ISSN 1346-5597

National Institute of Informatics

NII Technical Report

GMRES Methods for Least Squares Problems

Ken HAYAMI, Jun-Feng YIN, and Tokushi ITO

NII-2007-009E
July 2007

URL: http://www.nii.ac.jp/ Tokyo 101-8430, JAPAN +81-3-4212-2000 FAX: +81-3-3556-1916

GMRES METHODS FOR LEAST SQUARES PROBLEMS

KEN HAYAMI*, JUN-FENG YIN†, AND TOKUSHI ITO‡

Abstract.
The standard iterative method for solving large sparse least squares problems $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$, $A \in \mathbf{R}^{m\times n}$ is the CGLS method, or its stabilized version LSQR, which applies the (preconditioned) conjugate gradient method to the normal equation $A^{\mathrm{T}}A\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$.

In this paper, we will consider alternative methods using a matrix $B \in \mathbf{R}^{n\times m}$ and applying the Generalized Minimal Residual (GMRES) method to $\min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2$ or $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$.

Next, we give a sufficient condition concerning B for the GMRES methods to give a least squares solution without breakdown for arbitrary $\boldsymbol{b}$, for over-determined, under-determined and possibly rank-deficient problems. We then give a convergence analysis of the GMRES methods as well as the CGLS method.

Then, we propose using the robust incomplete factorization (RIF) for B.

Finally, we show by numerical experiments on over-determined and under-determined problems that, for ill-conditioned problems, the GMRES methods with RIF give least squares solutions faster than the CGLS and LSQR methods with RIF.

Key words. least squares problems, iterative method, Krylov subspace method, GMRES, CGLS, LSQR, robust incomplete factorization.

AMS subject classifications. Primary: 65F10, 65F20, 65F50; Secondary: 15A06, 15A09

1. Introduction.

Consider the least squares problem

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 \tag{1.1}$$

where $A \in \mathbf{R}^{m\times n}$, and $m \geq n$ or $m < n$. We also allow the rank-deficient case when the equality in $\operatorname{rank} A \leq \min(m, n)$ does not hold.

The least squares problem (1.1) is equivalent to the normal equation

$$A^{\mathrm{T}}A\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}. \tag{1.2}$$

For $m < n$,

$$AA^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{b}, \quad \boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{y} \tag{1.3}$$

gives the minimum norm solution of (1.1).

The standard direct method for solving the least squares problem (1.1) (where $m \geq n$ and A is full rank) is to use the QR decomposition: $A = QR$ where $Q \in \mathbf{R}^{m\times n}$ is an orthogonal matrix and $R \in \mathbf{R}^{n\times n}$ is an upper triangular matrix, which can be obtained using the Householder or Givens transformations, or by the modified Gram-Schmidt method. Then, equation (1.2) is transformed as $R^{\mathrm{T}}R\boldsymbol{x} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$. If $\operatorname{rank} A = n$, R is nonsingular, so that $R\boldsymbol{x} = Q^{\mathrm{T}}\boldsymbol{b}$, and back substitution gives the least squares solution $\boldsymbol{x}$ of (1.1). When A is large and sparse, techniques are used to save memory and computation time[4].

*National Institute of Informatics, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan, and Department of Informatics, The Graduate University for Advanced Studies (Sokendai), E-mail: hayami@nii.ac.jp

†National Institute of Informatics, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan, E-mail: yinjf@nii.ac.jp

‡Business Design Laboratory Co. Ltd., Design Center Bldg. 7th floor, NADYA Park, 18-1 Sakae 3-chome, Naka-ku, Nagoya-shi, Aichi 460-0008, Japan, E-mail: ito@business-design.co.jp

However, for very large and sparse problems, iterative methods become necessary, among which the (preconditioned) Conjugate Gradient Least Squares (CGLS) method[4] or its stabilized version LSQR[14] are most commonly used. This is based on the observation that, in the normal equation (1.2), the coefficient matrix $A^{\mathrm{T}}A$ is symmetric, and also positive definite if $\operatorname{rank} A = n$, so that it is natural to apply the conjugate gradient method to (1.2). The preferred implementation is the following[10, 4].

Method 1.1. **The CGLS(CGNR) method**

Choose $\boldsymbol{x}_0$.
$\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0, \quad \boldsymbol{p}_0 = \boldsymbol{s}_0 = A^T\boldsymbol{r}_0, \quad \gamma_0 = \|\boldsymbol{s}_0\|_2{}^2$
for $i = 0, 1, 2, \ldots$ *until* $\gamma_i < \epsilon$
  $\boldsymbol{q}_i = A\boldsymbol{p}_i$
  $\alpha_i = \gamma_i / \|\boldsymbol{q}_i\|_2{}^2$
  $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$
  $\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \alpha_i \boldsymbol{q}_i$
  $\boldsymbol{s}_{i+1} = A^T\boldsymbol{r}_{i+1}$
  $\gamma_{i+1} = \|\boldsymbol{s}_{i+1}\|_2{}^2$
  $\beta_i = \gamma_{i+1}/\gamma_i$
  $\boldsymbol{p}_{i+1} = \boldsymbol{s}_{i+1} + \beta_i \boldsymbol{p}_i$
endfor

In [5], different versions of the CGLS and the LSQR method are compared.

Here, let $A^\dagger$ be the generalized inverse of $A \in \mathbf{R}^{m\times n}$, $\operatorname{rank} A = r$, and let σ_1 and σ_r be the largest and smallest (nonzero) singular value of A, respectively. Then, the condition number of A is

$$\kappa(A) := \|A\|_2 \, \|A^\dagger\|_2 = \frac{\sigma_1}{\sigma_r},$$

and that of $A^{\mathrm{T}}A$ is

$$\kappa(A^{\mathrm{T}}A) = \left(\frac{\sigma_1}{\sigma_r}\right)^2 = \kappa(A)^2.$$

The convergence speed of the CGLS method is known to depend on $\kappa(A^{\mathrm{T}}A) = \kappa(A)^2$, including the case when A is rank deficient[4]. Hence, the convergence of the CGLS method may be slow for ill-conditioned problems, so that preconditioning becomes necessary.

A simple preconditioning is diagonal scaling using the diagonal elements of $A^{\mathrm{T}}A$. More sophisticated preconditionings are, for example, the incomplete Cholesky decomposition [13], incomplete QR decompositions using incomplete modified Gram-Schmidt methods e.g. [12, 17, 22] or incomplete Givens orthogonalizations[1, 15], and the robust incomplete factorization (RIF)[3].

For instance, the incomplete QR decomposition $A \sim QR$ can be used to precondition the normal equation (1.2) as

$$\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}, \tag{1.4}$$

where $\tilde{A} = R^{-\mathrm{T}}A^{\mathrm{T}}AR^{-1}$, $\tilde{\boldsymbol{x}} = R\boldsymbol{x}$, $\tilde{\boldsymbol{b}} = R^{-\mathrm{T}}A^{\mathrm{T}}\boldsymbol{b}$, after which the CGLS method is applied to equation (1.4).

Even with preconditioning, the convergence behaviour may deteriorate for highly ill-conditioned problems due to rounding errors, as will be observed in our numerical experiments.

On the other hand, Zhang and Oyanagi[23] proposed applying the Orthomin(k) method directly to the least squares problem (1.1), instead of treating the normal equation (1.2). This was done by introducing a mapping matrix $B \in \mathbf{R}^{n\times m}$ to transform the problem to a system with a square coefficient matrix $AB \in \mathbf{R}^{m\times m}$, and then applying the Krylov subspace method Orthomin(k) to this nonsymmetric system.

In [11, 8, 9], we further extended their method by applying the GMRES method instead of the Orthomin(k) method, and also introduced an alternative method of treating the system with a coefficient matrix $BA \in \mathbf{R}^{n\times n}$, and gave a sufficient condition concerning B for the methods to give the least squares solution without breakdown for over-determined full rank systems.

Similar methods were also proposed by Vuik et al. in [21]. A related idea can also be found in Tanabe[20] for linear stationary iterations.

In this paper, we will further extend the analysis to rank-deficient as well as under-determined systems, and derive a sufficient condition concerning B for the general case. We will also give some convergence analysis for the GMRES methods as well as the CGLS method. Then, we propose using the robust incomplete factorization of Benzi and Tuma[3] for B. Finally, we give numerical experiment results for over-determined and under-determined systems, showing that the GMRES based methods using RIF for B are faster than the CGLS or LSQR methods with RIF for ill-conditioned problems.

The rest of the paper is organized as follows. In section 2, we briefly review the GMRES method. In section 3, we present the AB- and BA-GMRES methods for over- and under-determined least squares problems, and give a sufficient condition for B. In section 4, we discuss the properties of the eigenvalues of AB and BA, and give some convergence analysis of the GMRES methods as well as the preconditioned CGLS method. In section 5, we propose using RIF for B. In section 6, numerical experiment results are presented. Section 7 concludes the paper.

2. The GMRES method. The Generalized Minimal Residual (GMRES) method[19] is an efficient and robust Krylov subspace iterative method for solving systems of linear equations $A\boldsymbol{x} = \boldsymbol{b}$, where $A \in \mathbf{R}^{n\times n}$ is nonsingular and nonsymmetric. Usually, the method is implemented with restarts in order to reduce storage and computation time, as in the GMRES(k) method below. ($k = \infty$ corresponds to the original full GMRES method.)

Method 2.1. **The GMRES(k) method**

Choose $\boldsymbol{x}_0$.
$*$ $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
$\boldsymbol{v}_1 = \boldsymbol{r}_0/\|\boldsymbol{r}_0\|_2$
for $i = 1, 2, \ldots, k$
  $\boldsymbol{w}_i = A\boldsymbol{v}_i$
  for $j = 1, 2, \ldots, i$
    $h_{j,i} = (\boldsymbol{w}_i, \boldsymbol{v}_j)$
    $\boldsymbol{w}_i = \boldsymbol{w}_i - h_{j,i}\boldsymbol{v}_j$
  end for

$h_{i+1,i} = \|\boldsymbol{w}_i\|_2$
$\boldsymbol{v}_{i+1} = \boldsymbol{w}_i / h_{i+1,i}$
Find $\boldsymbol{y}_i \in \mathbf{R}^i$ *which minimizes* $\|\boldsymbol{r}_i\|_2 = \left\| \|\boldsymbol{r}_0\|_2 \boldsymbol{e}_i - \bar{H}_i \boldsymbol{y} \right\|_2$.
if $\|\boldsymbol{r}_i\|_2 < \epsilon$ *then*
$\boldsymbol{x}_i = \boldsymbol{x}_0 + [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i$
stop
endif
endfor
$\boldsymbol{x}_0 = \boldsymbol{x}_k$
Go to $*$.

Here, $\bar{H}_i = (h_{pq}) \in \mathbf{R}^{(i+1)\times i}$ and $\boldsymbol{e}_i = (1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbf{R}^{i+1}$. The method is designed to minimizes the L_2 norm of the residual $\|\boldsymbol{r}_k\|_2$ for all $\boldsymbol{x}_k = \boldsymbol{x}_0 + \langle \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \rangle$ where $\langle \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \rangle = \langle \boldsymbol{r}_0, A\boldsymbol{r}_0, \cdots, A^{k-1}\boldsymbol{r}_0 \rangle$, and $(\boldsymbol{v}_i, \boldsymbol{v}_j) = \delta_{ij}$. Here, $\langle \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \rangle$ denotes the vector space spanned by the vectors $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$. The method is said to break down when $h_{i+1,i} = 0$.

In the following, we assume exact arithmetic.

When A is nonsingular, the GMRES method gives the exact solution for all $\boldsymbol{b} \in \mathbf{R}^n$ and $\boldsymbol{x}_0 \in \mathbf{R}^n$ within n steps[19].

When A is singular, the following holds[6], where $\mathcal{R}(A)$ denotes the range space of A.

THEOREM 2.1. *Let* $A \in \mathbf{R}^{n\times n}$. *The GMRES method gives a solution to* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *without breakdown for arbitrary* $\boldsymbol{b} \in \mathbf{R}^n$ *and* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *if and only if* $\mathcal{R}(A) = \mathcal{R}(A^{\mathrm{T}})$.

3. Solution of least squares problems using the GMRES method. Consider applying the GMRES method directly to the least squares problem (1.1). This would require multiplying $A \in \mathbf{R}^{m\times n}$ and the residual vector $\boldsymbol{r}_0 \in \mathbf{R}^m$, which is not feasible. In the following, we give two methods for overcoming this difficulty using a matrix $B \in \mathbf{R}^{n\times m}$.

3.1. The AB-GMRES method. The first method is to use the Krylov subspace $\mathcal{K}_i(AB, \boldsymbol{r}_0) = \langle\, \boldsymbol{r}_0, AB\boldsymbol{r}_0, \ldots, (AB)^{i-1}\boldsymbol{r}_0 \,\rangle$ in $\mathbf{R}^m$ generated by $AB \in \mathbf{R}^{m\times m}$, as in [23], and to solve the least squares problem $\min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2$ using the GMRES method.

First, we will give a theoretical justification for doing so. In the following, let $\mathcal{N}(A)$ denote the null space of A, and $V^{\perp}$ the orthogonal complement of subspace V, respectively.

THEOREM 3.1. $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 = \min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2$ *holds for all* $\boldsymbol{b} \in \mathbf{R}^m$ *if and only if* $\mathcal{R}(A) = \mathcal{R}(AB)$.

Proof. The sufficiency is obvious. The necessity is shown as follows.
$\mathcal{R}(A) \neq \mathcal{R}(AB) \Longrightarrow \mathcal{R}(A) \supset \mathcal{R}(AB) \Longrightarrow {}^{\exists}\tilde{\boldsymbol{b}} \in \mathcal{R}(A)\backslash\mathcal{R}(AB)$
$\Longleftrightarrow {}^{\exists}\tilde{\boldsymbol{x}} \in \mathbf{R}^n;\quad \tilde{\boldsymbol{b}} = A\tilde{\boldsymbol{x}},\quad \tilde{\boldsymbol{b}} \neq AB\boldsymbol{z} \quad {}^{\forall}\boldsymbol{z} \in \mathbf{R}^m$
$\Longleftrightarrow 0 = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|\tilde{\boldsymbol{b}} - A\boldsymbol{x}\|_2 < \min_{\boldsymbol{z}\in\mathbf{R}^m} \|\tilde{\boldsymbol{b}} - AB\boldsymbol{z}\|_2.$ □

LEMMA 3.2. $\mathcal{R}(AA^{\mathrm{T}}) = \mathcal{R}(A)$.

Proof. $\mathcal{R}(AA^{\mathrm{T}}) = \{A\boldsymbol{x} \,|\, \boldsymbol{x} \in \mathcal{R}(A^{\mathrm{T}})\} = \{A\boldsymbol{x} \,|\, \boldsymbol{x} \in \mathcal{N}(A)^{\perp}\}$
$= \{A\boldsymbol{x} \,|\, \boldsymbol{x} \in \mathcal{N}(A)^{\perp} \cup \mathcal{N}(A)\} = \mathcal{R}(A).$ □

This gives the following.

LEMMA 3.3. $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B) \Longrightarrow \mathcal{R}(A) = \mathcal{R}(AB)$.

For instance, if $\operatorname{rank} A = \operatorname{rank} B = n$, then, $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B) = \mathbf{R}^n$ holds, and hence $\mathcal{R}(A) = \mathcal{R}(AB)$ holds.

Thus, assume $\mathcal{R}(A) = \mathcal{R}(AB)$ holds. Then, for arbitrary $\boldsymbol{x}_0 \in \mathbf{R}^n$, there exists a $\boldsymbol{z}_0 \in \mathbf{R}^m$ such that $A\boldsymbol{x}_0 = AB\boldsymbol{z}_0$, and $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0 = \boldsymbol{r}_0 - AB\boldsymbol{z}_0$. Hence, consider solving the least squares problem

$$\min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2 = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$$

using the GMRES(k) method, letting the initial approximate solution $\boldsymbol{z} = \boldsymbol{z}_0$ such that $AB\boldsymbol{z}_0 = A\boldsymbol{x}_0$. Then, we have the following algorithm.

Method 3.1. **The AB-GMRES(k) method**

Choose $\boldsymbol{x}_0$ $(A\boldsymbol{x}_0 = AB\boldsymbol{z}_0)$.
$*$ $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$ $(= \boldsymbol{b} - AB\boldsymbol{z}_0)$
$\boldsymbol{v}_1 = \boldsymbol{r}_0/\|\boldsymbol{r}_0\|_2$
for $i = 1, 2, \ldots, k$
  $\boldsymbol{w}_i = AB\boldsymbol{v}_i$
  for $j = 1, 2, \ldots, i$
    $h_{j,i} = (\boldsymbol{w}_i, \boldsymbol{v}_j)$
    $\boldsymbol{w}_i = \boldsymbol{w}_i - h_{j,i}\boldsymbol{v}_j$
  end for
  $h_{i+1,i} = \|\boldsymbol{w}_i\|_2$
  $\boldsymbol{v}_{i+1} = \boldsymbol{w}_i/h_{i+1,i}$
  Find $\boldsymbol{y}_i \in \mathbf{R}^i$ *which minimizes* $\|\boldsymbol{r}_i\|_2 = \left\| \|\boldsymbol{r}_0\|_2 \boldsymbol{e}_i - \bar{H}_i \boldsymbol{y} \right\|_2$
  $\boldsymbol{x}_i = \boldsymbol{x}_0 + B[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i$ $(\Longleftrightarrow \boldsymbol{z}_i = \boldsymbol{z}_0 + [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i)$
  $\boldsymbol{r}_i = \boldsymbol{b} - A\boldsymbol{x}_i$
  if $\|A^{\mathrm{T}}\boldsymbol{r}_i\|_2 < \epsilon$ *stop*
endfor
$\boldsymbol{x}_0 = \boldsymbol{x}_k$ $(\Longleftrightarrow \boldsymbol{z}_0 = \boldsymbol{z}_k)$
Go to $*$.

Note here that the convergence is assessed by explicitly computing $\|A^{\mathrm{T}}\boldsymbol{r}_i\|_2$ since $\|\boldsymbol{r}_i\|_2$ does not necessary converge to 0 for the general inconsistent case ($\boldsymbol{b} \notin \mathcal{R}(A)$).

The following hold.

THEOREM 3.4. *If* $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$, *then* $\mathcal{R}(AB) = \mathcal{R}(B^{\mathrm{T}}A^{\mathrm{T}}) \Longleftrightarrow \mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ *holds.*

Proof. If $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$, Lemma 3.2 gives $\mathcal{R}(AB) = \mathcal{R}(AA^{\mathrm{T}}) = \mathcal{R}(A)$ and $\mathcal{R}(B^{\mathrm{T}}A^{\mathrm{T}}) = \mathcal{R}(B^{\mathrm{T}}B) = \mathcal{R}(B^{\mathrm{T}})$. □

Let AB-GMRES method be the AB-GMRES(k) method with $k = \infty$ (no restarts). Then, Theorems 2.1 and 3.4 give the following.

THEOREM 3.5. *If* $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$ *holds, then the AB-GMRES method determines a least squares solution of* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *for all* $\boldsymbol{b} \in \mathbf{R}^m$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown if and only if* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$.

As a corollary, we have the following sufficient condition.

COROLLARY 3.6. *If* $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$ *and* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$, *the AB-GMRES method determines a least squares solution of* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *for all* $\boldsymbol{b} \in \mathbf{R}^m$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown.*

Note that the condition $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ is satisfied if B can be expressed as

$$B = CA^{\mathrm{T}} \tag{3.1}$$

where C is a nonsingular matrix.

We note here that Calvetti, Lewis and Reichel[7] proposed a related method for solving over-determined ($m \geq n$) least squares problems using the GMRES method. Their method is to append $(m-n)$ zero column vectors to the right side of the matrix A, to obtain a square singular matrix $\tilde{A} = [A, 0] \in \mathbf{R}^{m\times m}$, and applying the GMRES method to

$$\min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - \tilde{A}\boldsymbol{z}\|_2 \left(= \min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2\right).$$

This corresponds to a special case of our AB-GMRES method with $B = [\,\mathrm{I}_n, 0\,] \in \mathbf{R}^{n\times m}$ where $\mathrm{I}_n \in \mathbf{R}^{n\times n}$ is an identity matrix, i.e. $\tilde{A} = AB$.

In this case, if $\operatorname{rank} A = n$, then $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B) = \mathbf{R}^n$ holds, but $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ does not necessarily hold. Hence, from Theorem 3.5, their method may break down before giving a least squares solution. In fact, in [16], Reichel and Ye propose a breakdown-free GMRES method to circumvent this difficulty.

3.2. The BA-GMRES method. The other alternative is to use the same matrix $B \in \mathbf{R}^{n\times m}$ to map the initial residual vector $\boldsymbol{r}_0 \in \mathbf{R}^m$ to $\tilde{\boldsymbol{r}}_0 = B\boldsymbol{r}_0 \in \mathbf{R}^n$, and then to construct the Krylov subspace $\mathcal{K}_i(BA, \tilde{\boldsymbol{r}}_0) = \langle \tilde{\boldsymbol{r}}_0, BA\tilde{\boldsymbol{r}}_0, \ldots, (BA)^{i-1}\tilde{\boldsymbol{r}}_0\rangle$ in $\mathbf{R}^n$ and to solve the least squares problem $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$ using the GMRES method.

First, we will give a theoretical justification for doing so.

THEOREM 3.7.

$$\|\boldsymbol{b} - A\boldsymbol{x}^*\|_2 = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$$

and

$$\|B\boldsymbol{b} - BA\boldsymbol{x}^*\|_2 = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$$

are equivalent for all $\boldsymbol{b} \in \mathbf{R}^m$*, if and only if* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}BA)$.

Proof. Note the following.

$$\|\boldsymbol{b} - A\boldsymbol{x}^*\|_2 = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 \Longleftrightarrow A^{\mathrm{T}}(\boldsymbol{b} - A\boldsymbol{x}^*) = \boldsymbol{0}.$$

$$\begin{aligned} \|B\boldsymbol{b} - BA\boldsymbol{x}^*\|_2 = \min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2 &\iff (BA)^{\mathrm{T}}B(\boldsymbol{b} - A\boldsymbol{x}^*) = \boldsymbol{0} \\ &\iff A^{\mathrm{T}}B^{\mathrm{T}}B(\boldsymbol{b} - A\boldsymbol{x}^*) = \boldsymbol{0}\,. \end{aligned}$$

Then, note that $A^{\mathrm{T}}(\boldsymbol{b}-A\boldsymbol{x}^*) = \boldsymbol{0}$ is equivalent to $A^{\mathrm{T}}B^{\mathrm{T}}B(\boldsymbol{b}-A\boldsymbol{x}^*) = \boldsymbol{0}$ for all $\boldsymbol{b} \in \mathbf{R}^m$, if and only if $\mathcal{N}(A^{\mathrm{T}}) = \mathcal{N}(A^{\mathrm{T}}B^{\mathrm{T}}B)$, which is equivalent to $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}BA)$. ∎

LEMMA 3.8. $\mathcal{R}(A) = \mathcal{R}(B^{T}) \Longrightarrow \mathcal{R}(BA) = \mathcal{R}(B)$.

Proof. $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) \Longrightarrow \mathcal{R}(BA) = \mathcal{R}(BB^{\mathrm{T}}) = \mathcal{R}(B)$ from Lemma 3.2. ∎

LEMMA 3.9. $\mathcal{R}(BA) = \mathcal{R}(B) \Longrightarrow \mathcal{R}(B^{T}BA) = \mathcal{R}(B^{T})$.

Proof. $\mathcal{R}(B^{\mathrm{T}}B) = \mathcal{R}(B^{\mathrm{T}})$. ∎

LEMMA 3.10. $\mathcal{R}(A) = \mathcal{R}(B^{T}) \Longrightarrow \mathcal{R}(A) = \mathcal{R}(B^{T}BA)$.

Proof. From Lemmas 3.8 and 3.9, $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) \Longrightarrow \mathcal{R}(BA) = \mathcal{R}(B)$
$\Longrightarrow \mathcal{R}(B^{\mathrm{T}}BA) = \mathcal{R}(B^{\mathrm{T}}) = \mathcal{R}(A)$. ∎

The following theorem also holds[1][9].

[1]This theorem is due to Professor Masaaki Sugihara.

THEOREM 3.11. *For all* $\boldsymbol{b} \in \mathbf{R}^m$*, the equation* $BA\boldsymbol{x} = B\boldsymbol{b}$ *has a solution, and the solution attains* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$*, if and only if* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$.

Proof. ($\Longrightarrow$) First, for all $\boldsymbol{b} \in \mathbf{R}^m$, $BA\boldsymbol{x} = B\boldsymbol{b}$ has a solution if and only if

$$\mathcal{R}(BA) = \mathcal{R}(B). \tag{3.2}$$

Next, let the equation $BA\boldsymbol{x} = B\boldsymbol{b}$ have a solution, and the solution attain $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$. That is, for all $\boldsymbol{b} \in \mathbf{R}^m$, if $B(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{0}$, then $A^{\mathrm{T}}(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{0}$, if and only if

$$\mathcal{N}(B) \subseteq \mathcal{N}(A^{\mathrm{T}}).$$

This is equivalent to

$$\mathcal{R}(A) \subseteq \mathcal{R}(B^{\mathrm{T}}). \tag{3.3}$$

Noting that

$$\operatorname{rank} BA \leq \min\{\operatorname{rank} A, \operatorname{rank} B\},$$

that is

$$\dim \mathcal{R}(BA) \leq \min\{\dim \mathcal{R}(B), \dim \mathcal{R}(A)\},$$

(3.2) gives $\dim \mathcal{R}(B) \leq \dim \mathcal{R}(A)$. This is equivalent to $\dim \mathcal{R}(B^{\mathrm{T}}) \leq \dim \mathcal{R}(A)$, and together with (3.3) gives $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$.

($\Longleftarrow$)

$$\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) \Longrightarrow \mathcal{R}(BA) = \mathcal{R}(BB^{\mathrm{T}}) = \mathcal{R}(B),$$

and

$$\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) \Longrightarrow \mathcal{R}(A) \subseteq \mathcal{R}(B^{\mathrm{T}}).$$

□

For instance, if $\operatorname{rank} A = \operatorname{rank} B = m$, then $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) = \mathbf{R}^m$ holds, and hence $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}BA)$ holds.

Thus, assume $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}BA)$ holds, and apply the GMRES(k) method to the least squares problem $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$, which gives the following algorithm.

Method 3.2. **The BA-GMRES(k) method**

Choose $\boldsymbol{x}_0$.
$*$ $\tilde{\boldsymbol{r}}_0 = B(\boldsymbol{b} - A\boldsymbol{x}_0)$
$\boldsymbol{v}_1 = \tilde{\boldsymbol{r}}_0 / \|\tilde{\boldsymbol{r}}_0\|_2$
for $i = 1, 2, \ldots, k$
  $\boldsymbol{w}_i = BA\boldsymbol{v}_i$
  for $j = 1, 2, \ldots, i$
    $h_{j,i} = (\boldsymbol{w}_i, \boldsymbol{v}_j)$
    $\boldsymbol{w}_i = \boldsymbol{w}_i - h_{j,i}\boldsymbol{v}_j$
  end for

$h_{i+1,i} = \|\boldsymbol{w}_i\|_2$
$\boldsymbol{v}_{i+1} = \boldsymbol{w}_i / h_{i+1,i}$
Find $\boldsymbol{y}_i \in \mathbf{R}^i$ *which minimizes* $\|\tilde{\boldsymbol{r}}_i\|_2 = \left\| \|\tilde{\boldsymbol{r}}_0\|_2 \, \boldsymbol{e}_i - \bar{H}_i \, \boldsymbol{y} \right\|_2$
$\boldsymbol{x}_i = \boldsymbol{x}_0 + [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i$
$\boldsymbol{r}_i = \boldsymbol{b} - A\boldsymbol{x}_i$
if $\|A^{\mathrm{T}}\boldsymbol{r}_i\|_2 < \epsilon$ *stop*
endfor
$\boldsymbol{x}_0 = \boldsymbol{x}_k$
Go to $*$.

Here, $\tilde{\boldsymbol{r}}_i = B\boldsymbol{r}_i$.

Similarly to Theorem 3.4, the following holds.

THEOREM 3.12. *If* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$*, then* $\mathcal{R}(BA) = \mathcal{R}(A^{\mathrm{T}}B^{\mathrm{T}}) \Longleftrightarrow \mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$ *holds.*

Proof. If $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ holds, Lemma 3.2 gives $\mathcal{R}(BA) = \mathcal{R}(BB^{\mathrm{T}}) = \mathcal{R}(B)$ and $\mathcal{R}(A^{\mathrm{T}}B^{\mathrm{T}}) = \mathcal{R}(A^{\mathrm{T}}A) = \mathcal{R}(A^{\mathrm{T}})$. □

Let BA-GMRES method be the BA-GMRES(k) method with $k = \infty$ (no restarts). Then, Theorems 2.1 and 3.12 give the following.

THEOREM 3.13. *If* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ *holds, then the BA-GMRES method determines a least squares solution of* $\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *for all* $\boldsymbol{b} \in \mathbf{R}^m$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown if and only if* $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$.

As a corollary, we have the following sufficient condition.

COROLLARY 3.14. *If* $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ *and* $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$*, the BA-GMRES method determines a least squares solution of* $\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *for all* $\boldsymbol{b} \in \mathbf{R}^m$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown.*

We note here that Reichel and Ye[16] proposed a related method for solving under-determined ($m \leq n$) least squares problems using the GMRES method. Their method is to append $(n - m)$ zero row vectors to the bottom of the matrix A, to obtain a square singular matrix

$$\tilde{A} = \begin{bmatrix} A \\ 0 \end{bmatrix} \in \mathbf{R}^{n \times n}$$

and let

$$\tilde{\boldsymbol{b}} = \begin{bmatrix} \boldsymbol{b} \\ 0 \end{bmatrix} \in \mathbf{R}^n$$

and apply the GMRES method to $\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\tilde{\boldsymbol{b}} - \tilde{A}\boldsymbol{x}\|_2$.

This corresponds to a special case of our BA-GMRES method with

$$B = \begin{bmatrix} \mathrm{I}_m \\ 0 \end{bmatrix} \in \mathbf{R}^{n \times m}$$

where $\mathrm{I}_m \in \mathbf{R}^{m \times m}$ is an identity matrix, i.e. $\tilde{A} = BA$ and $\tilde{\boldsymbol{b}} = B\boldsymbol{b}$.

Thus, in this case, if rank $A = m$, then $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) = \mathbf{R}^m$ holds, but $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$ does not necessarily hold. Hence, from Theorem 3.13, their method may break down before giving a least squares solution. In fact, in [16], Reichel and Ye propose a breakdown-free GMRES to circumvent this difficulty.

3.3. Summary on condition for B. Summing up the above, for the general case $\operatorname{rank} A \leq \min(m,n)$, including the rank deficient case, the following holds. If the condition:

$$\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}), \quad \mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B) \tag{3.4}$$

is satisfied, then from Corollaries 3.6 and 3.14, the AB-GMRES method and the BA-GMRES method determine a least squares solution of $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ for all $\boldsymbol{b} \in \mathbf{R}^m$ and for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ without breakdown.

The condition (3.4) is satisfied if $B = \alpha A^{\mathrm{T}}$, where $0 \neq \alpha \in \mathbf{R}$.

Next, consider the full rank case: $\operatorname{rank} A = \min(m,n)$.

First, consider the over-determined case: $m \geq n = \operatorname{rank} A$. Let

$$B = CA^{\mathrm{T}}, \tag{3.5}$$

where $C \in \mathbf{R}^{n\times n}$ is an arbitrary nonsingular matrix. Then, the following holds.

$$\begin{array}{c} B = CA^{\mathrm{T}}, \quad C \in \mathbf{R}^{n\times n} : \text{nonsingular} \\ \Updownarrow \\ B^{\mathrm{T}} = AC^{\mathrm{T}}, \quad C^{\mathrm{T}} : \text{nonsingular} \\ \Downarrow \\ \mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}). \end{array}$$

Hence,

$$n = \operatorname{rank} A^{\mathrm{T}} = \operatorname{rank} A = \operatorname{rank} B^{\mathrm{T}} = \operatorname{rank} B$$

gives

$$\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B) = \mathbf{R}^n.$$

Since $AB \in \mathbf{R}^{m\times m}, BA \in \mathbf{R}^{n\times n}$, with $m \geq n$, the amount of computation per iteration is less for the BA-GMRES method compared to the AB-GMRES method This is because the BA-GMRES works in a space with smaller dimension, so that the amount of computation for the modified Gram-Schmidt procedure is less. If $\operatorname{rank} A = n$, $A^{\mathrm{T}}A$ and $\operatorname{diag}(A^{\mathrm{T}}A)$ are nonsingular. Hence, a simple example for C is

$$C := \{\operatorname{diag}(A^{\mathrm{T}}A)\}^{-1}, \tag{3.6}$$

i.e. $B = \{\operatorname{diag}(A^{\mathrm{T}}A)\}^{-1}A^{\mathrm{T}}$, as in [23].

Next, for the full rank under-determined case; $\operatorname{rank} A = m \leq n$, let

$$B = A^{\mathrm{T}}C, \tag{3.7}$$

where $C \in \mathbf{R}^{m\times m}$ is an arbitrary nonsingular matrix. Then, the following holds for $B, A^{\mathrm{T}} \in \mathbf{R}^{n\times m}$.

$$\begin{array}{c} B = A^{\mathrm{T}}C, \quad C \in \mathbf{R}^{m\times m} : \text{nonsingular} \\ \Downarrow \\ \mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B). \end{array}$$

Hence, from

$$m = \operatorname{rank} A = \operatorname{rank} A^{\mathrm{T}} = \operatorname{rank} B = \operatorname{rank} B^{\mathrm{T}},$$

$$\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}}) = \mathbf{R}^m$$

also holds.

Note here that, since $AB \in \mathbf{R}^{m\times m}$, $BA \in \mathbf{R}^{n\times n}$, $m \leq n$, the amount of computation per iteration is less for the AB-GMRES method compared to the BA-GMRES method. If $\operatorname{rank} A = m$, AA^{T} and $\operatorname{diag}(AA^{\mathrm{T}})$ are nonsingular. Hence, a simple example for C is

$$C := \{\operatorname{diag}(AA^{\mathrm{T}})\}^{-1}, \tag{3.8}$$

i.e. $B = A^{\mathrm{T}}\{\operatorname{diag}(AA^{\mathrm{T}})\}^{-1}$.

Note also that when $\operatorname{rank} A = m, \mathcal{R}(A) = \mathbf{R}^m \ni \boldsymbol{b}$, so that

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 = \min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2 = \min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AA^{\mathrm{T}}C\boldsymbol{z}\|_2 = 0.$$

Hence, the AB-GMRES method with $B = A^{\mathrm{T}}C$ gives the minimum norm least squares solution $\boldsymbol{x}^* = B\boldsymbol{z}^* = A^{\mathrm{T}}(C\boldsymbol{z}^*)$ of (1.1), since $AA^{\mathrm{T}}(C\boldsymbol{z}^*) = \boldsymbol{b}$.

4. Convergence analysis. Next, we will analyze the convergence of the AB-GMRES method and the BA-GMRES method[2].

4.1. Over-determined case. First, we will consider the over-determined case $m \geq n$ with $B := CA^{\mathrm{T}}$ as in (3.5), where $C \in \mathbf{R}^{n\times n}$ is restricted to be symmetric and positive-definite.

THEOREM 4.1. *Let $A \in \mathbf{R}^{m\times n}$, $m \geq n$ and $B := CA^{\mathrm{T}}$ where $C \in \mathbf{R}^{n\times n}$ is symmetric and positive-definite. Let the singular values of $\tilde{A} := AC^{\frac{1}{2}}$ be $\sigma_i (1 \leq i \leq n)$. Then, ${\sigma_i}^2$ $(1 \leq i \leq n)$ are eigenvalues of AB and BA. If $m > n$, all the other eigenvalues of AB are* 0.

Proof. Let $\tilde{A} := AC^{\frac{1}{2}} = U\Sigma V^{\mathrm{T}}$ be the singular decomposition of $\tilde{A}$. Here, $U \in \mathbf{R}^{m\times m}, V \in \mathbf{R}^{n\times n}$ are orthogonal matrices, and

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \sigma_n \\ 0 & & \end{bmatrix} \in \mathbf{R}^{m\times n},$$

where $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$ are the singular values of $\tilde{A}$. Then,

$$AB = ACA^{\mathrm{T}} = \tilde{A}\,\tilde{A}^{\mathrm{T}} = U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}}, \tag{4.1}$$

$$BA = CA^{\mathrm{T}}A = C^{\frac{1}{2}}\tilde{A}^{\mathrm{T}}\tilde{A}C^{-\frac{1}{2}} = C^{\frac{1}{2}}V\Sigma^{\mathrm{T}}\Sigma(C^{\frac{1}{2}}V)^{-1}. \tag{4.2}$$

□

Note that AB is symmetric. It is also positive definite if $\operatorname{rank} A = n$.

If $\operatorname{rank} A = n$, then $C := \{\operatorname{diag}(A^{\mathrm{T}}A)\}^{-1}$ in the example in (3.6) is symmetric positive-definite. Also the $C := (LDL^{\mathrm{T}})^{-1}$ for the RIF preconditioner introduced in Section 5 is also symmetric positive-definite.

[2]We would like to thank Professor Michael Eiermann for discussions which lead to the following analysis.

Now note the following theorem. (See, for instance, [18].)

THEOREM 4.2. *Assume $A' \in \mathbf{R}^{n\times n}$ is diagonalizable and let $A' = X\Lambda X^{-1}$ where $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$.*

Then, the residual norm achieved by the k-th step of the GMRES applied to $A'\boldsymbol{x}' = \boldsymbol{b}'$ satisfies

$$\|\boldsymbol{r}'_m\|_2 \leq \kappa_2(X)\epsilon^{(k)}\|\boldsymbol{r}'_0\|_2$$

where $\kappa_2(X) = \|X\|_2\|X^{-1}\|_2$ and

$$\epsilon^{(k)} = \min_{p_k\in Q_k} \max_{i=1,\ldots,n} |p_k(\lambda_i)| \leq 2\left[\frac{\sqrt{\kappa(A')}-1}{\sqrt{\kappa(A')}+1}\right]^k$$

where $Q_k := \{p_k \mid p_k(x) : \text{polynomial of } x \text{ with degree} \leq k,\ p_k(0) = 1\}$ and $\kappa(A') = \dfrac{\lambda_1}{\lambda_n}$.

First, consider the AB-GMRES method. Let $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$. Note that $AB\boldsymbol{u}_i = {\sigma_i}^2\boldsymbol{u}_i \quad (i = 1, \ldots, n)$, $AB\boldsymbol{u}_i = \boldsymbol{0} \quad (i = n+1, \ldots, m)$, so that $\mathcal{R}(AB) = \mathcal{R}(A) = \mathrm{span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ and $\mathcal{N}(AB) = \mathcal{R}(AB)^\perp = \mathcal{R}(A)^\perp = \mathrm{span}\{\boldsymbol{u}_{n+1}, \ldots, \boldsymbol{u}_m\}$. Let

$$\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0 = \sum_{i=1}^{m} \rho_i\boldsymbol{u}_i \ . \tag{4.3}$$

The k-th residual vector $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k$ satisfies

$$\|\boldsymbol{r}_k\|_2 = \min_{\boldsymbol{\zeta}\in\mathcal{K}_k(AB,\boldsymbol{r}_0)} \|\boldsymbol{b} - AB(\boldsymbol{z}_0 + \boldsymbol{\zeta})\|_2 = \min_{p_k\in Q_k} \|p_k(AB)\boldsymbol{r}_0\|_2 \ .$$

Note from (4.1) that

$$p_k(AB) = Up_k(\Sigma\Sigma^{\mathrm{T}})U^{\mathrm{T}} = \sum_{i=1}^{n} p_k({\sigma_i}^2)\boldsymbol{u}_i{\boldsymbol{u}_i}^{\mathrm{T}} + p_k(0)\sum_{i=n+1}^{m} \boldsymbol{u}_i{\boldsymbol{u}_i}^{\mathrm{T}}.$$

Hence, (4.3) gives

$$p_k(AB)\boldsymbol{r}_0 = \sum_{i=1}^{n} \rho_i\, p_k({\sigma_i}^2)\boldsymbol{u}_i + p_k(0)\sum_{i=n+1}^{m} \rho_i\boldsymbol{u}_i.$$

Hence, the norm of the $\mathcal{R}(A)$-component of the residual is given by

$$\|\boldsymbol{r}_k|_{\mathcal{R}(A)}\|_2{}^2 = \min_{p_k\in Q_k} \sum_{i=1}^{n} {\rho_i}^2\{p_k({\sigma_i}^2)\}^2$$

where

$$\|\boldsymbol{r}_0|_{\mathcal{R}(A)}\|_2{}^2 = \sum_{i=1}^{n} {\rho_i}^2 \ .$$

Thus,

$$\begin{aligned}\frac{\|\boldsymbol{r}_k|_{\mathcal{R}(A)}\|_2}{\|\boldsymbol{r}_0|_{\mathcal{R}(A)}\|_2} &\leq \min_{p_k\in Q_k}\max_{1\leq i\leq n}|p_k({\sigma_i}^2)| \\ &\leq 2\left[\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right]^k \\ &= 2\left(\frac{\sigma_1-\sigma_n}{\sigma_1+\sigma_n}\right)^k\end{aligned}$$

where $\kappa = \left(\frac{\sigma_1}{\sigma_n}\right)^2$. Here we have assumed $\sigma_n > 0$, i.e. $\operatorname{rank} A = n$.

Thus, we have the following under the same assumptions and notations as in Theorem 4.1.

THEOREM 4.3. *The residual $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ achieved by the k-th step of AB-GMRES satisfies*

$$\|\boldsymbol{r}_k|_{\mathcal{R}(A)}\|_2 \leq 2\left(\frac{\sigma_1-\sigma_n}{\sigma_1+\sigma_n}\right)^k \|\boldsymbol{r}_0|_{\mathcal{R}(A)}\|_2.$$

Next, consider the BA-GMRES method. In Theorem 4.2, let $A' = BA, \boldsymbol{b}' = B\boldsymbol{b}$, and $\boldsymbol{x}' = \boldsymbol{x}$. Then, we have $\boldsymbol{r}' = B\boldsymbol{r} = CA^{\mathrm{T}}\boldsymbol{r}$, $\lambda_i = {\sigma_i}^2$ $(i = 1, \ldots, n), X = C^{\frac{1}{2}}V, \kappa_2(X) = \sqrt{\kappa(C)}$ where $\kappa(C) = \dfrac{\lambda_{\max}(C)}{\lambda_{\min}(C)}, \kappa(A') = \kappa(BA) = \left(\dfrac{\sigma_1}{\sigma_n}\right)^2$, $\dfrac{\sqrt{\kappa(A')}-1}{\sqrt{\kappa(A')}+1} = \dfrac{\sigma_1-\sigma_n}{\sigma_1+\sigma_n}$.

Thus, we have the following under the same assumptions and notations as in Theorem 4.1.

THEOREM 4.4. *The residual $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ achieved by the k-th step of the BA-GMRES method satisfies*

$$\|B\boldsymbol{r}_k\|_2 = \|CA^{T}\boldsymbol{r}_k\| \leq 2\sqrt{\kappa(C)}\left(\frac{\sigma_1-\sigma_n}{\sigma_1+\sigma_n}\right)^k \|B\boldsymbol{r}_0\|_2.$$

Next, we turn to the CGLS method. Consider a symmetric positive-definite preconditioner matrix $C = (\tilde{L}\tilde{L}^{\mathrm{T}})^{-1} \in \mathbf{R}^{n\times n}$. Here, C may be based on an incomplete factorization $LDL^{\mathrm{T}} \sim A^{\mathrm{T}}A$, $\tilde{L} = LD^{\frac{1}{2}}$ as in the robust incomplete factorization (RIF). For the diagonal scaling, we have $\operatorname{diag}(A^{\mathrm{T}}A) \sim A^{\mathrm{T}}A$, $\tilde{L} = (\operatorname{diag}(A^{\mathrm{T}}A))^{\frac{1}{2}}$.

Then, the natural way to precondition the CGLS(CGNR) method of (1.1) or the mathematically equivalent LSQR method, is to apply the CG method to

$$A'\boldsymbol{x}' = \boldsymbol{b}' \tag{4.4}$$

where $A' = \tilde{L}^{-1}A^{\mathrm{T}}A\tilde{L}^{-\mathrm{T}}$ is symmetric positive definite, $\boldsymbol{x}' = \tilde{L}^{\mathrm{T}}\boldsymbol{x}$ and $\boldsymbol{b} = \tilde{L}^{-1}A^{\mathrm{T}}\boldsymbol{b}$.

Note the following theorem. (See e.g. [18].)

THEOREM 4.5. *Let $\boldsymbol{x}_k$ be the k-th iterate of the CG method applied to $A'\boldsymbol{x}' = \boldsymbol{b}'$. Then,*

$$\|\boldsymbol{e}'_k\|_{A'} \leq 2\left\{\frac{\sqrt{\kappa(A')}-1}{\sqrt{\kappa(A')}+1}\right\}^k \|\boldsymbol{e}'_0\|_{A'}$$

where $\boldsymbol{e}'_k = \boldsymbol{x}'_* - \boldsymbol{x}'_k$, $\boldsymbol{x}'_* = {A'}^{-1}\boldsymbol{b}'$ *and* $\kappa(A') = \dfrac{\lambda_{\max}(A')}{\lambda_{\min}(A')}$.

Since $(\tilde{L}^{\mathrm{T}})^{-1}A'(\tilde{L}^{\mathrm{T}}) = CA^{\mathrm{T}}A = BA$, $\lambda_i(A') = \lambda_i(BA) = {\sigma_i}^2$ $(i = 1, \ldots, n)$. Hence, $\kappa(A') = \left(\dfrac{\sigma_1}{\sigma_n}\right)^2$, and $\dfrac{\sqrt{\kappa(A')}-1}{\sqrt{\kappa(A')}+1} = \dfrac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n}$.

Note also that

$$\|\boldsymbol{e}'_k\|_{A'} = \|\boldsymbol{e}_k\|_{A^{\mathrm{T}}A} = \|A^{\mathrm{T}}\boldsymbol{r}_k\|_{(A^{\mathrm{T}}A)^{-1}},$$

where $\boldsymbol{e}_k = \tilde{L}^{-\mathrm{T}}\boldsymbol{e}'_k$, $\boldsymbol{e}_k = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{r}_k$, $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k$.

Hence, we have the following under the same assumptions and notations as in Theorem 4.1.

THEOREM 4.6. *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *of the k-th step of the preconditioned CGLS (CG applied to (4.4)) satisfies*

$$\|A^{T}\boldsymbol{r}_k\|_{(A^{T}A)^{-1}} = \|\boldsymbol{e}_k\|_{A^{T}A} \leq 2\left(\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n}\right)^k \|A^{T}\boldsymbol{r}_0\|_{(A^{T}A)^{-1}}.$$

From the above analysis, we may expect that the AB-GMRES, BA-GMRES and the preconditioned CGLS methods exhibit similar convergence behaviours for the over-determined case.

4.2. Under-determined case. Similarly, consider the under-determined case $\operatorname{rank} A = m \leq n$ with $B = A^{\mathrm{T}}C$ as in (3.7), where $C \in \mathbf{R}^{m\times m}$ is restricted to be symmetric and positive-definite. The following holds.

THEOREM 4.7.
Let $A \in \mathbf{R}^{m\times n}$, $m \leq n$, *where* $B := A^{T}C$ *and* $C \in \mathbf{R}^{m\times m}$ *is symmetric and positive-definite. Let the singular values of* $\tilde{A} := C^{\frac{1}{2}}A$ *be* $\sigma_i (1 \leq i \leq m)$. *Then,* ${\sigma_i}^2 (1 \leq i \leq m)$ *are the eigenvalues of* AB *and* BA. *If* $m < n$, *all the other eigenvalues of* BA *are* 0.

Proof. Let $\tilde{A} := C^{\frac{1}{2}}A = U\Sigma V^{\mathrm{T}}$ be the singular value decomposition of $\tilde{A}$. Here, $U \in \mathbf{R}^{m\times m}, V \in \mathbf{R}^{n\times n}$ are orthogonal matrices, and

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_m & \end{bmatrix} \in \mathbf{R}^{m\times n},$$

where $\sigma_1 \geq \ldots \geq \sigma_m \geq 0$ are the singular values of $\tilde{A}$.

Then,

$$AB = AA^{\mathrm{T}}C = C^{-\frac{1}{2}}\tilde{A}\,\tilde{A}^{\mathrm{T}}C^{\frac{1}{2}} = C^{-\frac{1}{2}}U\Sigma\Sigma^{\mathrm{T}}(C^{-\frac{1}{2}}U)^{-1},$$

$$BA = A^{\mathrm{T}}CA = \tilde{A}^{\mathrm{T}}\tilde{A} = V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}.$$

□

First, consider the AB-GMRES method. In Theorem 4.2, let $A' = AB$, $\boldsymbol{b}' = \boldsymbol{b}$, $\boldsymbol{x}' = \boldsymbol{z}$. Then we have $\boldsymbol{r}' = \boldsymbol{b}' - A'\boldsymbol{x}' = \boldsymbol{b} - AB\boldsymbol{z} = \boldsymbol{r}$, $\lambda_i = {\sigma_i}^2$ $(i = 1, \ldots, m)$, $X = C^{-\frac{1}{2}}U$, $\kappa_2(X) = \sqrt{\kappa(C)}$ where $\kappa(C) = \dfrac{\lambda_{\max}(C)}{\lambda_{\min}(C)}$, $\kappa(A') = \kappa(AB) = \left(\dfrac{\sigma_1}{\sigma_m}\right)^2$.

Thus, we have the following theorem under the same assumptions and notations as in Theorem 4.7.

THEOREM 4.8. *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *achieved by the k-th step of the AB-GMRES method satisfies*

$$\|\boldsymbol{r}_k\|_2 \leq 2\sqrt{\kappa(C)}\left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \|\boldsymbol{r}_0\|_2.$$

Next, consider the BA-GMRES method. Note that $BA = A^{\mathrm{T}}CA$ is symmetric. Note $BA\boldsymbol{v}_i = {\sigma_i}^2\boldsymbol{v}_i$ $(i = 1, \ldots, m)$ and $BA\boldsymbol{v}_i = \boldsymbol{0}$ $(i = m+1, \ldots, n)$, so that $\mathcal{R}(BA) = \mathcal{R}(B) = \mathcal{R}(A^{\mathrm{T}}) = \mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ and $\mathcal{N}(BA) = \mathcal{R}(BA)^{\perp} = \mathcal{R}(B)^{\perp} = \mathrm{span}\{\boldsymbol{v}_{m+1}, \ldots, \boldsymbol{v}_n\}$.

Let $B\boldsymbol{r}_0 = \sum_{i=1}^{n} \rho_i \boldsymbol{v}_i \in \mathbf{R}^m$. Then,

$$p_k(BA)B\boldsymbol{r}_0 = \sum_{i=1}^{m} \rho_i p_k({\sigma_i}^2)\boldsymbol{v}_i + p_k(0)\sum_{i=m+1}^{n} \rho_i \boldsymbol{v}_i.$$

The k-th iterate of the BA-GMRES method satisfies

$$\|B\boldsymbol{r}_k\|_2 = \min_{\boldsymbol{\xi} \in \mathcal{K}_k(BA, B\boldsymbol{r}_0)} \|B\boldsymbol{v} - BA(\boldsymbol{x}_0 + \boldsymbol{\xi})\|_2 = \min_{p_k \in Q_k} \|p_k(BA)B\boldsymbol{r}_0\|_2.$$

Hence,

$$\|B\boldsymbol{r}_k|_{\mathcal{R}(B)}\|_2{}^2 = \min_{p_k \in Q_k} \sum_{i=1}^{m} {\rho_i}^2 p_k({\sigma_i}^2)^2,$$

$$\|B\boldsymbol{r}_0|_{\mathcal{R}(B)}\|_2{}^2 = \sum_{i=1}^{n} {\rho_i}^2,$$

so that

$$\begin{aligned} \frac{\|B\boldsymbol{r}_k|_{\mathcal{R}(B)}\|_2}{\|B\boldsymbol{r}_0|_{\mathcal{R}(B)}\|_2} &\leq \min_{p_k \in Q_k} \max_{1 \leq i \leq m} |p_k({\sigma_i}^2)| \\ &\leq 2\left[\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right]^k \\ &= 2\left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \end{aligned} \tag{4.5}$$

where $\kappa = \left(\dfrac{\sigma_1}{\sigma_m}\right)^2$, where we assume $\sigma_m > 0$ ($\mathrm{rank}\, A = m$).

Thus, we have the following theorem under the same assumptions and notations as in Theorem 4.7.

THEOREM 4.9. *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *achieved by the k-th step of BA-GMRES satisfies*

$$\|B\boldsymbol{r}_k|_{\mathcal{R}(B)}\|_2 \leq 2\left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \|B\boldsymbol{r}_0|_{\mathcal{R}(B)}\|_2.$$

Next, we turn to the CGLS(CGNE) method.

Consider a symmetric positive definite preconditioning matrix $C = (\tilde{L}\tilde{L}^{\mathrm{T}})^{-1} \in \mathbf{R}^{m\times m}$, where $\tilde{L} = LD^{\frac{1}{2}}$, $\tilde{L}\tilde{L}^{\mathrm{T}} = LDL^{\mathrm{T}} \sim AA^{\mathrm{T}}$ as in the RIF method, or $\tilde{L} = (\mathrm{diag}(AA^{\mathrm{T}}))^{\frac{1}{2}}$ in the diagonal scaling.

Then, one can precondition $AA^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{b}$ as

$$A'\boldsymbol{y}' = \boldsymbol{b}' \tag{4.6}$$

where $A' = \tilde{L}^{-1}AA^{\mathrm{T}}\tilde{L}^{-\mathrm{T}}$, $\boldsymbol{y}' = \tilde{L}^{\mathrm{T}}\boldsymbol{y}$, $\boldsymbol{b}' = \tilde{L}^{-1}\boldsymbol{b}$, $\boldsymbol{r}' = \tilde{L}^{-1}\boldsymbol{r}$.

Since $\tilde{L}A'\tilde{L}^{-1} = AA^{\mathrm{T}}C = AB$, $\lambda_i(A') = \lambda_i(AB) = {\sigma_i}^2$ $(i = 1, \ldots, m)$.

From Theorem 4.5, we have the following theorem under the same assumptions and notations as in Theorem 4.7.

THEOREM 4.10. *The residual $\boldsymbol{r} = \boldsymbol{b} - AA^{T}\boldsymbol{y}$ of the k-th step of the preconditioned CGLS(CGNE) method (CG applied to (4.6)) satisfies*

$$\|\boldsymbol{r}_k\|_{(AA^{T})^{-1}} = \|\boldsymbol{e}_k\|_{AA^{T}} \leq 2\left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \|\boldsymbol{r}_0\|_{(AA^{T})^{-1}}.$$

From the above analyses, we may expect that the AB-GMRES, BA-GMRES and the preconditioned CGLS(CGNE) methods exhibit similar convergence behaviours for the under-determined case also.

5. The choice of B. Besides satisfying the conditions $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ and $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$, it is desirable that B satisfies $AB \approx \mathrm{I}_m$ or $BA \approx \mathrm{I}_n$, in order to speed up the convergence.

Simple candidates, as mentioned before, are to take $C := \{\mathrm{diag}(A^{\mathrm{T}}A)\}^{-1}$ and $B = CA^{\mathrm{T}}$ when $m \geq n = \mathrm{rank}\, A$, and to take $C := \{\mathrm{diag}(AA^{\mathrm{T}})\}^{-1}$ and $B = A^{\mathrm{T}}C$ when $\mathrm{rank}\, A = m \leq n$.

More sophisticated preconditioners based on the incomplete QR decompositions of A may be considered. That is, $A = QR+E$, where $Q \in \mathbf{R}^{m\times n}$ is an (approximately) orthogonal matrix, $R \in \mathbf{R}^{n\times n}$ is an upper triangular matrix, and E is the error matrix.

Usually, the matrix R is used as a preconditioner for the CGLS method, as mentioned in (1.4). Similarly, we may let $B = R^{-1}Q^{\mathrm{T}}$ for the AB-GMRES(k) and BA-GMRES(k) methods, for instance when $m \geq n$, and in the case of the BA-GMRES(k) method, we may apply the GMRES(k) method to $R^{-1}Q^{\mathrm{T}}A\boldsymbol{x} = R^{-1}Q^{\mathrm{T}}\boldsymbol{b}$.

There are many approaches to construct the incomplete QR decomposition of the matrix A, e.g., the incomplete modified Gram-Schmidt method [12, 17, 22], the incomplete Householder reflection, and the incomplete Givens rotation [1, 15]. In [11, 8, 9], the IMGS(l) method, an approximation of the modified Gram-Schmidt method, was applied to the CGLS, AB-GMRES(k) and BA-GMRES(k) methods.

The robust incomplete factorization (RIF) is an attractive method with low memory requirements[2], and was also applied to the CGLS method for large sparse least squares problems[3]). The method gives an upper triangular matrix Z, a lower triangular matrix L, and a diagonal matrix $D = \mathrm{diag}(\boldsymbol{d})$ such that

$$A^{\mathrm{T}}A \approx LDL^{\mathrm{T}},$$

and

$$A^{\mathrm{T}}A \approx Z^{-\mathrm{T}}DZ^{-1},$$

where $A \in \mathbf{R}^{m\times n}$ with $m \geq n = \mathrm{rank}\, A$.

The method is based on a $A^{\mathrm{T}}A$-orthogonalization procedure, and will never breakdown. Moreover, by using a drop tolerance τ, the number of fill-ins of the matrices Z and L can be controlled. For $1 \leq j \leq n$, let $\mathbf{z}_j$, $\mathbf{e}_j$ and d_j denote the jth column vector of the matrix Z, the jth unit basis vector and the jth entry of the vector $\boldsymbol{d}$, respectively. The method is as follows.

Method 5.1. **The Robust Incomplete Factorization (RIF) Method**
Let $Z = L = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n]$
For $j = 1, \ldots, n$ Do
  Compute $\mathbf{u}_j = A\mathbf{z}_j$
  Compute $d_j = (\mathbf{u}_j, \mathbf{u}_j)$
  For $i = j+1, \ldots, n$ Do:
    Compute $\mathbf{v}_i = A\mathbf{e}_i$
    Compute $\theta_{ij} = \frac{(\mathbf{v}_i, \mathbf{u}_j)}{d_j}$
    IF $\theta_{ij} > \tau$
      Store $L(i,j) = \theta_{ij}$
    EndIF
    Compute $\mathbf{z}_i = \mathbf{z}_i - \theta_{ij}\mathbf{z}_j$
    Drop the elements in $\mathbf{z}_i$ which are smaller than τ
  EndDo
EndDo

If $\tau = 0$ in the above method, the complete factorization of the matrix $A^{\mathrm{T}}A$ is obtained, so that $A^{\mathrm{T}}A = LDL^{\mathrm{T}} = Z^{-\mathrm{T}}DZ^{-1}$ where $Z^{-1} = L^{\mathrm{T}}$. If we let $\widetilde{L} = LD^{1/2}$, we have $A^{\mathrm{T}}A = \widetilde{L}\widetilde{L}^{\mathrm{T}}$, which means that the RIF method is equivalent to the Cholesky factorization for the coefficient matrix of the normal equation.

The drop tolerance parameter τ plays an important role in the RIF method, since it determines not only how the sparse matrices L and Z approximate the corresponding complete factorization matrices, but also the amount of computation and storage required. The relative drop tolerance can also be used by replacing τ with $\tau\|\boldsymbol{a}_i\|_2$, where $\boldsymbol{a}_i$ is the i-th column of A [3]. In the numerical experiments in Section 6, we will use the relative drop tolerance.

The matrix LDL^{T} is guaranteed to be positive definite when A is full column rank, since $d_j > 0$ for all $1 \leq j \leq n$, and the matrix L is a sparse lower triangular matrix with all the diagonal elements being one.

Therefore, let $C := \{LDL^{\mathrm{T}}\}^{-1}$ and $B = CA^{\mathrm{T}}$ when $m \geq n = \operatorname{rank} A$. Since the matrix C is nonsingular, B satisfies the conditions $\mathcal{R}(A) = \mathcal{R}(B^{\mathrm{T}})$ and $\mathcal{R}(A^{\mathrm{T}}) = \mathcal{R}(B)$ (cf. Section 3.3 and Theorem 3.13). The approximation $BA \approx I_n$ improves as τ approaches 0.

Alternatively, let $C := ZD^{-1}Z^{\mathrm{T}}$ and $B = CA^{\mathrm{T}}$. Then, C and B satisfy the above conditions, respectively.

In the following numerical experiments, we used the latter: $C := ZD^{-1}Z^{\mathrm{T}}$.

For the under-determined case where $\operatorname{rank} A = m \leq n$, we can also construct the matrices $\widehat{Z}$, $\widehat{L}$ and $\widehat{D}$ using the RIF method, where $AA^{\mathrm{T}} \approx \widehat{L}\widehat{D}\widehat{L}^{\mathrm{T}}$ and $AA^{\mathrm{T}} \approx \widehat{Z}^{-\mathrm{T}}\widehat{D}\widehat{Z}^{-1}$. Then, let $B = A^{\mathrm{T}}C$ where $C := \{\widehat{L}\widehat{D}\widehat{L}^{\mathrm{T}}\}^{-1}$ or $C := \widehat{Z}\widehat{D}^{-1}\widehat{Z}^{\mathrm{T}}$. B satisfies the conditions in Theorem 3.5 and $AB \approx I_m$.

6. Numerical Experiments. Finally, we present numerical experiment results to show the performance of the AB-GMRES and BA-GMRES methods. We compare them to the preconditioned CGLS [4] (also called CGNR in [18]) and LSQR [14]

method for over-determined problems, and also compare them to the preconditioned CGNE [18] and LSQR method for under-determined problems.

In the experiments, we first generated the test matrices by the MATLAB routine "sprandn", so that we could specify the density (the ratio of non-zero elements) and the condition number of the matrices. The value of the non-zero elements are generated by a random number generator following the normal distribution and the pattern of the non-zero elements are also determined by a random number generator.

All the computations were run on the Dell Precision 690 where the CPU is 3.00 GHz and the memory is 16 GB, and the programming language and compiling environment was GNU C/C++ 3.4.3.

6.1. Over-determined case. First, consider the over-determined least squares problem

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2, \qquad A \in \mathbf{R}^{m\times n} \qquad (m \geq n).$$

We test the following matrices with $m = 30,000, n = 3,000$ and density 0.1%. The condition number (denoted by $\kappa(A)$) of these matrices are given in Table 6.1.

TABLE 6.1
The condition number of the test matrices.

Name	$\kappa(A)$
RANDL1	1.9×10
RANDL2	1.6×10^2
RANDL3	1.3×10^3
RANDL4	2.0×10^4
RANDL5	1.3×10^5
RANDL6	1.3×10^6
RANDL7	1.3×10^7

We set the initial approximate solution to $\boldsymbol{x}_0 = \mathbf{0}$, and the convergence criterion was $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2 < 10^{-6}$, where $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ is the residual. For the right hand vector $\boldsymbol{b}$, each of its components were generated by a random number generator following the normal distribution, such that $\boldsymbol{b} \in \mathbf{R}^m$ but $\boldsymbol{b} \notin \mathcal{R}(A)$.

In the experiments, we compare the CGLS (based on Method 1.1), LSQR, reorthogonalized CGLS, and the BA-GMRES methods with the diagonal scaling preconditioner and the RIF preconditioner. The methods are denoted by CGLS-diag., LSQR-diag., RCGLS-diag., BA-GMRES-diag. and CGLS-RIF, LSQR-RIF, RCGLS-RIF, BA-GMRES-RIF, respectively. We let $B = CA^{\mathrm{T}}$ for the BA-GMRES methods where $C = \{\mathrm{diag}(A^{\mathrm{T}}A)\}^{-1}$ for the diagonal scaling and $C = ZD^{-1}Z^{\mathrm{T}}$ for the RIF. In the reorthogonalized CGLS, the (preconditioned) residual vectors were reorthogonalized with respect to all the previous (preconditioned) residual vectors.

We compared with the reorthogonalized CGLS, since it was observed that the convergence of the CGLS and LSQR methods deteriorate as the problem becomes ill-conditioned.

For the BA-GMRES method, $\|B\boldsymbol{r}\|_2 = \|CA^{\mathrm{T}}\boldsymbol{r}\|_2$ is readily available in the GMRES process, so it is practical to use it to judge convergence. However, in the experiments below, we have compared all the methods by $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2$, and the extra time to compute $\|A^{\mathrm{T}}\boldsymbol{r}\|_2$ was neglected in the CPU times for the BA-GMRES method.

First, Figure 6.1 shows $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2$ vs. the number of iterations for the full AB-GMRES and BA-GMRES methods with diagonal scaling and the RIF preconditioning for the problem RANDL3. Here the relative drop tolerance for the RIF was set to the optimal value $\tau = 0.8$.

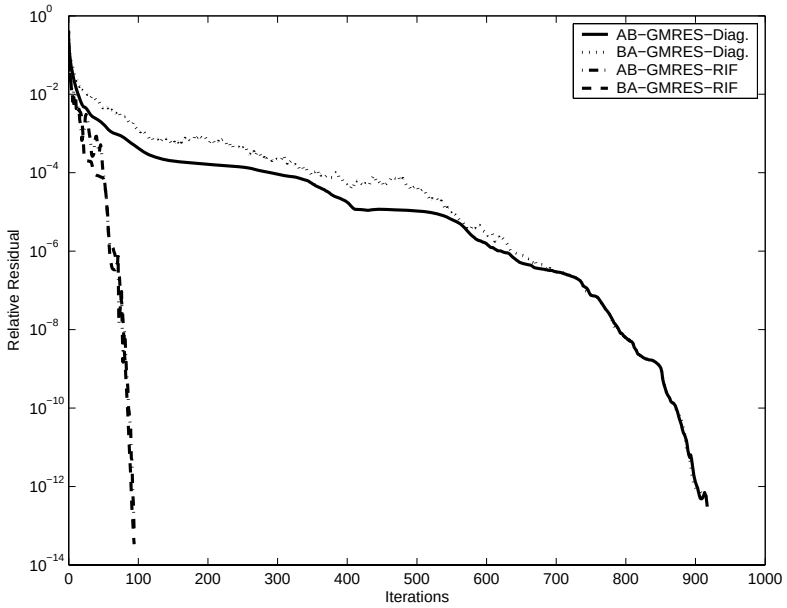

FIG. 6.1. *Comparison of AB-GMRES method and BA-GMRES method (RANDL3, $\tau = 0.8$ for RIF).*

The full AB-GMRES and BA-GMRES methods show similar convergence behaviours with both the diagonal scaling and the RIF preconditioner. This is in accordance with the convergence analysis in Section 4.1.

However, as mentioned in Section 3.3, the amount of computation per iteration for the BA-GMRES is less than the AB-GMRES method when $m \geq n$. Therefore, we compare the BA-GMRES method with the preconditioned CGLS, LSQR and reorthogonalized CGLS methods in the following experiments.

Next, we show the effect of changing the restart period k for the BA-GMRES(k) method with the RIF preconditioner for the problems RANDL5, RANDL6 and RANDL7 with optimal τ for each in Table 6.2.

In Table 6.2 and the tables below, $*$ indicates the fastest for each problem. In these examples, both the number of iterations and the computation time were minimum for the full GMRES method without restarts, though they require more storage to store the orthogonal vectors. Hence, in the experiments below, we adopt the full BA-GMRES method.

Next, we focus on the choice of the optimal relative drop tolerance parameter τ in the RIF preconditioner. Table 6.3 and Table 6.4 give the number of iterations and total CPU time (the sum of the preconditioning time and iteration time) for each method with different τ, for the problems RANDL3 and RANDL6, respectively. $\dagger$ indicates the fastest for each method. For RANDL3, $\tau = 0.8$ was optimal, whereas for RANDL6, $\tau = 0.07$ was optimal. The optimal τ was the same for the four methods. The optimal parameter τ for the other problems are given in Table 6.5. In

TABLE 6.2
The effect of the restart period k in the BA-GMRES(k)-RIF method (over-determined problem).

RANDL5	k	100	140	180	220	260	≥ 295
($\tau = 0.8$)	iter	755	668	587	498	453	295
	time	25.05	24.81	24.14	22.62	22.15	$*$14.27
RANDL6	k	160	200	240	280	300	≥ 318
($\tau = 0.07$)	iter	2,470	2,085	1,856	1,311	599	318
	time	144.68	130.08	121.72	92.88	48.92	$*$26.73
RANDL7	k	200	280	320	340	360	≥ 362
($\tau = 0.02$)	iter	2,594	2,168	1,567	900	506	362
	time	179.83	165.79	129.07	80.13	50.99	$*$37.26

k: restart period, iter: number of iterations, time: computation time (sec.).
Convergence criterion: $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2 < 10^{-6}$.

the experiments below, we will use these optimal values for τ for each problem.

TABLE 6.3
The effect of the RIF parameter τ for problem RANDL3.

	τ	0.9	0.8	0.7	0.6	0.5	0
CGLS-RIF	iter	83	72	77	65	67	1
	time	5.39	$*$5.38	5.42	5.43	5.59	68.00
LSQR-RIF	iter	83	73	78	66	68	1
	time	5.39	$*$5.38	5.42	5.43	5.60	68.00
RCGLS-RIF	iter	79	70	73	64	66	1
	time	5.68	†5.62	5.69	5.63	5.76	67.99
BA-GMRES-RIF	iter	70	60	73	64	66	1
	time	5.54	†5.47	5.64	5.59	5.63	67.99

iter: number of iterations, time: computation time (sec.).
Convergence criterion: $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2 < 10^{-6}$.

TABLE 6.4
The effect of the RIF parameter τ for problem RANDL6.

	τ	0.09	0.08	0.07	0.06	0.05	0
CGLS-RIF	iter	670	674	615	706	658	1
	time	34.38	35.39	†33.67	38.17	37.31	60.00
LSQR-RIF	iter	680	685	645	736	701	1
	time	34.87	35.50	†34.15	37.21	37.41	60.00
RCGLS-RIF	iter	333	324	317	313	307	1
	time	27.19	27.00	†26.93	27.26	27.59	60.01
BA-GMRES-RIF	iter	335	325	318	315	309	1
	time	27.03	26.77	$*$26.73	27.16	27.45	60.01

iter: number of iterations, time: computation time (sec.).
Convergence criterion: $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2 < 10^{-6}$.

When $\tau = 0$, the $A^{\mathrm{T}}A$ orthogonal factorization is complete and the iterative methods act as direct methods and converges in only one iteration. Table 6.3 and Table 6.4 show that the preconditioned Krylov subspace iterative methods are faster

than this "direct method".

Figures 6.2, 6.3, 6.4 and 6.5 show $\|A^{\mathrm{T}}\boldsymbol{r}\|_2/\|A^{\mathrm{T}}\boldsymbol{b}\|_2$ vs. the number of iterations for the CGLS, LSQR, reorthogonalized CGLS and BA-GMRES methods, with diagonal scaling and RIF preconditioners, for the problems RANDL5 and RANDL6. τ for RIF was set to the optimal value 0.8 and 0.07, respectively.

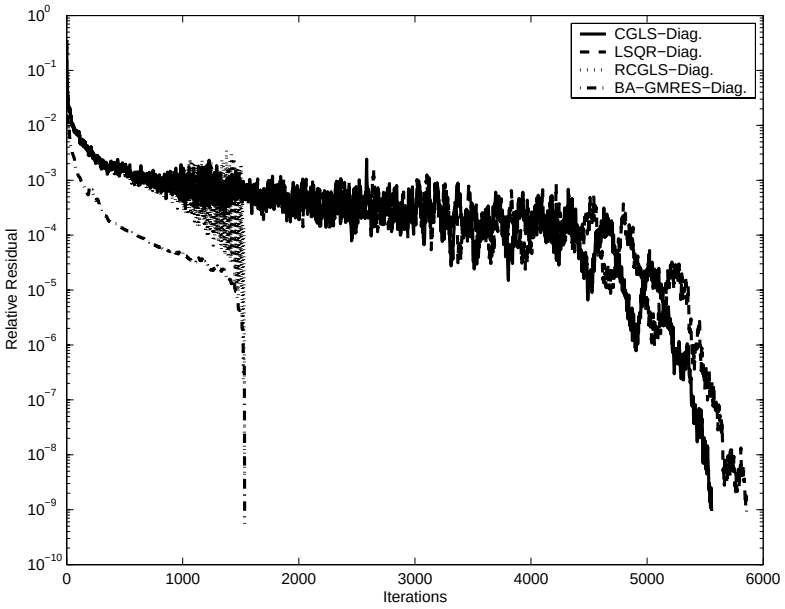

FIG. 6.2. $\|A^T\boldsymbol{r}\|_2/\|A^T\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL5, diagonal scaling).*

The figures show that the CGLS and LSQR methods are slow to converge compared to the BA-GMRES and reorthogonalized CGLS methods for these ill-conditioned problems. This can be explained by the fact that the CGLS and LSQR methods rely on three-term recurrence and suffer from loss of orthogonality due to rounding errors especially for ill-conditioned problems, whereas the BA-GMRES and reorthogonalized CGLS methods are more robust against loss of orthogonality because they perform explicit orthogonalization by the modified Gram-Schmidt procedure and reorthogonalization, respectively.

The BA-GMRES and the reorthogonalized CGLS show similar convergence behaviours. This is in accordance with the convergence analysis in Section 4.1, where we obtained similar upper bounds for the BA-GMRES and the similarly preconditioned CGLS in the absence of rounding errors.

The RIF preconditioning significantly improves convergence over the diagonal scaling.

The BA-GMRES converges more smoothly compared to the reorthogonalized CGLS.

In Table 6.5, we compare the methods for the problems in Table 6.1. The first row in each cell gives the number of iterations required for convergence, and the second row gives the total computation time in seconds. The value of the optimal relative drop tolerance parameter τ for the RIF preconditioning is also indicated for each problem.

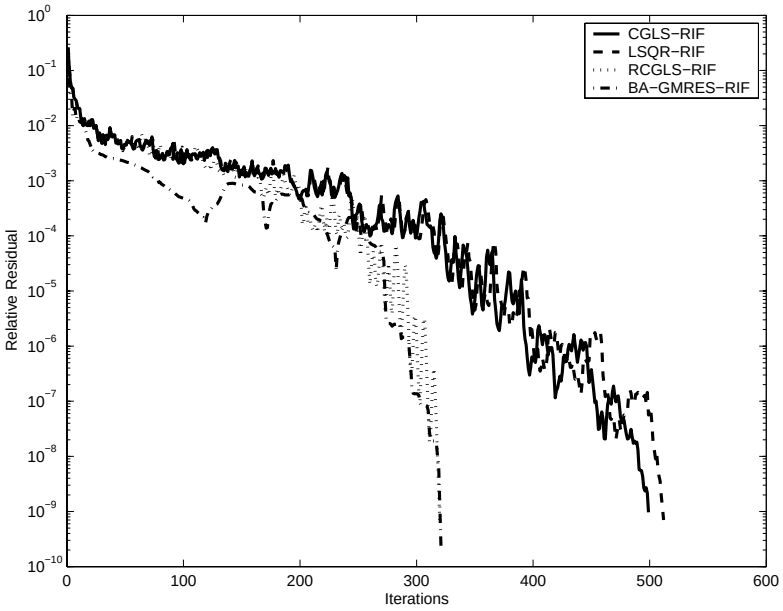

FIG. 6.3. $\|A^T\boldsymbol{r}\|_2/\|A^T\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL5, RIF,* $\tau = 0.8$*).*

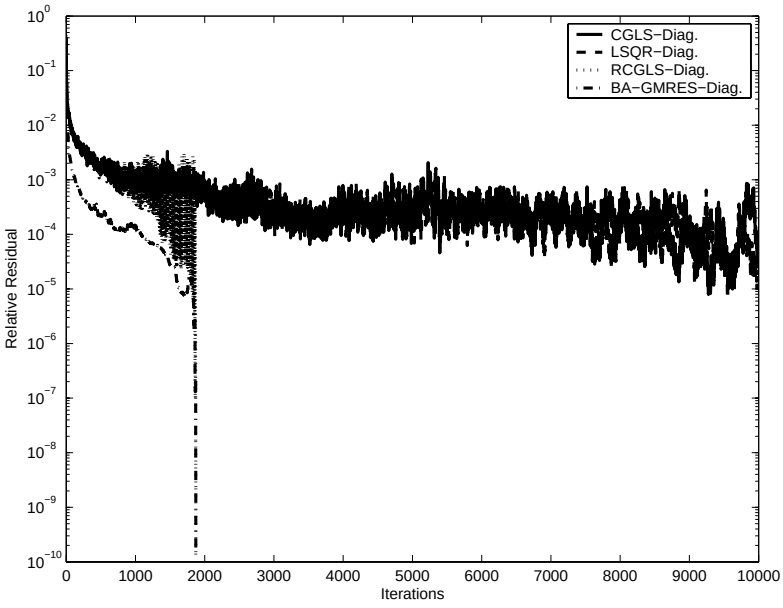

FIG. 6.4. $\|A^T\boldsymbol{r}\|_2/\|A^T\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL6, diagonal scaling).*

For the problems RANDL1 to RANDL4, the CGLS or LSQR method with diagonal scaling were the fastest.

As the condition number increases, the number of iterations for CGLS (and its stabilized version, LSQR) increases much more rapidly than the correspondingly preconditioned BA-GMRES method.

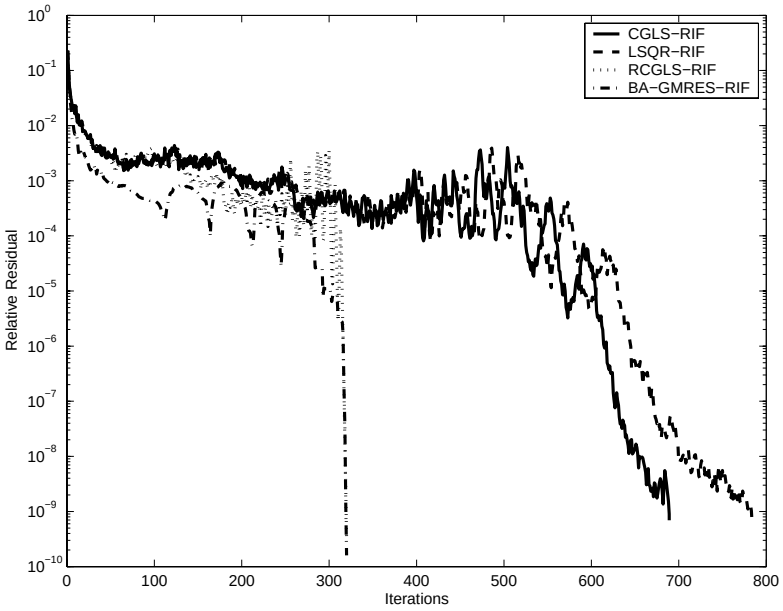

FIG. 6.5. $\|A^T\boldsymbol{r}\|_2/\|A^T\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL6, RIF,* $\tau = 0.07$*).*

TABLE 6.5
Comparison of the iterative methods.

	CGLS		LSQR		RCGLS		BA-GMRES	
	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF
RANDL1	35	14	35	14	35	14	35	14
($\tau = 0.5$)	$*$0.10	4.98	$*$0.10	4.98	0.16	4.99	0.14	4.99
RANDL2	214	21	214	21	208	21	193	21
($\tau = 0.7$)	$*$0.61	5.10	$*$0.61	5.10	2.77	5.13	2.28	5.11
RANDL3	742	72	740	73	697	70	622	60
($\tau = 0.8$)	2.08	5.38	$*$2.07	5.38	26.30	5.62	20.70	5.47
RANDL4	1,147	85	1,154	85	1,062	84	1,069	82
($\tau = 0.5$)	$*$3.22	6.38	3.38	6.38	59.35	7.17	59.41	6.62
RANDL5	4,897	470	5,064	401	1,521	305	1,522	299
($\tau = 0.9$)	13.74	$*$12.78	14.03	11.79	119.70	14.42	118.87	13.91
RANDL6	10,551	615	11,088	645	1,861	317	1,862	318
($\tau = 0.07$)	29.65	33.67	30.21	34.15	177.94	26.93	176.93	$*$26.73
RANDL7	32,143	1,951	35,034	2,443	1,914	371	1,899	362
($\tau = 0.02$)	89.93	102.28	91.31	128.40	195.63	40.07	183.90	$*$37.26

First row: number of iterations, second row: computation time (seconds).
Convergence criterion: $\|A^T r\|_2/\|A^T b\|_2 < 10^{-6}$.

For RANDL5, CGLS with RIF preconditioning is the fastest.

For the strongly ill-conditioned problems RANDL6 and RANDL7, the BA-GMRES method with RIF requires far less iterations than the CGLS with RIF, so that it is the fastest method with respect to computation time as well as the number of iterations.

6.2. Under-determined problems. Next, we show numerical experiment results for the under-determined least squares problem

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2, \qquad A \in \mathbf{R}^{m\times n} \qquad (m < n).$$

Here, we compare the AB-GMRES (BA-GMRES) methods with $B = A^{\mathrm{T}}C$ where C is obtained by diagonal scaling or RIF, with the corresponding preconditioned CGNE, LSQR and reorthogonalized CGNE (RCGNE) methods.

The coefficient matrices A were obtained by transposing the matrices RANDLn in Table 6.1 for the over-determined problems, and are denoted by RANDLnT. The density and condition number of RANDLnT is the same as that of the corresponding matrix RANDLn.

Since $\operatorname{rank} A = m < n$, the systems are consistent. Hence, the right-hand side vector $\boldsymbol{b}$ was given by $\boldsymbol{b} = A\boldsymbol{x}^*$ where $\boldsymbol{x}^* = (1, \ldots, 1)^{\mathrm{T}}$, and the initial approximate solution was set to $\boldsymbol{x}_0 = \boldsymbol{0}$. Hence, the convergence of the methods were judged by $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ where $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$.

Note that for the AB-GMRES method, $\|\boldsymbol{r}\|_2$ is available at each iteration without extra computational cost.

Figure 6.6 shows $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ vs. the number of iterations for the full AB-GMRES and BA-GMRES methods with diagonal scaling and RIF preconditioning, for the problem RANDL3T. τ for RIF was set to the optimal value 0.8. In the figure, the AB-GMRES and BA-GMRES methods show similar convergence behaviours, as predicted in Section 4.2.

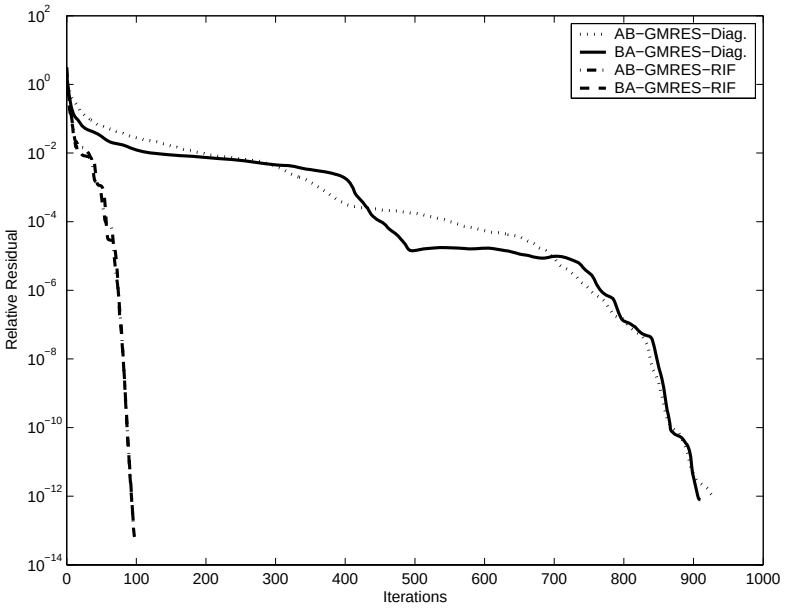

FIG. 6.6. *Comparison of GMRES-AB method and GMRES-BA method (RANDL3T, $\tau = 0.8$ for RIF).*

In the experiments below, we compare the AB-GMRES method with the CGNE, LSQR and reorthogonalized CGNE methods. The reason why the AB-GMRES method is used instead of the BA-GMRES method is because the former gives the minimum

norm solution, and requires less computation per iteration when $m < n$. (See section 3.3.) The full AB-GMRES method without restart was used. Similar to the over-determined systems, we chose the optimal τ for the RIF preconditioner for all the problems.

Note that, although the CGNE guarantees to give the minimum norm least squares solution, the LSQR does not. However, it was observed that the approximate solutions converge to the same solution for both methods in the following numerical experiments.

Figures 6.7, 6.8, 6.9 and 6.10 show the relative residual $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ vs. the number of iterations for the different methods for the problems RANDL5T and RANDL6T. τ for RIF was set to the optimal value 0.9 and 0.01, respectively. The figures show that the AB-GMRES method converges faster than the corresponding CGNE and LSQR methods, for diagonal scaling and RIF.

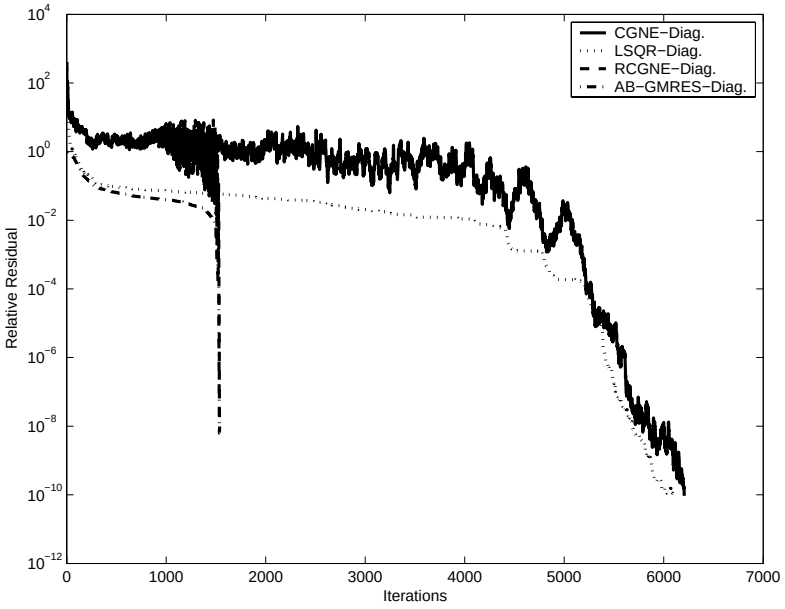

FIG. 6.7. $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL5T, diagonal scaling).*

Table 6.6 compares the methods for under-determined systems RANDLnT. The first row in each box gives the number of iterations required for convergence, and the second row gives the total computation time in seconds. It is observed that the AB-GMRES-RIF method is the fastest method for the ill-conditioned problems RANDL6T and RANDL7T.

6.3. Required memory. One drawback of the GMRES based methods is that they require increasingly more memory with the number of iterations or the restarting cycle k, whereas the CG based methods (without reorthogonalization) require constant memory. This is because the GMRES based methods require storing the orthonormal vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ in the modified Gram-Schmidt process, as well as the Hessenberg matrix.

Table 6.7 shows the memory required other than the coefficient matrix A and the preconditioner for each method. $\boldsymbol{r}, \boldsymbol{p}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{e}$ and $\boldsymbol{y}$ are the intermediate vectors,

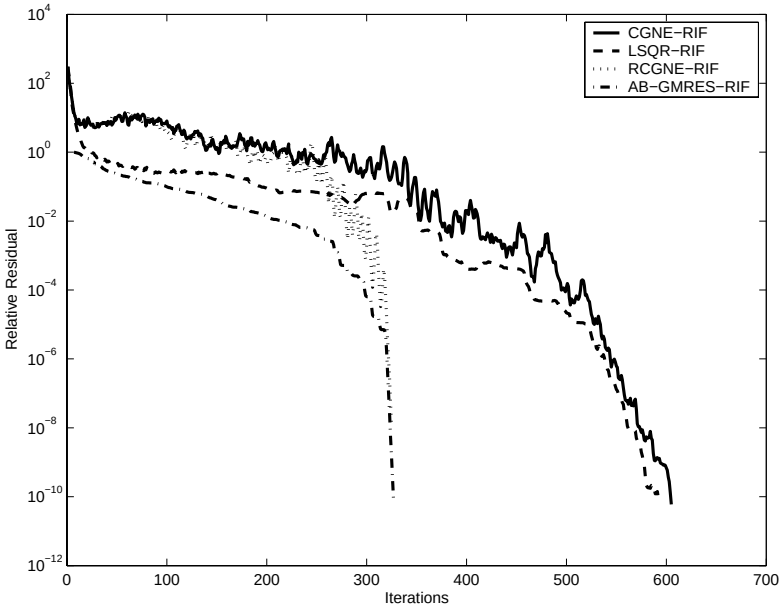

FIG. 6.8. $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL5T, RIF,* $\tau = 0.9$*).*

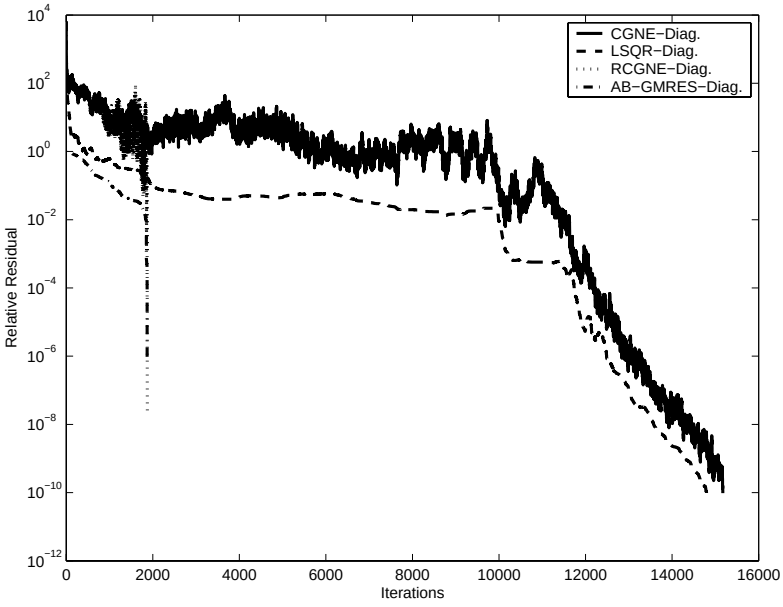

FIG. 6.9. $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL6T, diagonal scaling).*

and V denotes the k orthonormal vectors in the modified Gram-Schmidt process. For the LSQR method, we used the notation of variables according to [14].

If one can keep the number of iterations k of the GMRES type methods sufficiently small compared to m or n with the use of an efficient preconditioner like RIF, they may be faster compared to the CG type methods, and the memory required may not

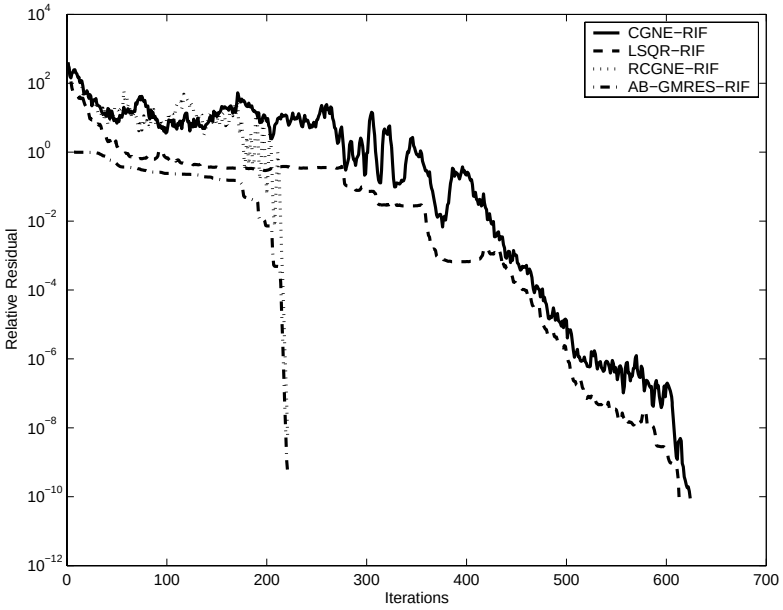

FIG. 6.10. $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ *vs. number of iterations (RANDL6T, RIF,* $\tau = 0.01$*).*

TABLE 6.6
Comparison of the iterative methods.

	CGNE		LSQR		RCGNE		AB-GMRES	
	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF
RANDL1T	37	12	37	12	37	12	37	12
($\tau = 0.4$)	$*$0.15	5.01	$*$0.15	5.02	0.23	5.02	0.17	5.01
RANDL2T	259	25	256	25	252	25	247	25
($\tau = 0.7$)	$*$1.06	5.16	$*$1.06	5.16	4.22	5.17	3.75	5.17
RANDL3T	838	81	823	80	783	77	754	75
($\tau = 0.8$)	$*$3.43	5.53	$*$3.43	5.53	33.87	5.81	30.56	5.67
RANDL4T	1,464	116	1,407	114	1,223	106	1,187	106
($\tau = 0.5$)	5.97	6.94	$*$5.96	6.94	79.67	7.45	73.67	7.21
RANDL5T	5,548	544	5,414	539	1,535	322	1,533	322
($\tau = 0.9$)	22.61	13.88	22.71	$*$13.86	123.80	15.58	121.58	15.11
RANDL6T	12,837	514	12,486	502	1,873	219	1,871	218
($\tau = 0.01$)	52.38	39.04	50.10	38.99	182.51	27.43	179.85	$*$26.99
RANDL7T	38,397	3,078	37,792	2,979	2,240	451	2,238	450
($\tau = 0.04$)	156.01	94.19	152.07	91.69	366.88	44.49	353.92	$*$43.76

First row: number of iterations, second row: computation time (seconds).
Convergence criterion: $\|r\|_2/\|b\|_2 < 10^{-6}$.

be prohibitive.

7. Conclusions. We proposed two methods for applying the GMRES method to linear least squares problems with $m \times n$ coefficient matrix A, using an $n \times m$ matrix B. The first method is to apply GMRES to $\min_{\boldsymbol{z} \in \mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2$ (AB-GMRES

TABLE 6.7
Intermediate memory required for each method for k iterations.

	dim(m)	dim(n)	dim(k)	total
CGLS	$\boldsymbol{r}, A\boldsymbol{p}$	$\boldsymbol{x}, \boldsymbol{p}, A^{\mathrm{T}}\boldsymbol{r}$		$2m + 3n$
CGNE	$\boldsymbol{r}, A\boldsymbol{p}$	$\boldsymbol{x}, \boldsymbol{p}, A^{\mathrm{T}}\boldsymbol{r}$		$2m + 3n$
LSQR	$\boldsymbol{u}$	$\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}$		$m + 3n$
AB-GMRES(k)	$V, \boldsymbol{w}$	$\boldsymbol{x}$	$\boldsymbol{y}, \boldsymbol{e}$	$(k+1)m + n + 2k + k^2/2$
BA-GMRES(k)		$V, \boldsymbol{w}, \boldsymbol{x}$	$\boldsymbol{y}, \boldsymbol{e}$	$(k+2)n + 2k + k^2/2$

method), and the second method is to apply GMRES to $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$ (BA-GMRES method).

Then, we derived a sufficient condition for B, such that the methods give a least squares solution for arbitrary $\boldsymbol{b}$ and $\boldsymbol{x}_0$ without breakdown.

Next, we showed that, theoretically, one may expect similar convergence behaviours for the AB- and BA- GMRES methods as well as the corresponding CGLS type methods.

Further, we proposed using the robust incomplete factorization (RIF) method for B in the GMRES methods.

Numerical experiments on over-determined problems with full column rank showed that the BA-GMRES method with the RIF preconditioner was faster than the RIF preconditioned CGLS, LSQR and reorthogonalized CGLS methods, for ill-conditioned problems.

For under-determined problems with full row rank, the AB-GMRES method with RIF preconditioning was faster than the RIF preconditioned CGNE, LSQR and reorthogonalized CGNE methods, for ill-conditioned problems.

Acknowledgments. We would like to thank Professors Shao-Liang Zhang, Kokichi Sugihara, Yimin Wei, Yohsuke Hosoda, Kunio Tanabe, Takashi Tsuchiya, Michael Eiermann, Masaaki Sugihara, Zhong-Zhi Bai, Michele Benzi, Lothar Reichel, Gene Golub, Michael Saunders and Yang-Feng Su for valuable remarks and discussions.

This research was supported by the Grants-in-Aid for Scientific Research (C) No. 17560056 of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

[1] Z. -Z. BAI, I. S. DUFF, AND A. J. WATHEN, *A class of incomplete orthogonal factorization methods. I: Methods and theories*, BIT, 41 (2001), pp. 53–70.
[2] M. BENZI AND M. TUMA, *A robust incomplete factorization preconditioner for positive definite matrices*, Numer. Linear Algebra Appl., 10 (2003), pp. 385–400.
[3] M. BENZI AND M. TUMA, *A robust preconditioner with low memory requirements for large sparse least squares problems*, SIAM J. Sci. Comput., 25 (2003), pp. 499–512.
[4] A. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[5] A. BJÖRCK, T. ELFVING, AND Z. STRAKOŠ, *Stability of conjugate gradient and Lanczos methods for linear least squares problems*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 720–736.
[6] P. N. BROWN AND H. F. WALKER, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.
[7] D. CALVETTI, B. LEWIS, AND L. REICHEL, *GMRES-type methods for inconsistent systems*, Linear Algebra Appl., 316 (2000), pp. 157–169.
[8] K. HAYAMI AND T. ITO, *Application of the GMRES method to singular systems and least squares problems*, in Proceedings of the Seventh China-Japan Seminar on Numerical Math-

ematics and Scientific Computing, Z. -C. Shi and H. Okamoto, eds., Zhangjiajie, 2004, Science Press, Beijing, 2006, pp. 31–44.

[9] K. Hayami and T. Ito, *Solution of least squares problems using GMRES methods*, Proc. Inst. Statist. Math., 53 (2005), pp. 331–348 (in Japanese).

[10] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, B49 (1952), pp. 409–436.

[11] T. Ito and K. Hayami, *Preconditioned GMRES methods for least squares problems*, NII Technical Report, NII-2004-006E, National Institute of Informatics, Tokyo, Japan, May, 2004, pp. 1–29.

[12] A. Jennings and M. A. Ajiz, *Incomplete methods for solving $A^TAx = b$*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 978–987.

[13] J. A. Meijerink and H. van der Vorst, *An iterative method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comp., 31 (1977), pp. 148–162.

[14] C. C. Paige and M. A. Saunders, *A robust incomplete factorization preconditioner for positive definite matrices*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.

[15] A. Papadopoulos, I. S. Duff, and A. J. Wathen, *A class of incomplete orthogonal factorization methods II: implementation and results*, BIT, 45 (2005), pp. 159–179.

[16] L. Reichel and Q. Ye, *Breakdown-free GMRES for singular systems*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1001–1021.

[17] Y. Saad, *Preconditioning techniques for nonsymmetric and indefinite linear systems*, J. Comput. Appl. Math., 24 (1988), pp. 89–105.

[18] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Second ed., SIAM, Philadelphia, 2003.

[19] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual method for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[20] K. Tanabe, *Characterization of linear stationary iterative processes for solving a singular system of linear equations*, Numer. Math., 22 (1974), pp. 349–359.

[21] K. Vuik, A. G. J. Sevink, and G. C. Herman, *A preconditioned Krylov subspace method for the solution of least squares problems in inverse scattering*, J. Comput. Phys., 123 (1996), pp. 330–340.

[22] X. Wang, K. A. Gallivan, and R. Bramley, *CIMGS: An incomplete orthogonal factorization preconditioner*, SIAM J. Sci. Comput., 18 (1997), pp. 516–536.

[23] S. -L. Zhang and Y. Oyanagi, *Orthomin(k) method for linear least squares problem*, J. Inf. Process., 14 (1991), pp. 121–125.