methods. We therefore only report errors for standard and augmented RRGMRES.

Since the solution (25) of (24) is a fairly small relative perturbation of a constant, it is natural to augment the Krylov subspaces generated by standard RRGMRES by the space \mathcal{W}_1 defined by (21). Table II displays the index j^* defined by (20) and the norm of the error in the computed approximation \mathbf{x}_{j^*} of $\hat{\mathbf{x}}$ for standard and augmented RRGMRES. The computed approximate solution \mathbf{x}_{j^*} determined by augmented RRGMRES lives in $\mathcal{K}_{j^*}(A, A\mathbf{b}) \cup \mathcal{W}_1$ and its computation requires the evaluation of $j^* + 2$ matrix–vector products with the matrix A . The vector $\hat{\mathbf{x}}$ and the approximate solutions \mathbf{x}_{j^*} computed by standard and augmented RRGMRES are shown in Figure 2. Table II displays the error in the most accurate computed iterates and shows augmentation by \mathcal{W}_1 to reduce this error by about a factor $\frac{1}{13}$.

In the above example, as well as in the examples below, the index j^* is assumed to be known, and we display the smallest errors achieved by standard and augmented RRGMRES. For many linear discrete ill-posed problems that arise in applications, j^* has to be estimated. Several methods are available for this purpose, see, e.g. [13, 15, 16], and can be used in conjunction with the iterative methods of the present paper.

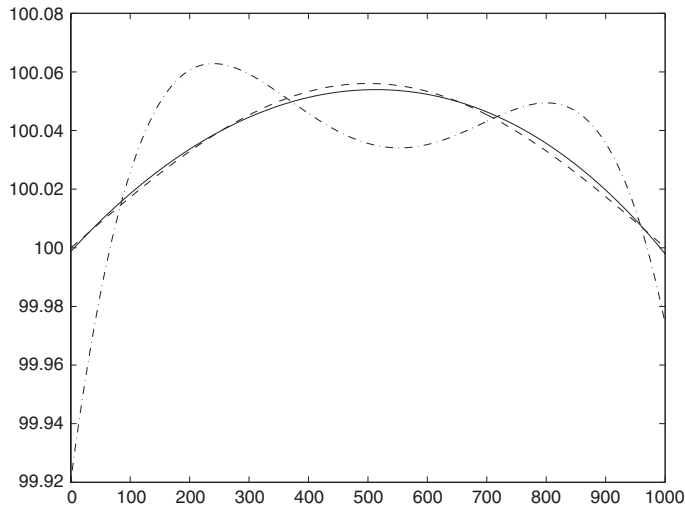


Figure 2. Example 4.3. Iterate \mathbf{x}_2 determined by RRGMRES augmented by \mathcal{W}_1 (continuous curve), iterate \mathbf{x}_4 determined by standard RRGMRES (dash-dotted curve), and exact solution $\hat{\mathbf{x}}$ (dashed curve). The iterate \mathbf{x}_2 is seen to provide a more accurate approximation of $\hat{\mathbf{x}}$ than \mathbf{x}_4 .

Example 4.4

Consider the Fredholm integral equation of the first kind

$$\int_0^1 k(s, t)x(t) dt = \exp(s) + (1 - e)s - 1, \quad 0 \leq s \leq 1 \tag{26}$$

where

$$k(s, t) = \begin{cases} s(t - 1), & s < t \\ t(s - 1), & s \geq t \end{cases}$$

We discretize the integral equation by a Galerkin method with orthonormal box functions as test and trial functions using the Matlab program `deriv2` from [20]. The program yields a symmetric indefinite matrix $A \in \mathbb{R}^{200 \times 200}$ and a scaled discrete approximation $\hat{\mathbf{x}} \in \mathbb{R}^{200}$ of the solution $x(t) = \exp(t)$ of (26). The error-free right-hand side vector is given by $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$, and the right-hand side vector \mathbf{b}' in (1) is determined by (18) with $\varepsilon = 1 \times 10^{-3}$ in (23).

Table III displays the performance of standard and augmented RRGMRES, and Figure 3 shows the best approximations of $\hat{\mathbf{x}}$ computed by RRGMRES with and without augmentation. Clearly, augmentation reduces the error significantly.

The success of augmentation depends on the augmenting spaces chosen. For some linear discrete ill-posed problems, the approximate solution determined by standard GMRES or RRGMRES may suggest a suitable augmenting space. In the present example, the computed approximate solution \mathbf{x}_{10} determined by standard RRGMRES is increasing if we ignore the wiggles; see Figure 3(a). This suggests that RRGMRES augmented by \mathcal{W}_2 may yield a better approximation of $\hat{\mathbf{x}}$ than standard RRGMRES, because linearly increasing functions can be represented by elements in \mathcal{W}_2 . Moreover, since when ignoring the wiggles, \mathbf{x}_{10} increases faster than linearly, augmentation

Table III. Example 4.4. Comparison of standard and augmented RRGMRES.

Augmentation	$\ \hat{\mathbf{x}} - \mathbf{x}_{j^*}\ $	j^*	Mat.-vec. prod.
—	2.7×10^{-1}	10	11
\mathcal{W}_1	2.1×10^{-1}	10	12
\mathcal{W}_2	1.7×10^{-2}	7	10
\mathcal{W}_3	4.4×10^{-3}	2	6

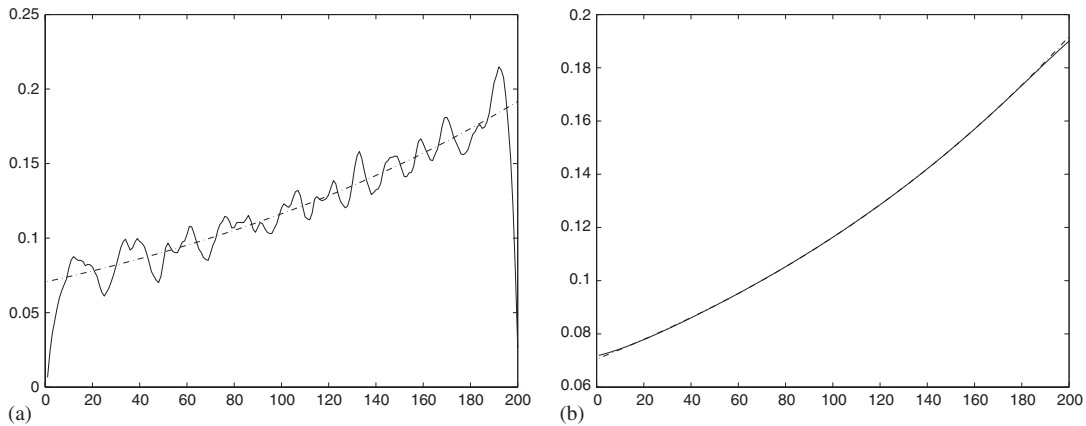


Figure 3. Example 4.4. (a) Approximate solution \mathbf{x}_{10} computed by standard RRGMRES (continuous curve) and exact solution $\hat{\mathbf{x}}$ (dashed curve); and (b) approximate solution \mathbf{x}_2 computed by RRGMRES augmented by \mathcal{W}_3 (continuous curve) and exact solution $\hat{\mathbf{x}}$ (dashed curve).

by \mathcal{W}_3 may give an even better approximation of $\hat{\mathbf{x}}$ than augmentation by \mathcal{W}_2 . Table III shows this indeed to be the case.

We remark that augmented RRGMRES does not utilize the symmetry of the matrix A . It is possible to develop a symmetry-exploiting method using the approach described in [11].

Example 4.5

A standard test example for linear discrete ill-posed problems is the solution of equation (26) with the right-hand side replaced by $(s^3 - s)/6$. Then the solution is given by $x(t) = t$. We determine the right-hand side \mathbf{b}' as in Example 4.4. Standard RRGMRES yields $j^* = 10$ and $\|\hat{\mathbf{x}} - \mathbf{x}_{j^*}\| = 9.0 \times 10^{-2}$, and RRGMRES augmented by \mathcal{W}_2 gives $j^* = 1$ and $\|\hat{\mathbf{x}} - \mathbf{x}_{j^*}\| = 2.6 \times 10^{-4}$. We remark that $\mathbf{x}_{j^*} \neq \hat{\mathbf{x}}$ because of the error \mathbf{e} in \mathbf{b}' and round-off errors introduced during the computations.

5. CONCLUSION

The computed examples illustrate that it can be worthwhile to augment Krylov subspaces generated by GMRES and RRGMRES by linear spaces that are not defined by Ritz vectors. When the

linear system of equations arises from the discretization of a well-posed problem, augmentation can reduce the number of iterations required to determine an approximate solution of desired accuracy. Augmented RRGMRRES applied to the solution of linear discrete ill-posed problems with a contaminated right-hand side can yield approximations of the desired solution of the error-free linear system of higher accuracy than standard RRGMRRES.

ACKNOWLEDGEMENTS

We would like to thank a referee for valuable comments.

REFERENCES

1. Saad Y. *Iterative Methods for Sparse Linear Systems* (2nd edn). SIAM: Philadelphia, 2003.
2. Saad Y, Schultz MH. GMRES: a generalized minimal residual method for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:856–869.
3. Calvetti D, Lewis B, Reichel L. GMRES-type methods for inconsistent systems. *Linear Algebra and its Applications* 2000; **316**:157–169.
4. Calvetti D, Lewis B, Reichel L. On the choice of subspace for iterative methods for linear discrete ill-posed problems. *International Journal of Applied Mathematics and Computer Science* 2001; **11**:1069–1092.
5. Reichel L, Ye Q. Breakdown-free GMRES for singular systems. *SIAM Journal on Matrix Analysis and Applications* 2005; **26**:1001–1021.
6. Chapman A, Saad Y. Deflated and augmented Krylov subspace techniques. *Numerical Linear Algebra with Applications* 1997; **4**:43–66.
7. Morgan RB. A restarted GMRES method augmented with eigenvectors. *SIAM Journal on Matrix Analysis and Applications* 1995; **16**:1154–1171.
8. Morgan RB. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1112–1135.
9. Morgan RB. GMRES with deflated restarting. *SIAM Journal on Scientific Computing* 2002; **24**:20–37.
10. Saad Y. Analysis of augmented Krylov subspace methods. *SIAM Journal on Matrix Analysis and Applications* 1997; **18**:435–449.
11. Calvetti D, Reichel L, Shuibi A. Enriched Krylov subspace methods for ill-posed problems. *Linear Algebra and its Applications* 2003; **362**:257–273.
12. Baglama J, Reichel L. Decomposition methods for large linear discrete ill-posed problems. *Journal of Computational and Applied Mathematics* 2007; **198**:332–343.
13. Hansen PC. *Rank Deficient and Discrete Ill-Posed Problems*. SIAM: Philadelphia, 1998.
14. Calvetti D, Lewis B, Reichel L. On the regularizing properties of the GMRES method. *Numerische Mathematik* 2002; **91**:605–625.
15. Calvetti D, Lewis B, Reichel L. GMRES, L-curves and discrete ill-posed problems. *BIT* 2002; **42**:44–65.
16. Hanke M. *Conjugate Gradient Type Methods for Ill-Posed Problems*. Longman: Harlow, 1995.
17. Hanke M, Hansen PC. Regularization methods for large-scale problems. *Surveys on Mathematics for Industry* 1993; **3**:253–315.
18. Hansen PC, Mosegaard K. Piecewise polynomial solutions without *a priori* break points. *Numerical Linear Algebra with Applications* 1996; **3**:513–524.
19. Baart ML. The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems. *IMA Journal of Numerical Analysis* 1982; **2**:241–247.
20. Hansen PC. Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms* 1994; **6**:1–35.