

The adaptive augmented GMRES method for solving ill-posed problems

Nao Kuroiwa¹

Takashi Nodera²

(Received 14 August 2008; revised 18 December 2008)

Abstract

The GMRES method is an iterative method that provides better solutions when dealing with large linear systems of equations with a non-symmetric coefficient matrix. The GMRES method generates a Krylov subspace for the solution, and the augmented GMRES method allows augmentation of the Krylov subspaces by a user supplied subspace which represents certain known features of the desired solution. The augmented GMRES method performs well with suitable augmentation, but performs poorly with unsuitable augmentation. The adaptive augmented GMRES method automatically selects a suitable subspace from a set of candidates supplied by the user. This study shows that this method maintains the performance level of augmented GMRES and lightens the burden it puts on its users. Numerical experiments compare robustness as well as the efficiency of various heuristic strategies.

Contents

1	Introduction	C655
2	Iterative methods for solving linear discrete ill-posed problems	C656
3	Adaptive augmented GMRES and RRGMRES method	C658
4	Numerical experiments	C660
4.1	Experiment 1	C662
4.2	Experiment 2	C664
5	Conclusion	C666
	References	C666

1 Introduction

This article explores iterative methods for solving large linear systems of equations,

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

which have a coefficient matrix \mathbf{A} of ill-determined rank. That is, matrix \mathbf{A} has many singular values of different orders of magnitude close to the origin. In particular, \mathbf{A} is severely ill-conditioned. Some of the singular values of \mathbf{A} may be vanishing.

When linear ill-posed problems such as the first kind Fredholm integral equations are discretized it results in a linear system of equations like (1) with a coefficient matrix of ill-determined rank. These arise in many inverse problems; for example, in computations of the magnetization inside a volcano,

and in computations of sharp images from blurred ones. In this study, the linear systems of equations with a matrix of ill-determined rank are referred to as linear discrete ill-posed problems and are based on that developed by Hansen [7].

In many linear discrete ill-posed problems, the right-hand side \mathbf{b} is contaminated by a measurement and discretization error $\mathbf{e} \in \mathbb{R}^n$, that is,

$$\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}, \quad (2)$$

where $\hat{\mathbf{b}} \in \mathbb{R}^n$ is an unknown error-free vector. Assume that the linear system of equations has an unknown error-free right-hand side where,

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}, \quad (3)$$

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

When obtaining a solution $\hat{\mathbf{x}}$ of equation (3), for instance, the least-squares solution of the minimal Euclidean norm, the right-hand side $\hat{\mathbf{b}}$ of equation (3) is unknown. It becomes necessary to determine an approximation of $\hat{\mathbf{x}}$ by computing a j th iterate solution \mathbf{x}_j of the available linear system of equations (1). The final approximate solution \mathbf{x}_j^* is determined by

$$\|\mathbf{x}_j^* - \hat{\mathbf{x}}\| = \min \|\mathbf{x}_j - \hat{\mathbf{x}}\|, \quad (4)$$

where $\|\cdot\|$ denotes the Euclidean norm.

2 Iterative methods for solving linear discrete ill-posed problems

Recently, iterative methods for creating an approximate solution for large scale linear discrete ill-posed problems like (1) have received a lot of attention. Calvetti et al. [4] explored this extensively.

The GMRES method by Saad and Schulz [8] is one of the popular iterative methods for solving linear systems of equations (1) with a non-symmetric matrix \mathbf{A} .

Let \mathbf{x}_0 be an initial approximate solution of (1) and $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$ be the initial residual vector. The Krylov subspaces are

$$\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) = \text{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \dots, \mathbf{A}^{j-1}\mathbf{r}_0\}. \quad (5)$$

The j th iterate \mathbf{x}_j satisfies

$$\|\mathbf{A}\mathbf{x}_j - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|, \quad (6)$$

and is an approximate solution.

The range restricted GMRES (RRGMRES) method by Calvetti et al. [3] differs from the normal GMRES method in that the Krylov subspace is restricted to the range of the coefficient matrix \mathbf{A} . To obtain an approximate solution \mathbf{x} we solve the least-squares problem

$$\|\mathbf{A}\mathbf{x}_j - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{A}\mathbf{r}_0)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|, \quad (7)$$

where $\mathcal{K}_j(\mathbf{A}, \mathbf{A}\mathbf{r}_0) = \mathbf{A}\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$. Calvetti et al. [4] showed that RRGMRRES works better than GMRES when solving linear discrete ill-posed problems.

Baglama and Reichel [2] presented an augmented GMRES method where the approximate solution \mathbf{x}_j satisfies

$$\|\mathbf{A}\mathbf{x}_j - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) + \mathcal{W}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|, \quad (8)$$

and an augmented RRGMRRES method

$$\|\mathbf{A}\mathbf{x}_j - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{A}\mathbf{r}_0) + \mathcal{W}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|. \quad (9)$$

These methods add a user supplied space \mathcal{W} to the Krylov subspaces generated by normal GMRES and RRGMRRES respectively. These methods create

an approximate solution with augmented Krylov subspaces. Baglama [2] describes the details of user supplied spaces. This helps the Krylov subspaces to represent features of the solution and to create an approximation which is sufficiently accurate. Based on Baglama [2], we prepare the spaces for augmentation as

$$W_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}, \quad W_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix}, \quad (10)$$

$$W_i = \text{range } W_k, \quad 1 \leq i \leq 3. \quad (11)$$

An algorithm of the augmented GMRES and RRGMRES is shown by Baglama [2, Algorithm 2.1]. In this article, the first $\mathbf{x}_0 = \mathbf{0}$, but the following $\mathbf{x}_0 \neq \mathbf{0}$ when we restart.

The problem with these methods is that it is difficult to choose the best space for augmentation. First of all, the choice of the space depends on the features of the exact solution. Secondly, if we added an inadequate space, we may achieve a less acceptable result than the normal GMRES or RRGMRES methods. This means that if the features of the exact solution are not known, it is a waste of time to use the augmented GMRES and RRGMRES methods. As a solution for overcoming these issues, an adaptive augmented GMRES and RRGMRES method, and a restarted version of them were developed.

3 Adaptive augmented GMRES and RRGMRES method

This section explains how our adaptive augmented GMRES and RRGMRES method are implemented.

After preparing some spaces, the adaptive method automatically selects the space that is augmented through certain selecting conditions, even if the features of the exact solution are unknown. The user needs to define a number of candidate spaces. Based on this, the adaptive algorithm chooses the best spaces. Here is an example of how this works

The first step is to prepare some spaces \mathcal{W}_i , $1 \leq i \leq l$, for augmentation and apply the skinny QR-factorization [5]

$$A\mathcal{W}_i = \mathbf{V}_{p_i} \mathbf{R}_i, \quad (12)$$

where $\mathcal{W}_i \in \mathbb{R}^{n \times p_i}$ is the basis of \mathcal{W}_i , the matrices $\mathbf{V}_{p_i} \in \mathbb{R}^{n \times p_i}$ and have orthonormal columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{p_i}$, and $\mathbf{R}_i \in \mathbb{R}^{p_i \times p_i}$ is upper triangular.

The second step is to use the initial vector of the Modified Arnoldi's method for choosing the augmentation space, which determines our selecting condition. GMRES generates Krylov subspaces by Modified Arnoldi's method and creates an approximate solution. The final residual norm of this approximate solution is minimal. Thus, the smaller the norm of the initial vector of Modified Arnoldi's method, the smaller the final residual norm is.

We then calculate the norm of the initial vector of Modified Arnoldi's method utilising the augmented GMRES

$$\|(\mathbf{I} - \mathbf{V}_{p_i} \mathbf{V}_{p_i}^T) \mathbf{r}_0\|,$$

for each of the l spaces. After this, the magnitudes and the initial residual \mathbf{r}_0 that we use if we augment no space, are compared. If the space \mathcal{W}_i is the minimum norm, we augment with \mathcal{W}_i and set it so that $p = p_i$. If the norm of initial residual \mathbf{r}_0 is the minimum, normal GMRES is implemented.

At this point, we introduce another least-squares problem by modification of (8). We chose the space \mathcal{W} for augmentation in the second step, so we execute the modified Arnoldi's method. From this, we obtain

$$A[\mathcal{W}_i \mathbf{V}_{p+1:p+j}] = \mathbf{V}_{p+j+1} \mathbf{H}, \quad (13)$$

where $\mathbf{V}_{p+j+1} = [\mathbf{V}_p \mathbf{V}_{p+1:p+j+1}] \in \mathbb{R}^{n \times (p+j+1)}$ has orthonormal columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$, and the first column of submatrix $\mathbf{V}_{p+1:p+j+1}$ is $(\mathbf{I} - \mathbf{V}_p \mathbf{V}_p^T) \mathbf{b} / \|(\mathbf{I} - \mathbf{V}_p \mathbf{V}_p^T) \mathbf{b}\|$ for adaptive augmented GMRES and $(\mathbf{I} - \mathbf{V}_p \mathbf{V}_p^T) \mathbf{A} \mathbf{b} / \|(\mathbf{I} - \mathbf{V}_p \mathbf{V}_p^T) \mathbf{A} \mathbf{b}\|$ for the adaptive augmented RRGMRRES. The modified Arnoldi process determines the remaining columns of $\mathbf{V}_{p+1:p+j+1}$. \mathbf{H} is the upper Hessenberg matrix, and its $p \times p$ principal submatrix is \mathbf{R} in the QR-factorization (12).

Since for all $\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) + \mathcal{W}$ is represented as

$$\mathbf{x} = \mathbf{x}_0 + [\mathbf{W} \mathbf{V}_{p+1:p+j+1}] \mathbf{y}, \quad \mathbf{y} \in \mathbb{R}^n, \tag{14}$$

we translate residual norms by using this Arnoldi decomposition (13):

$$\begin{aligned} \|\mathbf{b} - \mathbf{A} \mathbf{x}\| &= \|\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + [\mathbf{W} \mathbf{V}_{p+1:p+j+1}] \mathbf{y})\| \\ &= \|\mathbf{r}_0 - \mathbf{V}_{p+j+1} \mathbf{H} \mathbf{y}\| \\ &= \|\mathbf{V}_{p+j+1}^T \mathbf{r}_0 - \mathbf{H} \mathbf{y}\|. \end{aligned}$$

The iterate \mathbf{x}_j is determined as

$$\mathbf{x}_j = \mathbf{x}_0 + [\mathbf{W} \mathbf{V}_{p+1:p+j+1}] \mathbf{y}_j, \tag{15}$$

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

$$\min_{\mathbf{y} \in \mathbb{R}^{p+j}} \|\mathbf{V}_{p+j+1}^T \mathbf{r}_0 - \mathbf{H} \mathbf{y}\|. \tag{16}$$

When restarting the adaptive algorithm, it is necessary to reset the initial guess \mathbf{x}_0 to be the obtained approximate solution \mathbf{x}_j and start by selecting the space for augmenting. In this article the restarted method is referred to as the adaptive augmented GMRES(j) and RRGMRRES(j) method.

4 Numerical experiments

The purpose of this section is to illustrate the performance of the adaptive augmented GMRES and RRGMRRES method for providing solutions to ill-posed

Algorithm 1 Adaptive augmented GMRES and RRGMRES method.

Input: $\mathbf{x}_0, A \in \mathbb{R}^{n \times n}, \mathbf{b} \in \mathbb{R}^n, W_i \in \mathbb{R}^{n \times p_i}, 1 \leq i \leq l, j$
Output: \mathbf{x}_j
 $\mathbf{r}_0 := \mathbf{b} - A\mathbf{x}_0, \quad \mathbf{v}_{p+1} := \mathbf{r}_0$
if RRGMRES **then**
 $\mathbf{v}_{p+1} := A\mathbf{r}_0$
end if
for $i = 1, \dots, l$ **do**
 $AW_i = V_{p_i} H_i$
 $\mathbf{v}_{p_i+1} := (I - V_{p_i} V_{p_i}^T) \mathbf{v}_{p+1}$
end for
 $m := \arg \min_{1 \leq i \leq l} \|\mathbf{v}_{p_i+1}\|$
if $\|\mathbf{v}_{p_l+1}\| < \|\mathbf{v}_{p+1}\|$ **then**
 $p := p_m$
 $\mathbf{v}_{p+1} := \mathbf{v}_{p_m+1}$
 $V_p := V_{p_m}$
else
 $p = 0$
 $V_p :=$ matrix with no columns

end if
for $k = p + 1, \dots, p + j$ **do**
 $\mathbf{w}_k := A\mathbf{v}_k$
for $i = 1, \dots, k$ **do**
 $h_{ik} := (\mathbf{w}_k, \mathbf{v}_i)$
 $\mathbf{w}_k := \mathbf{w}_k - h_{ik} \mathbf{v}_i$
end for
 $h_{k+1,k} := \|\mathbf{w}_k\|$
 $\mathbf{v}_{k+1} := \mathbf{w}_{k+1}/h_{k+1,k}, V_{k+1} := [V_k \mathbf{v}_{k+1}]$
end for
Compute: \mathbf{y}_j of $\min_{\mathbf{y} \in \mathbb{R}^{p+j}} \|V_{p+j+1}^T \mathbf{b} - H\mathbf{y}\|$
 $\mathbf{x}_j := \mathbf{x}_0 + [WV_{p+1:p+j}] \mathbf{y}_j$

problems, particularly linear discrete ill-posed problems in numerical experiments.

We run restarted algorithms on C using a hp Compaq ProLiant DL145 machine with $\epsilon_M = 2.22 \times 10^{-16}$. The maximum iteration is set at 100.

4.1 Experiment 1

The first experiment focuses on the Fredholm integral equation of the first kind

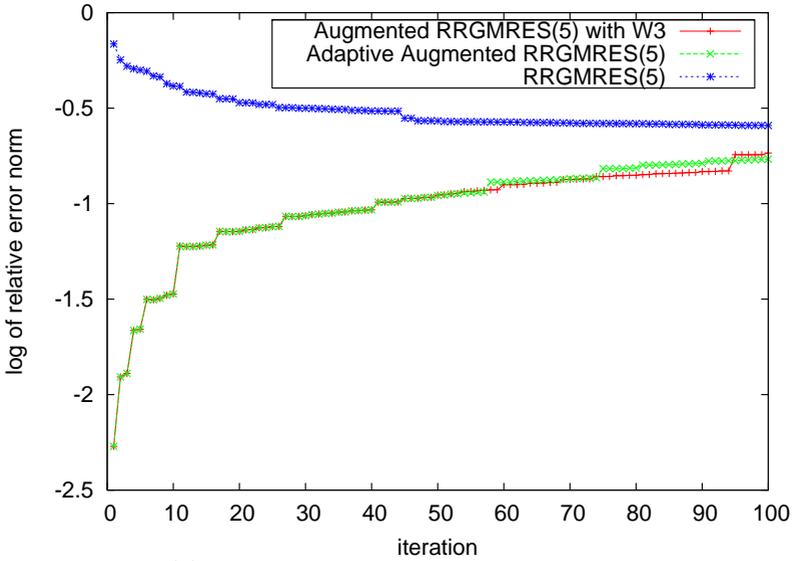
$$\int_0^1 K(s, t)x(t) dt = e^s + (1 - e)s - 1, \quad 0 \leq s \leq 1, \quad (17)$$

$$K(s, t) = \begin{cases} s(t - 1), & s < t, \\ t(s - 1), & s \geq t, \end{cases} \quad (18)$$

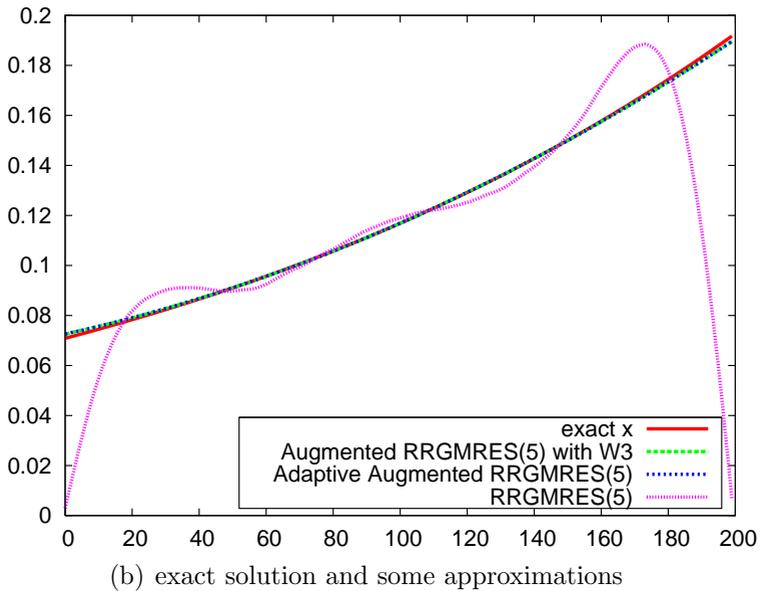
where the solution $x(t) = \exp(t)$. We discretize the equation using the Matlab program `deriv2` from test functions in the Regularization Tools [6]. Then we obtain a symmetric matrix $A \in \mathbb{R}^{200 \times 200}$ and $\hat{\mathbf{x}} \in \mathbb{R}^{200}$. The error vector is set to $\mathbf{e} \in \mathbb{R}^{200}$. This would consist of normally distributed random entries of zero mean and variance $1/200^2$. $\|\mathbf{e}\| = 3.56 \times 10^{-4}$. Set it so that $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$, and the right-hand side $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}$ in (1).

Since the RRGMRRES method works better than the GMRES method, we compare adaptive augmented RRGMRRES (5), augmented RRGMRRES (5) augmented by \mathcal{W}_3 , which is the best space for augmentation, and standard RRGMRRES (5).

Table 1 and Figure 1(a) show that the results of the adaptive method is almost equal to that of the normal augmented method and better than standard RRGMRRES. In Figure 1(b), adaptive augmented RRGMRRES(5) and augmented RRGMRRES(5) with \mathcal{W}_3 approximate the exact solution well.



(a) relative error norm versus iterations



(b) exact solution and some approximations

FIGURE 1: Experiment 1

TABLE 1: Experiment 1

Method	Iteration	$\ \mathbf{x}_j^* - \hat{\mathbf{x}}\ $
RRGMRES(5)	100	4.59×10^{-1}
Augmented RRGMRRES(5) with \mathcal{W}_3	1	9.55×10^{-3}
Adaptive Augmented RRGMRRES(5)	1	9.57×10^{-3}

4.2 Experiment 2

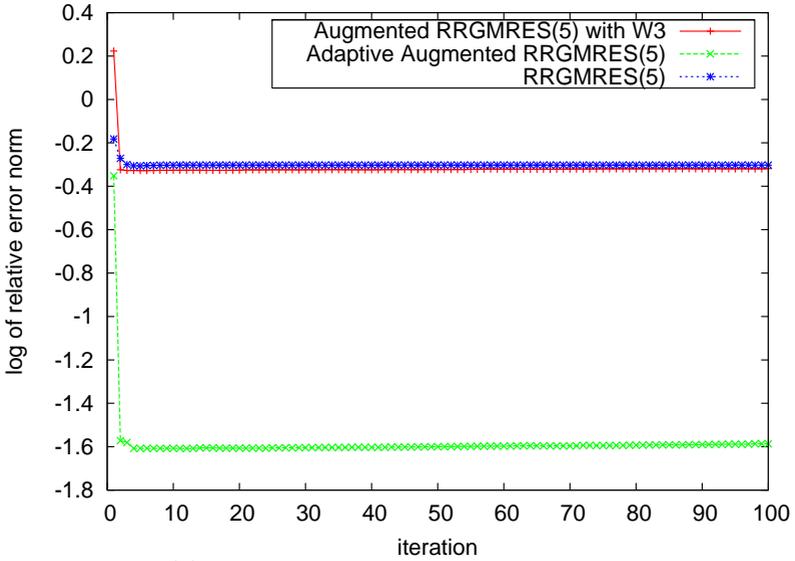
The second experiment has to do with a Fredholm integral equation of the first kind by Baart [1]:

$$\int_0^\pi \exp[s \cos(t)] \chi(t) dt = \mathbf{b}(s) = 2 \frac{\sinh(s)}{s}, \quad 0 \leq s \leq \frac{\pi}{2}, \quad (19)$$

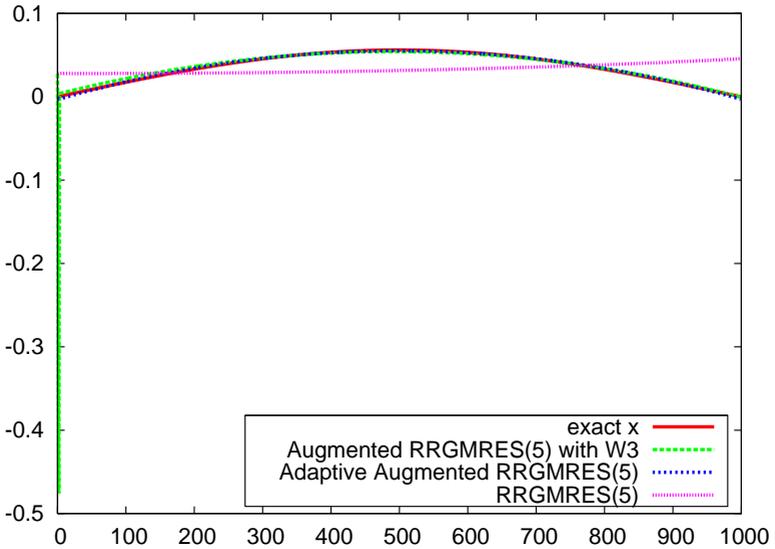
The solution is $\chi(t) = \sin(t)$. It is discretized using the Matlab program `baart` from test functions in the Regularization Tools [6], after which a non-symmetric coefficient matrix $\mathbf{A} \in \mathbb{R}^{1000 \times 1000}$ and $\hat{\mathbf{x}} \in \mathbb{R}^{1000}$ are obtained. An error vector $\mathbf{e} \in \mathbb{R}^{1000}$ consisting of normally distributed random entries of zero mean and variance $1/1000^2$, and $\|\mathbf{e}\| = 3.04 \times 10^{-5}$ were determined. Here, $\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}}$, and the right-hand side $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}$ in equation (1).

Since the performance of the RRGMRRES method is better than GMRES, we make a comparison using the same methods as Experiment 1.

Figure 2(a) indicates that our adaptive method performs better than the other methods. In Table 2, the error norm of adaptive augmented RRGMRRES(5) is about 1/10 that of augmented RRGMRRES with \mathcal{W}_3 . Figure 2(b) illustrates that the approximate solution of the adaptive augmented RRGMRRES(5) is closer to the exact solution than that of the augmented RRGMRRES(5) with \mathcal{W}_3 .



(a) relative error norm versus iterations



(b) exact solution and some approximations

FIGURE 2: Experiment 2

TABLE 2: Experiment 2

method	iteration	$\ \mathbf{x}_j^* - \hat{\mathbf{x}}\ $
RRGMRES(5)	4	6.20×10^{-1}
Augmented RRGMRRES(5) with \mathcal{W}_3	6	5.90×10^{-1}
Adaptive augmented RRGMRRES(5)	12	3.09×10^{-2}

5 Conclusion

Exploring the adaptive augmented GMRES and RRGMRRES method, we find that the numerical experiments for linear discrete ill-posed problems in Section 4 ensure that our adaptive augmented method maintains at the least and even improves on the performance of the normal augmented method in some cases.

References

- [1] M. L. Baart. The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned linear least-squares problems. *IMA Journal of Numerical Analysis*, 2:241–247, 1982. doi:10.1093/imanum/2.2.241 C664
- [2] J. Baglama and L. Reichel. Augmented GMRES-type method. *Numerical Linear Algebra with Applications*, 14:337–350, 2007. doi:10.1002/nla.518 C657, C658
- [3] D. Calvetti, B. Lewis, and L. Reichel. GMRES-type methods for inconsistent systems. *Linear Algebra and its Applications*, 316:157–169, 2000. doi:10.1016/S0024-3795(00)00064-1 C657
- [4] D. Calvetti, B. Lewis, and L. Reichel. On the choice of subspace for iterative methods for linear discrete ill-posed problems. *Int. J. Appl.*

- Math. Comput. Sci*, 11:1069–1092, 2001. <http://www.math.kent.edu/~reichel/publications/subspaceselect.pdf>
C656, C657
- [5] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 2nd edition, 1989. C659
- [6] P. C. Hansen. Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6:1–35, 1994. [doi:10.1007/BF02149761](https://doi.org/10.1007/BF02149761) C662, C664
- [7] P. C. Hansen. *Rank Deficient and Discrete Ill-Posed Problems*. SIAM, Philadelphia, 1998. C656
- [8] Y. Saad. and M. H. Schultz. GMRES: A generalized minimal residual method for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.
<http://www.stanford.edu/class/cme324/saad-schultz.pdf> C657

Author addresses

1. **Nao Kuroiwa**, Graduate School of Science and Technology, Keio University, JAPAN.
<mailto:kuroiwa@math.keio.ac.jp>
2. **Takashi Nodera**, Department of Mathematics, Faculty of Science and Technology, Keio University, JAPAN.
<mailto:nodera@math.keio.ac.jp>