ISSN 1346-5597

National Institute of Informatics

NII Technical Report

Convergence of Inner-iteration GMRES Methods for Least Squares Problems (Revised Version)

Keiichi Morikuni and Ken Hayami

NII-2013-004E
Nov. 2013

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

CONVERGENCE OF INNER-ITERATION GMRES METHODS FOR LEAST SQUARES PROBLEMS*

KEIICHI MORIKUNI† AND KEN HAYAMI‡

Abstract. We develop a general convergence theory for the generalized minimal residual method preconditioned by inner iterations for solving least squares problems. The inner iterations are performed by stationary iterative methods. We also present theoretical justifications for using specific inner iterations such as the Jacobi and SOR-type methods. The theory improves previous work [K. Morikuni and K. Hayami, SIAM J. Matrix Appl. Anal., 34 (2013), pp. 1–22], particularly in the rank-deficient case. We also characterize the spectrum of the preconditioned coefficient matrix by the spectral radius of the iteration matrix for the inner iterations, and give a convergence bound for the proposed methods. Finally, numerical experiments show that the proposed methods are more robust and efficient compared to previous methods for some rank-deficient problems.

Key words. least squares problems, iterative methods, preconditioner, inner-outer iteration, GMRES method, stationary iterative method, rank-deficient problem

AMS subject classifications. 65F08, 65F10, 65F20, 65F50

1. Introduction.

Consider solving least squares problems

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

where $A\in\mathbf{R}^{m\times n}$ is not necessarily of full rank and $\boldsymbol{b}\in\mathbf{R}^m$ is not necessarily in $\mathcal{R}(A)$, the range space of A. The least squares problem (1.1) is equivalent to the normal equations

$$A^{\mathsf{T}}A\boldsymbol{x}=A^{\mathsf{T}}\boldsymbol{b}, \tag{1.2}$$

and in general has an infinite number of solutions. If $\boldsymbol{b}\in\mathcal{R}(A)$, the problem

$$\min_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{x}\|_2,\ \mathcal{S}=\{\boldsymbol{x}\in\mathbf{R}^n : A\boldsymbol{x}=\boldsymbol{b}\} \tag{1.3}$$

has a unique solution called the minimum-norm solution and is equivalent to the normal equations of the second kind

$$\boldsymbol{x}=A^{\mathsf{T}}\boldsymbol{u} \text{ subject to } AA^{\mathsf{T}}\boldsymbol{u}=\boldsymbol{b}. \tag{1.4}$$

By applying $B\in\mathbf{R}^{n\times m}$, we may transform the problem (1.1) to equivalent problems.

THEOREM 1.1 ([14, 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* $\mathcal{R}(AB)=\mathcal{R}(A)$.

THEOREM 1.2 ([14, Theorem 3.11]). $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b}-A\boldsymbol{x}\|_2$ *and* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b}-BA\boldsymbol{x}\|_2$ *are equivalent for all* $\boldsymbol{b}\in\mathbf{R}^m$ *if and only if* $\mathcal{R}(B^{\mathsf{T}}BA)=\mathcal{R}(A)$.

*This work was supported by the Grants-in-Aid for Scientific Research (C) of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

†National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan (morikuni@nii.ac.jp).

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

Thus, the original problem (1.1) may be reduced to least squares problems with a square matrix AB or BA. Based on these transformations, the generalized minimal residual method (GMRES) [24] was applied to solve least squares problems (1.1) in [14]. The right- and left-preconditioned GMRES for least squares problems were called AB- and BA-GMRES, respectively. Sufficient conditions under which these methods determine a least squares solution for arbitrary $\boldsymbol{b}$ were shown.

In [19], these methods were preconditioned by several iterations of stationary iterative methods such as variants of the Jacobi overrelaxation (JOR) and successive overrelaxation (SOR) methods, which may be considered as inner iterations. In [20], we assumed that A should be of full-column rank for the convergence theory for BA-GMRES with the Cimmino-NR (Normal Residual) and NR-SOR inner iterations, but numerical experiments showed that these methods actually converge also for rank-deficient problems. In this paper, we give theoretical justifications for the convergence also in the rank-deficient case, for which only few preconditioners such as diagonal scaling and the Greville's method [6] were known to work.

The outline of the paper is as follows. In section 2, we give a convergence condition for GMRES, which will be used later. In section 3, we introduce BA-GMRES, correct its convergence theorem in [14], analyze the spectrum of the preconditioned matrix, and give a convergence bound for BA-GMRES. In section 4, we introduce inner-iteration preconditioning for BA-GMRES, and give main results on sufficient conditions for convergence in terms of the inner iterations. We also give a convergence bound for the method. In section 5, we introduce AB-GMRES, extend its convergence theorem in [14], and give a convergence theory also for AB-GMRES preconditioned by inner iterations. In section 6, we show numerical results comparing these methods with previous methods. In section 7, we conclude the paper.

Throughout this paper, we use bold letters for column vectors. $\boldsymbol{e}_j$ denotes the jth column of the identity matrix. We denote quantities related to the kth inner iteration by using a superscript with brackets, e.g., $\boldsymbol{x}^{(k)}$, and the kth outer iteration by using a subscript without brackets, e.g., $\boldsymbol{x}_k$. $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{b}$ between real vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. $\mathcal{N}(A)$ denotes the null space of A. $\mathcal{S}^{\perp}$ denotes the orthogonal complement of a subspace $\mathcal{S}$.

2. GMRES method on linear systems including the singular case. We first explain GMRES applied to the linear system $\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$ with initial approximate solution $\tilde{\boldsymbol{x}}_0 \in \mathbf{R}^N$, where $\tilde{A} \in \mathbf{R}^{N\times N}$ is not necessarily nonsingular. GMRES determines the kth approximate solution $\tilde{\boldsymbol{x}}_k$ in $\tilde{\boldsymbol{x}}_0 + \mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0)$ which minimizes $\|\tilde{\boldsymbol{b}} - \tilde{A}\tilde{\boldsymbol{x}}_k\|_2$, where

$$\mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0) = \mathrm{span}\{\tilde{\boldsymbol{r}_0}, \tilde{A}\tilde{\boldsymbol{r}}_0, \ldots, \tilde{A}^{k-1}\tilde{\boldsymbol{r}}_0\}$$

is the Krylov subspace of order k and $\tilde{\boldsymbol{r}}_0 = \tilde{\boldsymbol{b}} - \tilde{A}\tilde{\boldsymbol{x}}_0$ is the initial residual.

We next present the convergence condition for GMRES for the linear system $\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$, independent of any particular implementation of the algorithm. Note that, in general, $\dim \tilde{A}\mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0) \le \dim \mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0) \le k$ holds for each k.

Definition 2.1 ([4, p. 38]). *GMRES is said to break down at some step k if*

$$\dim \tilde{A}\mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0) < \dim \mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0) \text{ or } \dim \mathcal{K}_k(\tilde{A}, \tilde{\boldsymbol{r}}_0) < k.$$

Theorem 2.2. *GMRES determines a solution of $\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$ without breakdown for all $\tilde{\boldsymbol{b}} \in \mathcal{R}(\tilde{A})$ and for all $\tilde{\boldsymbol{x}}_0 \in \mathbf{R}^N$ if and only if $\mathcal{R}(\tilde{A}) \cap \mathcal{N}(\tilde{A}) = \{\boldsymbol{0}\}$.*

Proof. The sufficiency was shown in [4, Theorem 2.6]. Now we show the necessity. Assume $\mathcal{R}(\tilde{A})\cap\mathcal{N}(\tilde{A}) \neq \{\mathbf{0}\}$, or equivalently let $\tilde{A}_{11}$ be singular, where $\tilde{A}_{11} = \tilde{Q}_1^{\mathsf{T}}\tilde{A}\tilde{Q}_1$, $\mathcal{R}(\tilde{Q}_1) = \mathcal{R}(\tilde{A})$, $\tilde{Q}_1^{\mathsf{T}}\tilde{Q}_1 = \mathrm{I}_r$, and I_r is the identity matrix of size $r = \mathrm{rank}A$ [13, Theorem 2.3]. Then, there exists $\boldsymbol{s}^1 \neq \mathbf{0}$ such that $\tilde{A}_{11}\boldsymbol{s}^1 = \mathbf{0}$. Let $\tilde{\boldsymbol{b}} = \tilde{Q}_1\boldsymbol{s}^1 + \tilde{A}\tilde{\boldsymbol{x}}_0 \in \mathcal{R}(\tilde{A})$. Then, $\tilde{\boldsymbol{r}}_0 = \tilde{Q}_1\boldsymbol{s}^1 \neq \mathbf{0}$.

In step $k = 1$, $\tilde{\boldsymbol{x}}_1 = \tilde{\boldsymbol{x}}_0 + c\tilde{\boldsymbol{r}}_0$, $c \in \mathbf{R}$ and $\tilde{\boldsymbol{r}}_1 = \tilde{\boldsymbol{b}} - \tilde{A}\tilde{\boldsymbol{x}}_1 = \tilde{\boldsymbol{r}}_0 - c\tilde{A}\tilde{\boldsymbol{r}}_0$. Let $\tilde{Q}_2 \in \mathbf{R}^{N\times(N-r)}$ such that $\mathcal{R}(\tilde{Q}_2) = \mathcal{R}(\tilde{A})^{\perp}$ and $\tilde{Q}_2^{\mathsf{T}}\tilde{Q}_2 = \mathrm{I}_{N-r}$, and $\tilde{Q} = \left[\tilde{Q}_1, \tilde{Q}_2\right]$. Since

$$\tilde{Q}^{\mathsf{T}}\tilde{\boldsymbol{r}}_1 = \tilde{Q}^{\mathsf{T}}\tilde{\boldsymbol{r}}_0 - c(\tilde{Q}^{\mathsf{T}}\tilde{A}\tilde{Q})\tilde{Q}^{\mathsf{T}}\tilde{\boldsymbol{r}}_0 = \tilde{Q}^{\mathsf{T}}\tilde{\boldsymbol{r}}_0 - c\begin{bmatrix}\tilde{A}_{11} & \tilde{A}_{12}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{s}^1\\ \mathbf{0}\end{bmatrix} = \tilde{Q}^{\mathsf{T}}\tilde{\boldsymbol{r}}_0$$

and $\tilde{\boldsymbol{r}}_1 = \tilde{\boldsymbol{r}}_0 \neq \mathbf{0}$, $\boldsymbol{x}_1$ is not a solution of $\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$. Moreover, since $\dim\tilde{A}\mathcal{K}_1(\tilde{A}, \tilde{\boldsymbol{r}}_0) = 0 < \dim\mathcal{K}_1(\tilde{A}, \tilde{\boldsymbol{r}}_0) = 1$, GMRES breaks down at the first step. □

This theorem is similar to [13, Theorem 2.8], which was for the standard GMRES method using Gram-Schmidt orthogonalization, whose breakdown was defined as Definition A.3. Here, we have generalized [13, Theorem 2.8] by using the more general definition of breakdown given in Definition 2.1. We give a discussion on the relation between the two breakdowns in Appendix A.

3. BA-GMRES method. Consider solving (1.1). BA-GMRES [14] applies GMRES to $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$ and works in a smaller n-dimensional space than the m-dimensional space for AB-GMRES in the overdetermined case $m > n$. The algorithm of BA-GMRES with the modified Gram-Schmidt orthogonalization is given as follows.

ALGORITHM 3.1. *BA-GMRES method.*

1. *Let $\boldsymbol{x}_0$ be the initial approximate solution and $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.*
2. *$\boldsymbol{z}_0 := B\boldsymbol{r}_0$, $\beta := \|\boldsymbol{z}_0\|_2$, $\boldsymbol{v}_1 := \boldsymbol{z}_0/\beta$*
3. *For $k = 1, 2, \ldots$ until convergence, Do*
4. * $\boldsymbol{w}_k := BA\boldsymbol{v}_k$*
5. * For $i = 1, 2, \ldots, k$, Do*
6. * $h_{i,k} := (\boldsymbol{w}_k, \boldsymbol{v}_i)$, $\boldsymbol{w}_k := \boldsymbol{w}_k - h_{i,k}\boldsymbol{v}_i$*
7. * EndDo*
8. * $h_{k+1,k} := \|\boldsymbol{w}_k\|_2$, $\boldsymbol{v}_{k+1} := \boldsymbol{w}_k/h_{k+1,k}$*
9. *EndDo*
10. *$\boldsymbol{y}_k := \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}\|_2$, $\boldsymbol{x}_k := \boldsymbol{x}_0 + [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k]\,\boldsymbol{y}_k$*

Here, $\bar{H}_k = \{h_{i,j}\} \in \mathbf{R}^{(k+1)\times k}$.

We say that BA-GMRES breaks down at some step k if $\dim BA\mathcal{K}_k(BA, B\boldsymbol{r}_0) < \dim\mathcal{K}_k(BA, B\boldsymbol{r}_0)$ or $\dim\mathcal{K}_k(BA, B\boldsymbol{r}_0) < k$ (cf. Definition 2.1). Applying Theorem 2.2 to BA-GMRES gives the following.

THEOREM 3.2. *Assume $\mathcal{R}(B^{\mathsf{T}}) = \mathcal{R}(A)$. Then, BA-GMRES determines a solution of $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ without breakdown for all $\boldsymbol{b} \in \mathbf{R}^m$ and for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ if and only if $\mathcal{R}(B) \cap \mathcal{N}(A) = \{\mathbf{0}\}$.*

Proof. Substitute BA, $\boldsymbol{x}$, and $B\boldsymbol{b}$ into $\tilde{A}$, $\tilde{\boldsymbol{x}}$, and $\tilde{\boldsymbol{b}}$, respectively, in Theorem 2.2. $\mathcal{R}(B^{\mathsf{T}}) = \mathcal{R}(A)$ gives $\mathcal{N}(BA) = \mathcal{R}(A^{\mathsf{T}}B^{\mathsf{T}})^{\perp} = \mathcal{R}(A^{\mathsf{T}}A)^{\perp} = \mathcal{R}(A^{\mathsf{T}})^{\perp} = \mathcal{N}(A)$ and $\mathcal{R}(BA) = \mathcal{R}(BB^{\mathsf{T}}) = \mathcal{R}(B)$ Hence, "for all $B\boldsymbol{b} \in \mathcal{R}(BA) = \mathcal{R}(B)$" is equivalent to "for all $\boldsymbol{b} \in \mathbf{R}^m$". $\mathcal{R}(B^{\mathsf{T}}) = \mathcal{R}(A)$ also gives $\mathcal{R}(B^{\mathsf{T}}BA) = \mathcal{R}(B^{\mathsf{T}}B) = \mathcal{R}(B^{\mathsf{T}}) = \mathcal{R}(A)$. Therefore, Theorem 1.2 completes the proof. □

This theorem corrects [14, Theorem 3.18].

3.1. Spectrum of the preconditioned matrix. Next, we analyze the spectrum of the preconditioned matrix BA. Assume $\mathcal{R}(B^{\mathsf{T}}) = \mathcal{R}(A)$. Then, $B\boldsymbol{b} \in \mathcal{R}(BA) = \mathcal{R}(B)$ holds, and $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$ is equivalent to $BA\boldsymbol{x} = B\boldsymbol{b}$. Let $r = \mathrm{rank}A$, $Q_1 \in \mathbf{R}^{n\times r}$ such that $\mathcal{R}(Q_1) = \mathcal{R}(BA)$, $Q_2 \in \mathbf{R}^{n\times(n-r)}$ such that $\mathcal{R}(Q_2) = \mathcal{R}(BA)^{\perp}$, and $Q = [Q_1, Q_2]$, where the columns of Q are orthonormal. Then, GMRES applied to $BA\boldsymbol{x} = B\boldsymbol{b}$ is equivalent to GMRES applied to

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^1 \\ \boldsymbol{x}^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}^1 \\ \boldsymbol{b}^2 \end{bmatrix},$$

where $A_{11} = Q_1^{\mathsf{T}}(BA)Q_1 \in \mathbf{R}^{r\times r}$, $A_{12} = Q_1^{\mathsf{T}}(BA)Q_2 \in \mathbf{R}^{r\times(n-r)}$, $\boldsymbol{x}^1 = Q_1^{\mathsf{T}}\boldsymbol{x}$, $\boldsymbol{x}^2 = Q_2^{\mathsf{T}}\boldsymbol{x}$, $\boldsymbol{b}^1 = Q_1^{\mathsf{T}}B\boldsymbol{b}$, and $\boldsymbol{b}^2 = Q_2^{\mathsf{T}}B\boldsymbol{b} = \boldsymbol{0}$ since $B\boldsymbol{b} \in \mathcal{R}(BA)$. As shown in [13], if $\boldsymbol{x}_0 \in \mathcal{R}(BA) = \mathcal{R}(B)$, then the $\mathcal{R}(BA)$ component of GMRES applied to $BA\boldsymbol{x} = B\boldsymbol{b}$, is equivalent to GMRES applied to $A_{11}\boldsymbol{x}^1 = \boldsymbol{b}^1$. On the other hand, in the $\mathcal{R}(BA)^{\perp}$ component, $\boldsymbol{x}_k^2 = \boldsymbol{x}_0^2$ for all iterates $\boldsymbol{x}_k$.

Now note the following.

THEOREM 3.3. *A_{11} is nonsingular if and only if $\mathcal{R}(BA) \cap \mathcal{N}(BA) = \{\boldsymbol{0}\}$.*

Proof. See [13, Theorem 2.3]. □

THEOREM 3.4. *Assume $\mathcal{R}(BA) \cap \mathcal{N}(BA) = \{\boldsymbol{0}\}$. Then, $\lambda \neq 0$ is an eigenvalue of BA if and only if $\lambda \neq 0$ is an eigenvalue of A_{11}.*

Proof. Let $Q = [Q_1, Q_2] \in \mathbf{R}^{n\times n}$ be as given above. Then,

$$\begin{aligned} \det(BA - \lambda\,\mathrm{I}) &= \det Q^{\mathsf{T}} \det(BA - \lambda\,\mathrm{I}) \det Q = \det\bigl(Q^{\mathsf{T}}BAQ - \lambda\,\mathrm{I}\bigr) \\ &= \det\left(\begin{bmatrix} A_{11} - \lambda\,\mathrm{I}_r & A_{12} \\ 0 & \lambda\,\mathrm{I}_{n-r} \end{bmatrix}\right) = (-\lambda)^{n-r} \det(A_{11} - \lambda\,\mathrm{I}_r) \end{aligned}$$

Hence, Theorem 3.3 completes the proof. □

3.2. Convergence bound for BA-GMRES. Next, we give a convergence bound for (BA-)GMRES.

THEOREM 3.5. *Let $\boldsymbol{z}_k = B(\boldsymbol{b} - A\boldsymbol{x}_k)$ be the kth residual for GMRES applied to $BA\boldsymbol{x} = B\boldsymbol{b}$ and T be the Jordan basis of BA. Assume $\mathcal{R}(B^{\mathsf{T}}) = \mathcal{R}(A)$, $\mathcal{R}(B) \cap \mathcal{N}(A) = \{\boldsymbol{0}\}$, $\boldsymbol{x}_0 \in \mathcal{R}(B)$, and that all the nonzero eigenvalues of BA are located in a disk in the open right (left) half plane with center c and radius a, excluding the origin. Then, we have*

$$\|\boldsymbol{z}_k\|_2 \leq \kappa(T) \left(\frac{a}{|c|}\right)^k \sum_{i=0}^{\tau(k,d)} \binom{k}{i} a^{-i} \|\boldsymbol{z}_0\|_2 \tag{3.1}$$

for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ and for all $\boldsymbol{b} \in \mathbf{R}^m$, where $\kappa(T) = \|T\|_2\|T^{-1}\|_2$, d is the size of the largest Jordan block of BA corresponding to a nonzero eigenvalue of BA, and $\tau(k,d) = \min\{k, d-1\}$.

Proof. Theorem 3.2 ensures that GMRES determines a solution of $BA\boldsymbol{x} = B\boldsymbol{b}$ without breakdown for all $\boldsymbol{b} \in \mathbf{R}^n$ and for all $\boldsymbol{x}_0 \in \mathbf{R}^n$. From [2, Theorem 1], we have

$$\|\boldsymbol{z}_k\|_2 = \min_{\substack{p\in P_k \\ p(0)=1}} \|p(BA)\boldsymbol{z}_0\|_2 \leq \kappa(T) \left(\min_{\substack{p\in P_k \\ p(0)=1}} \max_{1\leq i\leq M} \|p(J_i)\|_2 \right) \|\boldsymbol{z}_0\|_2,$$

where P_k is the set of all polynomials of degree not exceeding k and J_i is a Jordan block of BA corresponding to a nonzero eigenvalue, $i = 1, 2, \ldots, M$. From [2, Theorems 2,

5], the second factor is bounded as

$$\min_{\substack{p\in P_k\\ p(0)=1}} \max_{1\le i\le M} \|p(J_i)\|_2 \le \left(\frac{a}{|c|}\right)^k \sum_{i=0}^{\tau(k,d)} \binom{k}{i} a^{-i}.$$

This gives (3.1). □

The particular spectrum of BA assumed in this theorem is satisfied for inner-iteration preconditioning (see Theorem 4.9). We will use this theorem in section 4.3. Note that the residual $\|\boldsymbol{z}_k\|_2$ does not necessarily depend only on the eigenvalues of BA when $\kappa(T)$ is large [12], [1], [27]. A similar argument can be applied to AB-GMRES.

4. BA-GMRES preconditioned by stationary iterative methods as inner iterations. Instead of applying B explicitly as in Algorithm 3.1, consider using inner iterations as follows [20].

ALGORITHM 4.1. *BA-GMRES method preconditioned by inner iterations.*

1. *Let* $\boldsymbol{x}_0$ *be the initial approximate solution and* $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.
2. *Apply* ℓ *steps of a stationary iterative method to* $A^\mathsf{T}A\boldsymbol{z} = A^\mathsf{T}\boldsymbol{r}_0$ *to obtain* $\boldsymbol{z}_0 := B^{(\ell)}\boldsymbol{r}_0$.
3. $\beta := \|\boldsymbol{z}_0\|_2$, $\boldsymbol{v}_1 := \boldsymbol{z}_0/\beta$
4. *For* $k = 1, 2, \ldots$ *until convergence, Do*
5. $\quad \boldsymbol{u}_k := A\boldsymbol{v}_k$
6. $\quad$ *Apply* ℓ *steps of a stationary iterative method to* $A^\mathsf{T}A\boldsymbol{z} = A^\mathsf{T}\boldsymbol{u}_k$ *to obtain* $\boldsymbol{z}_k := B^{(\ell)}\boldsymbol{u}_k$.
7. $\quad$ *For* $i = 1, 2, \ldots, k$, *Do*
8. $\quad\quad h_{i,k} := (\boldsymbol{z}_k, \boldsymbol{v}_i)$, $\boldsymbol{z}_k := \boldsymbol{z}_k - h_{i,k}\boldsymbol{v}_i$
9. $\quad$ *EndDo*
10. $\quad h_{k+1,k} := \|\boldsymbol{z}_k\|_2$, $\boldsymbol{v}_{k+1} := \boldsymbol{z}_k/h_{k+1,k}$
11. *EndDo*
12. $\boldsymbol{y}_k := \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}\|_2$, $\boldsymbol{x}_k := \boldsymbol{x}_0 + [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k]\,\boldsymbol{y}_k$

Here, $B^{(\ell)}$ denotes the preconditioning matrix for ℓ inner iterations.

In lines 2 and 6 in Algorithm 4.1, stationary iterative methods are applied to the normal equations. We now introduce a stationary iterative method for the normal equations $A^\mathsf{T}A\boldsymbol{z} = A^\mathsf{T}\boldsymbol{c}$. Consider the splitting $A^\mathsf{T}A = M - N$, where M is nonsingular. Then, consider a class of iterative methods of the form

$$\boldsymbol{z}^{(\ell)} = M^{-1}N\boldsymbol{z}^{(\ell-1)} + M^{-1}A^\mathsf{T}\boldsymbol{c}.$$

Let $H = M^{-1}N = \mathrm{I} - M^{-1}A^\mathsf{T}A$ be the iteration matrix. In practice, there is no need to form $A^\mathsf{T}A$, M^{-1}, and N explicitly, as will be seen in the Richardson-NR, Cimmino-NR, NR-SOR, and NR-SSOR methods [23] in section 4.1.

Here, we define the following, e.g., [18].

DEFINITION 4.2. *A matrix* C *is called semi-convergent if* $\lim_{i\to\infty} C^i$ *exists.*

The semi-convergence is algebraically characterized by the following.

THEOREM 4.3 ([15], [21, Theorem 1], [26, Theorem 2]). *The following are equivalent.*

1. C *is semi-convergent.*
2. *For any eigenvalue* λ *of* C, *either*
   (a) $|\lambda| < 1$ *or*

(b) $\lambda = 1$ *and* $\mathrm{index}(\mathrm{I} - C) = 1$

holds.

Here, $\mathrm{index}(C)$ denotes the smallest nonnegative integer i such that $\mathcal{R}(C^i) = \mathcal{R}(C^{i+1})$. Thus, $\mathrm{index}(C)$ is equal to the size of the largest Jordan block corresponding to the zero eigenvalue of C.

This property has been used for analyzing the convergence of stationary iterative methods for singular linear systems [17], [26], [18], [9], and it will also play an important role in analyzing the convergence of our methods.

4.1. Convergence theory. The convergence analysis of BA-GMRES preconditioned by specific inner iterations in [20] was incomplete, especially in the rank-deficient case. The following convergence analysis for general inner iterations leads to a better understanding of our methods.

We first give an explicit expression for the preconditioned matrix $B^{(\ell)}A$ for BA-GMRES with ℓ inner iterations. Assume that the initial approximate solution for the inner iteration is $\boldsymbol{z}^{(0)} = \mathbf{0}$. Then, the ℓth iterate for the inner iteration is

$$\boldsymbol{z}^{(\ell)} = H\boldsymbol{z}^{(\ell-1)} + M^{-1}A^{\mathsf{T}}\boldsymbol{c} = \sum_{i=0}^{\ell-1} H^i M^{-1} A^{\mathsf{T}} \boldsymbol{c}. \tag{4.1}$$

Hence, if we define the preconditioning matrix by

$$B^{(\ell)} = \sum_{i=0}^{\ell-1} H^i M^{-1} A^{\mathsf{T}}, \tag{4.2}$$

we have $\boldsymbol{z}^{(\ell)} = B^{(\ell)}\boldsymbol{c}$. If $C^{(\ell)} = \sum_{i=0}^{\ell-1} H^i M^{-1}$, then $B^{(\ell)} = C^{(\ell)}A^{\mathsf{T}}$. Hence, the preconditioned matrix is expressed as

$$B^{(\ell)}A = C^{(\ell)}A^{\mathsf{T}}A = \sum_{i=0}^{\ell-1} H^i(\mathrm{I} - H) = \mathrm{I} - H^\ell. \tag{4.3}$$

We prepare the following.

LEMMA 4.4. *Let* $C \in \mathbf{R}^{n\times n}$. *Then,* $\mathrm{index}(C) \le 1$ *is equivalent to* $\mathcal{R}(C) \cap \mathcal{N}(C) = \{\mathbf{0}\}$.

Proof. Let $\boldsymbol{u} \in \mathcal{R}(C) \cap \mathcal{N}(C)$. Then, there exits a $\boldsymbol{w} \in \mathbf{R}^n$ such that $C\boldsymbol{w} = \boldsymbol{u}$. From the definition of the index, we have $\mathcal{N}(\mathrm{I}) = \mathcal{N}(C)$ or $\mathcal{N}(C) = \mathcal{N}(C^2)$. Since $C^2\boldsymbol{w} = C\boldsymbol{u} = \mathbf{0}$, $\boldsymbol{w} \in \mathcal{N}(C^2) = \mathcal{N}(C)$. Therefore, $\boldsymbol{u} = \mathbf{0}$. The converse is also true. □

LEMMA 4.5. *Let* $A \in \mathbf{R}^{m\times n}$ *and* $B^{(\ell)}$ *be given by* (4.2). *Assume that* H *is semi-convergent. Then,* $\mathrm{index}(B^{(\ell)}A) \le 1$ *for all* $\ell \ge 1$.

Proof. Let $J = S^{-1}(\mathrm{I} - H)S$ be the Jordan canonical form of $(\mathrm{I} - H)$. Assume that H is semi-convergent. Then, from Theorem 4.3, $\mathrm{index}(\mathrm{I} - H) = \mathrm{index}(J) \le 1$. Without loss of generality, we denote J by $J = \mathrm{diag}(\tilde{J}, 0_{n-r}) \in \mathbf{C}^{n\times n}$, where $r = \mathrm{rank}A$, $\tilde{J}$ has no eigenvalues equal to zero, and 0_{n-r} is the zero matrix of size $n - r$. Using (4.3), we have

$$B^{(\ell)}A = S \begin{bmatrix} \mathrm{I}_r - \left(\mathrm{I}_r - \tilde{J}\right)^\ell & 0 \\ 0 & 0_{n-r} \end{bmatrix} S^{-1}$$

for all $\ell \geq 1$. Let ν_i be an eigenvalue of H such that $|\nu_i| < 1$, and $\lambda_i = 1 - \nu_i$ be a corresponding nonzero eigenvalue of $(\mathrm{I} - H)$. Since $|1 - \lambda_i| = |\nu_i| < 1$, the corresponding eigenvalue of $\mathrm{I}_r - (\mathrm{I}_r - \tilde{J})^\ell$ is $\mu_i = 1 - (1 - \lambda_i)^\ell$ for all $\ell \geq 1$. If $\mu_i = 0$, then $(1 - \lambda_i)^\ell = 1$ for all $\ell \geq 1$, which contradicts $|1 - \lambda_i| < 1$. Hence, $\mathrm{I}_r - (\mathrm{I}_r - \tilde{J})^\ell$ is nonsingular for all $\ell \geq 1$. Therefore, $\mathrm{index}(B^{(\ell)}A) \leq 1$ for all $\ell \geq 1$. □

LEMMA 4.6. *Using the notations and the assumption of Lemma 4.5,* $\mathcal{R}\left(B^{(\ell)\mathsf{T}}\right) = \mathcal{R}(A)$ *holds for all* $\ell \geq 1$.

Proof. If $C^{(\ell)}$ is nonsingular, then $\mathcal{R}\left(B^{(\ell)\mathsf{T}}\right) = \mathcal{R}\left(AC^{(\ell)\mathsf{T}}\right) = \mathcal{R}(A)$. Hence, we show that $C^{(\ell)}$ is nonsingular

Assume that H is semi-convergent. Then, we have

$$C^{(\ell)} = \sum_{i=0}^{\ell-1} (\mathrm{I} - SJS^{-1})^i M^{-1} = S \begin{bmatrix} \left[\mathrm{I}_r - \left(\mathrm{I}_r - \tilde{J}\right)^\ell\right] \tilde{J}^{-1} & 0 \\ 0 & \ell\, \mathrm{I}_{n-r} \end{bmatrix} S^{-1} M^{-1}.$$

As in Lemma 4.5, $\left[\mathrm{I}_r - (\mathrm{I}_r - \tilde{J})^\ell\right]$ is nonsingular. Hence, $C^{(\ell)}$ is nonsingular for all $\ell \geq 1$. Therefore, we have $\mathcal{R}\left(B^{(\ell)\mathsf{T}}\right) = \mathcal{R}(A)$ for all $\ell \geq 1$. □

Hence, we obtain the main result.

THEOREM 4.7. *Assume that H is semi-convergent. Then, BA-GMRES with the inner-iteration preconditioning of the form* (4.1) *determines a least squares solution of* $\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *without breakdown for all* $\boldsymbol{b} \in \mathbf{R}^m$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$.

Proof. Assume that H is semi-convergent. Then, from Lemmas 4.4 and 4.5, we have $\mathrm{index}(B^{(\ell)}A) \leq 1$, or equivalently $\mathcal{R}(B^{(\ell)}A) \cap \mathcal{N}(B^{(\ell)}A) = \{0\}$. Moreover, since $\mathcal{R}\left(B^{(\ell)\mathsf{T}}\right) = \mathcal{R}(A)$ from Lemma 4.6, we have $\mathcal{R}(B^{(\ell)}A) = \mathcal{R}\left(B^{(\ell)}B^{(\ell)\mathsf{T}}\right) = \mathcal{R}(B^{(\ell)})$ and $\mathcal{N}(B^{(\ell)}A) = \mathcal{R}\left(A^\mathsf{T}B^{(\ell)\mathsf{T}}\right)^\perp = \mathcal{R}\left(A^\mathsf{T}A\right)^\perp = \mathcal{R}(A^\mathsf{T})^\perp = \mathcal{N}(A)$. Hence, Theorem 3.2 completes the proof. □

We remark that this theorem holds whether A is of full rank or rank-deficient, and whether A is overdetermined or underdetermined, i.e., unconditionally with respect to A.

Now, we consider applying Theorem 4.7 to BA-GMRES preconditioned by specific inner-iteration methods as follows. The inner-iteration preconditioning matrices for the Richardson-NR, Cimmino-NR, NR-SOR, and NR-SSOR methods are respectively obtained from (4.2) by setting

$$M = \begin{cases} \omega I & : \text{Richardson-NR}, \\ \omega D & : \text{Cimmino-NR}, \\ \frac{1}{\omega}(D + \omega L) & : \text{NR-SOR}, \\ \omega^{-1}(2-\omega)^{-1}(D + \omega L)D^{-1}(D + \omega L^\mathsf{T}) & : \text{NR-SSOR}, \end{cases} \tag{4.4}$$

where $A^\mathsf{T}A = L + D + L^T$, L is a strictly lower triangular matrix, D is a diagonal matrix, and ω is the relaxation parameter. Note that M is nonsingular in the last three cases if A has no zero columns. Cimmino-NR is mathematically equivalent to JOR applied to $A^\mathsf{T}A\boldsymbol{x} = A^\mathsf{T}\boldsymbol{b}$. NR-SOR is mathematically equivalent to SOR applied to $A^\mathsf{T}A\boldsymbol{x} = A^\mathsf{T}\boldsymbol{b}$ [23], [3], [20]. NR-SSOR is a symmetric version of NR-SOR. These methods can be implemented without explicitly forming $A^\mathsf{T}A$. See Appendix B for the algorithms of the methods.

According to [9], the iteration matrix H for

$$\left\{ \begin{array}{l} \text{Richardson-NR with } 0 < \omega < 2/\rho(A^{\mathsf{T}}A) \\ \text{Cimmino-NR with } 0 < \omega < 2/\rho(D^{-1/2}A^{\mathsf{T}}AD^{-1/2}) \\ \text{NR-SOR with } 0 < \omega < 2 \\ \text{NR-SSOR with } 0 < \omega < 2 \end{array} \right\} \tag{4.5}$$

is semi-convergent, where $\rho(C)$ is the spectral radius of C. Here, we assume that A has no zero columns. Hence, from Theorem 4.7, we obtain the following theorem which guarantees that these methods can serve as the inner iterations for BA-GMRES.

THEOREM 4.8. *BA-GMRES preconditioned by inner iterations* (4.5) *respectively determines a solution of* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *without breakdown for all* $\boldsymbol{b} \in \mathbf{R}^m$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$.

4.2. Spectrum of the matrix preconditioned by inner iterations. Next, we analyze the spectrum of the matrix preconditioned by ℓ inner iterations.

THEOREM 4.9. *Let* $r = \mathrm{rank}A$. *Assume that* H *is semi-convergent. Then, there exist* r *eigenvalues of* $B^{(\ell)}A$ *in a disk with center at* 1 *and radius* $\rho(H)^\ell < 1$, *and the remaining* $n - r$ *eigenvalues are zero.*

Proof. If ν is an eigenvalue of H, then from (4.3), $B^{(\ell)}A$ has an eigenvalue $\mu = 1 - \nu^\ell$. Assume that H is semi-convergent. Then, from Theorem 4.3, H has r eigenvalues such that $|\nu| < 1$ and $n - r$ eigenvalues such that $\nu = 1$. For $\nu = 1$, we have $\mu = 0$. For $|\nu| < 1$, we obtain $|\mu - 1| = |\nu|^\ell \le \rho(H)^\ell < 1$. □

This theorem shows that if H is semi-convergent, then $B^{(\ell)}A$ satisfies the condition for the spectrum in Theorem 3.5, and that the r nonzero eigenvalues of $B^{(\ell)}A$ approach 1 as ℓ increases.

We illustrate this observation for a test matrix

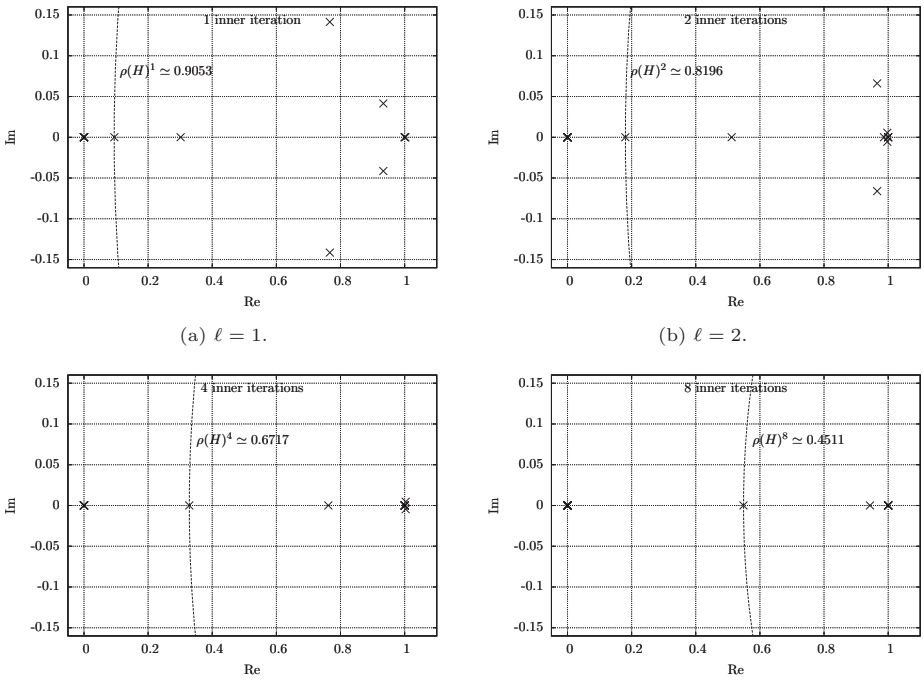
$$A = U \begin{bmatrix} 1 & 1 & & & & & 0 & \\ & & 0.9 & 0.9 & & & & \\ & & & & \ddots & \ddots & & \\ & & & & & 0.1 & 0.1 & \\ & & 0 & & & & & \end{bmatrix} V^{\mathsf{T}} \in \mathbf{R}^{100\times 20}, \tag{4.6}$$

where U and V are orthogonal matrices computed with the QR factorization of random matrices. Hence, A is rank-deficient. The computations were done using MATLAB 2011b. Figure 4.1 shows the spectrum of the preconditioned matrix $B^{(\ell)}A$ with the NR-SOR inner iterations for $\ell = 1$, 2, 4, and 8. The relaxation parameter was set to $\omega = 1$. Hence, the iteration matrix H is semi-convergent (4.5). The circles with radius $\rho(H)^\ell$ with center at 1 are also plotted. As the number of inner iterations ℓ increases, the eigenvalues of $B^{(\ell)}A$ approaches 1.

Next, we use a matrix arising in an application, called Maragal_3 [7] of size 1,690 × 860, number of nonzero elements 18,391, nonzero density 1.27%, and rank 613. Figure 4.2 shows the spectrum of the preconditioned matrix $B^{(k)}A$ with the NR-SOR inner iterations for $k = 1$, 2, 4, and 8. Here also, the eigenvalues tend to approach 1, as ℓ increases.

4.3. Convergence bound. Next, we give a convergence bound for BA-GMRES preconditioned by inner iterations.

THEOREM 4.10. *Let* $\boldsymbol{r}_k$ *be the* k*th residual of BA-GMRES preconditioned by* ℓ *inner iterations* $B^{(\ell)}$ (4.2) *and* T *be the Jordan basis of* $B^{(\ell)}A$. *Assume that* H *is*



(a) $\ell = 1$. (b) $\ell = 2$.

(c) $\ell = 4$. (d) $\ell = 8$.

FIG. 4.1. *Spectrum of the preconditioned matrix $B^{(\ell)}A$ with NR-SOR inner iterations for the matrix defined in* (4.6). *ℓ is the number of inner iterations.*

semi-convergent and $\boldsymbol{x}_0 \in \mathcal{R}(B^{(\ell)})$. Then, we have

$$\|B^{(\ell)}\boldsymbol{r}_k\|_2 \leq \kappa(T) \sum_{i=0}^{\tau(k,d)} \binom{k}{i} \rho(H)^{k\ell-i} \|B^{(\ell)}\boldsymbol{r}_0\|_2.$$

for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ and for all $\boldsymbol{b} \in \mathbf{R}^m$, where d is the size of the largest Jordan block of $B^{(\ell)}A$ corresponding to a nonzero eigenvalue of $B^{(\ell)}A$, and $\kappa(T) = \|T\|_2\|T^{-1}\|_2$.

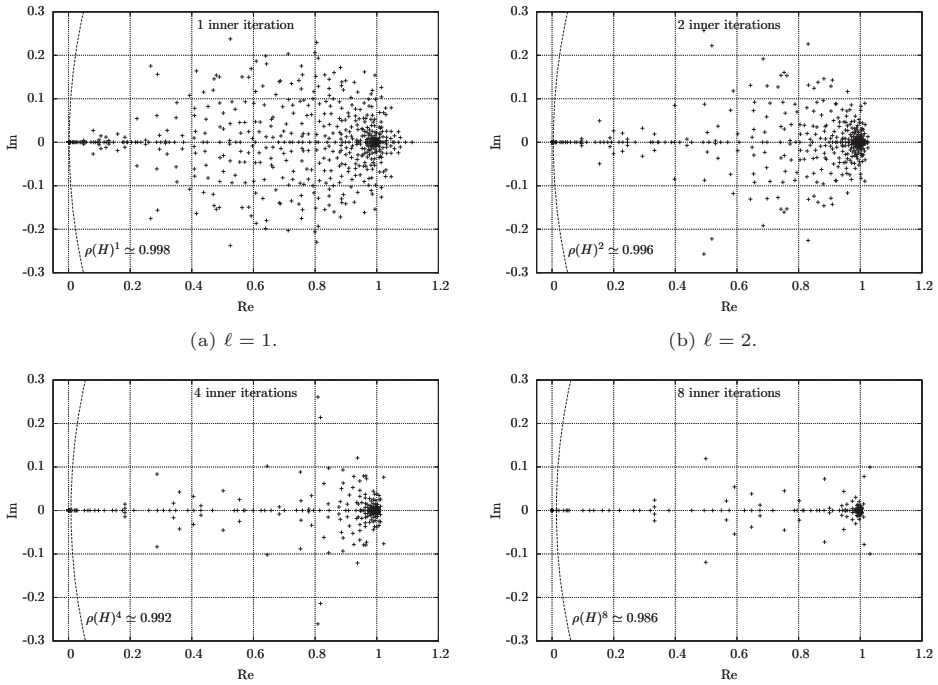
Proof. The theorem is a direct consequence of Theorems 3.5, 4.7, and 4.9. □

Note that the convergence of $\|B^{(\ell)}\boldsymbol{r}_k\|_2$ is not necessarily governed only by the eigenvalues of $B^{(\ell)}A$ when $\kappa(T)$ is large [12], [1], [27].

5. AB-GMRES method. Consider solving (1.3). AB-GMRES [14] applies GMRES to $\min_{\boldsymbol{u}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{u}\|_2$ with $\boldsymbol{x} = B\boldsymbol{u}$, and works in a smaller m-dimensional space than the n-dimensional space for BA-GMRES in the underdetermined case $m < n$. The algorithm of AB-GMRES with the modified Gram-Schmidt orthogonalization is given as follows.

ALGORITHM 5.1. *AB-GMRES method.*

1. *Let $\boldsymbol{x}_0$ be the initial approximate solution and $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.*
2. *$\beta := \|\boldsymbol{r}_0\|_2$, $\boldsymbol{v}_1 := \boldsymbol{r}_0/\beta$*
3. *For $k = 1, 2, \ldots$ until convergence, Do*
4. *$\quad \boldsymbol{w}_k := AB\boldsymbol{v}_k$*
5. *$\quad$ For $i = 1, 2, \ldots, k$, Do*
6. *$\quad\quad h_{i,k} := (\boldsymbol{w}_k, \boldsymbol{v}_i)$, $\boldsymbol{w}_k := \boldsymbol{w}_k - h_{i,k}\boldsymbol{v}_i$*



(a) $\ell = 1$. (b) $\ell = 2$.

(c) $\ell = 4$. (d) $\ell = 8$.

FIG. 4.2. *Spectrum of the preconditioned matrix $B^{(\ell)}A$ with NR-SOR inner iterations for Maragal_3. ℓ is the number of inner iterations.*

7. *EndDo*
8. $h_{k+1,k} := \|\boldsymbol{w}_k\|_2$, $\boldsymbol{v}_{k+1} := \boldsymbol{w}_k / h_{k+1,k}$
9. *EndDo*
10. $\boldsymbol{y}_k := \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta \boldsymbol{e}_1 - \bar{H}_k \boldsymbol{y}\|_2$, $\boldsymbol{x}_k := \boldsymbol{x}_0 + B\left[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\right] \boldsymbol{y}_k$

We say that AB-GMRES breaks down at some step k if $\dim AB\mathcal{K}_k(AB, \boldsymbol{r}_0) < \dim \mathcal{K}_k(AB, \boldsymbol{r}_0)$ or $\dim \mathcal{K}_k(AB, \boldsymbol{r}_0) < k$ (cf. Definition 2.1). We have the following.

THEOREM 5.2. *Assume $\mathcal{R}(B) = \mathcal{R}(A^\mathsf{T})$. Then, AB-GMRES determines the minimum-norm solution of $A\boldsymbol{x} = \boldsymbol{b}$ without breakdown for all $\boldsymbol{b} \in \mathcal{R}(A)$ and for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ if and only if $\mathcal{R}(A) \cap \mathcal{N}(B) = \{\boldsymbol{0}\}$.*

Proof. Let $\hat{\boldsymbol{u}}$ be a solution of $AB\boldsymbol{u} = \boldsymbol{b}$. Then $\hat{\boldsymbol{x}} = B\hat{\boldsymbol{u}} \in \mathcal{R}(A^\mathsf{T}) = \mathcal{N}(A)^\perp$ is a solution of $A\boldsymbol{x} = \boldsymbol{b}$. Any solution of $A\boldsymbol{x} = \boldsymbol{b}$ is given by $\boldsymbol{x} = \hat{\boldsymbol{x}} + \boldsymbol{t}$, $\boldsymbol{t} \in \mathcal{N}(A)$. Since $\hat{\boldsymbol{x}} \perp \mathcal{N}(A)$ and $\|\boldsymbol{x}\|_2{}^2 = \|\hat{\boldsymbol{x}}\|_2{}^2 + \|\boldsymbol{t}\|_2{}^2$, $\hat{\boldsymbol{x}}$ is the unique solution of $A\boldsymbol{x} = \boldsymbol{b}$, whose Euclidean-norm is minimum.

Next, substitute AB, $\boldsymbol{u}$, and $\boldsymbol{b}$ into $\tilde{A}$, $\tilde{\boldsymbol{x}}$, and $\tilde{\boldsymbol{b}}$, respectively, in Theorem 2.2. $\mathcal{R}(B) = \mathcal{R}(A^\mathsf{T})$ gives $\mathcal{R}(AB) = \mathcal{R}(AA^\mathsf{T}) = \mathcal{R}(A)$ and $\mathcal{N}(AB) = \mathcal{R}(B^\mathsf{T}A^\mathsf{T})^\perp = \mathcal{R}(B^\mathsf{T}B)^\perp = \mathcal{R}(B^\mathsf{T})^\perp = \mathcal{N}(B)$. Theorem 1.1 completes the proof. □

This theorem extends [14, Theorem 3.7] to the consistent case.

5.1. AB-GMRES preconditioned by stationary iterative methods as inner iterations. Similarly to obtaining Algorithm 4.1 from Algorithm 3.1, we obtain the algorithm of AB-GMRES preconditioned by inner iterations as follows.

ALGORITHM 5.3. *AB-GMRES method preconditioned by inner iterations.*

1. *Let* $\boldsymbol{x}_0$ *be the initial approximate solution and* $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.
2. $\beta := \|\boldsymbol{r}_0\|_2$, $\boldsymbol{v}_1 := \boldsymbol{r}_0/\beta$
3. *For* $k = 1, 2, \ldots$ *until convergence, Do*
4. *Apply* ℓ *steps of a stationary iterative method to* $AA^\mathsf{T}\boldsymbol{y} = \boldsymbol{v}_k$, $\boldsymbol{z} = A^\mathsf{T}\boldsymbol{y}$ *to obtain* $\boldsymbol{z}_k := B^{(\ell)}\boldsymbol{v}_k$.
5. $\boldsymbol{w}_k := A\boldsymbol{z}_k$
6. *For* $i = 1, 2, \ldots, k$, *Do*
7. $h_{i,k} := (\boldsymbol{w}_k, \boldsymbol{v}_i)$, $\boldsymbol{w}_k := \boldsymbol{w}_k - h_{i,k}\boldsymbol{v}_i$
8. *EndDo*
9. $h_{k+1,k} := \|\boldsymbol{w}_k\|_2$, $\boldsymbol{v}_{k+1} := \boldsymbol{w}_k/h_{k+1,k}$
10. *EndDo*
11. $\boldsymbol{y}_k := \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}\|_2$, $\boldsymbol{u}_k = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k]\,\boldsymbol{y}_k$
12. *Apply* ℓ *steps of a stationary iterative method to* $AA^\mathsf{T}\boldsymbol{y} = \boldsymbol{u}_k$, $\boldsymbol{z} = A^\mathsf{T}\boldsymbol{y}$ *to obtain* $\boldsymbol{z}' := B^{(\ell)}\boldsymbol{u}_k$
13. $\boldsymbol{x}_k := \boldsymbol{x}_0 + \boldsymbol{z}'$

The preconditioning matrix with ℓ inner iterations for AB-GMRES is given by

$$B^{(\ell)} = A^\mathsf{T} \sum_{i=0}^{\ell-1} \hat{H}^i \hat{M}^{-1}, \tag{5.1}$$

where $AA^\mathsf{T} = \hat{M} - \hat{N}$, $\hat{M}$ is nonsingular, and $\hat{H} = \hat{M}^{-1}\hat{N}$ is the inner-iteration matrix. From Theorem 5.2, we obtain the following lemmas.

LEMMA 5.4. *Let* $B^{(\ell)}$ *be given by* (5.1). *Assume that* $\hat{H}$ *is semi-convergent. Then,* $\mathrm{index}(AB^{(\ell)}) \le 1$ *for all* $\ell \ge 1$.

LEMMA 5.5. *Using the notations and the assumption of Lemma 5.4,* $\mathcal{R}(B^{(\ell)}) = \mathcal{R}\left(A^\mathsf{T}\right)$ *holds for all* $\ell \ge 1$.

Hence, we obtain the following theorems.

THEOREM 5.6. *Assume that* $\hat{H}$ *is semi-convergent. Then, AB-GMRES with the inner-iteration preconditioning* (5.1) *determines the minimum-norm solution of* $A\boldsymbol{x} = \boldsymbol{b}$ *without breakdown for all* $\boldsymbol{b} \in \mathcal{R}(A)$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$.

THEOREM 5.7. *AB-GMRES preconditioned by*

$$\left\{ \begin{array}{l} \textit{Richardson-NE inner iterations with } 0 < \omega < 2/\rho(AA^\mathsf{T}) \\ \textit{Cimmino-NE inner iterations with } 0 < \omega < 2/\rho(\hat{D}^{-1/2}AA^\mathsf{T}\hat{D}^{-1/2}) \\ \textit{NE-SOR inner iterations with } 0 < \omega < 2 \\ \textit{NE-SSOR inner iterations with } 0 < \omega < 2 \end{array} \right\}$$

respectively determines the minimum-norm solution of $A\boldsymbol{x} = \boldsymbol{b}$ *without breakdown for all* $\boldsymbol{b} \in \mathcal{R}(A)$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$, *where* $\hat{D} = \mathrm{diag}(AA^\mathsf{T})$.

The Cimmino-NE method is mathematically equivalent to JOR applied to $AA^\mathsf{T}\boldsymbol{u} = \boldsymbol{b}$ with $\boldsymbol{x} = A^\mathsf{T}\boldsymbol{u}$. The normal-error (NE-)SOR method is mathematically equivalent to SOR applied to $AA^\mathsf{T}\boldsymbol{u} = \boldsymbol{b}$ with $\boldsymbol{x} = A^\mathsf{T}\boldsymbol{u}$ [23], [3]. See Appendix B for the algorithms of the methods.

6. Numerical experiments. We compare the proposed methods BA- and AB-GMRES preconditioned by the NR- and NE-SOR inner iterations with previous methods in terms of the CPU time by numerical experiments on overdetermined and underdetermined problems, respectively. We omit results on the Richardson and Cimmino-type inner-iteration preconditioning since they were not as efficient as the SOR-type ones. However, the Richardson-type inner-iteration preconditioning can potentially

be useful for problems for which the quantity $\|\boldsymbol{a}_j\|_2$ cannot not be efficiently computed.[1] The Cimmino-type inner-iteration preconditioning can potentially be useful for parallel implementations.

The proposed methods require two preconditioning parameters: the number of inner iterations ℓ and the relaxation parameter ω. Since CPU time for the proposed method vary with the values of these parameters, it is desirable to determine the values automatically for any given problem. In order to determine these parameters, we perform the following procedure using the NR- and or NE-SOR iterations alone before starting the outer iterations [20]:

PROCEDURE 6.1. *Automatic parameter tuning.*

1. *Set* $\omega := 1$.
2. *Starting from* $\ell := 0$*, find the minimum* ℓ *that satisfies*

$$\|\boldsymbol{z}^{(\ell-1)} - \boldsymbol{z}^{(\ell)}\|_\infty \leq 10^{-1}\|\boldsymbol{z}^{(\ell)}\|_\infty.$$

3. *Find* ω *which first minimizes* $\|\boldsymbol{r}^{(\ell)}\|_2$*, searching* $\omega = 1.9, 1.8, \ldots, 0.1$ *in this order.*

In NR-SOR, the approximate residual $\boldsymbol{r}^{(\ell)}$ is given in the algorithm. In NE-SOR, $\boldsymbol{r}^{(\ell)} = \boldsymbol{c} - A\boldsymbol{z}^{(\ell)}$ is explicitly evaluated since it is not given in the algorithm.

Table 6.1 gives information on the test matrices from [7], including the number of rows m, the number of columns n, the number of nonzero elements nnz, and the density of the nonzero elements, the rank, and the condition number $\kappa(A)$ which is the ratio of the largest singular value of the matrices to the smallest positive one. These matrices were appropriately transposed to form overdetermined and underdetermined problems, i.e., $m > n$ and $m < n$, respectively. The name of a matrix is denoted by using T if the matrix is transposed. Table 6.1 shows the effective size of the matrices after removing all zero columns and zero rows. (If A has a zero column (row), then the diagonal scaling $D = \mathrm{diag}(A^\mathsf{T}A)$ ($\hat{D} = \mathrm{diag}(AA^\mathsf{T})$) and the splitting matrix M of NR-SOR (NE-SOR) are singular.) The condition number was computed by dividing the largest singular value by the smallest nonzero one, where the singular values were computed by using the MATLAB function `svd` and the number of nonzero singular values was determined by using the MATLAB function `spnrank` [11]. (The rank and condition number of Maragal_8 could not be computed on our computer due to insufficient memory.)

For all the CPU times, an average was taken over ten measurements. The initial solution for the inner and outer iterations was set to zero. No restarts were used for the GMRES-type methods.

All computations were done on a PC workstation with an Intel Xeon X5492 3.4 GHz CPU, 16 GB RAM, Scientific Linux version 6.4, and double precision floating-

TABLE 6.1
Information of the matrices.

Name	m	n	nnz	density [%]	rank	$\kappa(A)$
landmark	71,952	2,673	1,146,848	0.596	2,671	$1.02{\cdot}10^8$
lp_cre_a	3,428	7,248	18,168	0.073	3,423	$2.11{\cdot}10^4$
lp_dfl001	6,071	12,230	35,632	0.048	6,058	$3.49{\cdot}10^2$
Maragal_6	21,251	10,144	537,694	0.249	8,331	$2.91{\cdot}10^6$
Maragal_7	46,845	26,525	1,200,537	0.096	20,843	$8.98{\cdot}10^6$
Maragal_8	33,093	60,845	1,308,415	0.065	–	–

[1] Private communication with Dr. Wei Xu of Tongji University.

point arithmetic. All programs for the iterative methods in our tests were coded in Fortran 95 and compiled by Intel Fortran version 13.1.0. For reproducibility, the compiler options we used were `-fp-model precise -fimf-arch-consistency=true`, i.e., IEEE 754 standard binary64 with no compiler optimization affecting accuracy. For the direct methods, we used Matlab 2013b and SuiteSparseQR version 1.3.1 [8].

6.1. Overdetermined problems. We first present numerical experiment results on overdetermined problems ($m > n$). The proposed BA-GMRES preconditioned by the NR-SOR inner iterations was compared with previous methods: the preconditioned CGLS, LSMR, and BiCGSTAB-LS methods [16], [10], [28]. We call the BiCGSTAB method applied to the normal equations the BiCGSTAB-LS method. The initial dual (or so-called shadow) residual for BiCGSTAB-LS was set to the initial primal residual. Note that these methods work in n-dimensional space.

We used the diagonal scaling $D = \mathrm{diag}(A^{\mathsf{T}}A)$ for preconditioning the CGLS, LSMR, BA-GMRES, and BiCGSTAB-LS methods, i.e., the conjugate gradient (CG) [16] and MINRES [22] methods were applied to $D^{-\frac{1}{2}}A^{\mathsf{T}}AD^{-\frac{1}{2}}\boldsymbol{u} = D^{-\frac{1}{2}}A^{\mathsf{T}}\boldsymbol{b}$, $\boldsymbol{x} = D^{-\frac{1}{2}}\boldsymbol{u}$ and GMRES and BiCGSTAB were applied to $D^{-1}A^{\mathsf{T}}A\boldsymbol{x} = D^{-1}A^{\mathsf{T}}\boldsymbol{b}$, which corresponds to $B = D^{-1}A^{\mathsf{T}}$.

On the other hand, we used the NR-SOR inner iterations for preconditioning BA-GMRES and BiCGSTAB-LS, i.e., GMRES and BiCGSTAB were applied to $B^{(\ell)}A\boldsymbol{x} = B^{(\ell)}\boldsymbol{b}$, where $B^{(\ell)}$ is given by (4.2) and (4.4). We used the NE-SSOR inner iterations for preconditioning CGLS and LSMR to obtain a symmetric preconditioner. See Appendix C for their algorithms. BiCGSTAB-LS preconditioned by inner iterations can be implemented similarly to BA-GMRES preconditioned by inner iterations.

The elements of $\boldsymbol{b}$ were randomly generated using the Fortran built-in subroutine `random_number`. Therefore, the test problems were not necessarily consistent, i.e., $\boldsymbol{b}$ may not be in $\mathcal{R}(A)$.

In exact arithmetic, the CGLS, LSMR, BA-GMRES methods with diagonal scaling and preconditioned by the NR-(S)SOR inner iterations with $0 < \omega < 2$ determine a solution of $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ and for all $\boldsymbol{b} \in \mathbf{R}^m$. However, this is not necessarily the case for the BiCGSTAB-type methods.

The stopping criterion for the kth (outer) iteration was

$$\left\|A^{\mathsf{T}}(\boldsymbol{b} - A\boldsymbol{x}_k)\right\|_2 < 10^{-8}\left\|A^{\mathsf{T}}\boldsymbol{b}\right\|_2. \tag{6.1}$$

This means that we explicitly compute the residual 2-norm $A^{\mathsf{T}}(\boldsymbol{b}-A\boldsymbol{x}_k)$ for the normal equations (1.2) from $\boldsymbol{x}_k$ at each iteration. In the numerical experiments, the CPU time for checking (6.1) was excluded from the total CPU time. The left-hand side of (6.1) can converge to zero because of the equivalence between (1.1) and (1.2).

Table 6.2 gives the CPU time in seconds taken by direct methods, namely, the "backslash" solver in MATLAB and SuiteSparseQR [8], and the resulting relative residual 2-norm $\|A^{\mathsf{T}}\boldsymbol{r}\|_2/\|A^{\mathsf{T}}\boldsymbol{b}\|_2$. The ‡ indicates that the direct method did not achieve the criterion (6.1).

Table 6.3 gives the CPU time in seconds for the iterative methods to achieve the stopping criterion (6.1) for each problem. Diag stands for the diagonal scaling. The first column in each cell gives the number of (outer) iterations. The second column gives the total CPU time including the time to set up the preconditioning matrix for the diagonal scaling and the parameter tuning time for the NR-SOR inner-iteration preconditioning for BA-GMRES. The third column gives the number of inner iterations ℓ which was optimal in terms of the CPU time except for BA-GMRES

TABLE 6.2
Results of the direct solvers for overdetermined problems.

	landmark		lp_cre_aT		lp_dfl001T	
	Time	Rel. res.	Time	Rel. res.	Time	Rel. res.
Backslash	0.405	$5.18{\cdot}10^{-13}$	0.017	$1.08{\cdot}10^{-14}$	1.214	$2.14{\cdot}10^{-15}$
SuiteSparseQR	11.46	$5.18{\cdot}10^{-13}$	0.003	$1.79{\cdot}10^{-15}$	1.344	$1.86{\cdot}10^{-15}$

	Maragal_6		Maragal_7		Maragal_8T	
	Time	Rel. res.	Time	Rel. res.	Time	Rel. res.
Backslash	384.4	‡$1.26{\cdot}10^{-1}$	329.1	‡$7.98{\cdot}10^{-2}$	72.22	$1.28{\cdot}10^{-9}$
SuiteSparseQR	256.0	‡$1.35{\cdot}10^{-1}$	467.3	‡$8.22{\cdot}10^{-2}$	67.08	$9.72{\cdot}10^{-10}$

TABLE 6.3
Results for overdetermined problems.

Solver	Precon.	landmark				lp_cre_aT				lp_dfl001T			
		Iter	Time	ℓ	ω	Iter	Time	ℓ	ω	Iter	Time	ℓ	ω
CGLS	diag	311	2.10	-	-	2,288	0.44	-	-	411	0.22	-	-
	NE-SSOR	73	1.00	1	1.3	894	0.29	1	0.6	154	0.13	1	1.1
LSMR	diag	290	2.10	-	-	2,041	0.38	-	-	399	0.20	-	-
	NE-SSOR	71	*0.99	1	1.3	739	*0.23	1	0.8	151	0.13	1	1.1
BA-GMRES	diag	296	2.15	-	-	1,190	6.28	-	-	388	1.31	-	-
	NR-SOR	143	1.86	2	1.0	1,931	16.18	10	1.7	57	*0.09	5	1.4
optimal	NR-SOR	143	1.82	2	1.0	363	0.88	11	0.6	21	0.07	13	1.8
BiCGSTAB	diag	†	†	-	-	†	†	-	-	328	0.25	-	-
-LS	NR-SOR	31	4.26	15	1.1	1,006	0.30	1	0.7	30	*0.09	5	1.8

Solver	Precon.	Maragal_6				Maragal_7				Maragal_8T			
		Iter	Time	ℓ	ω	Iter	Time	ℓ	ω	Iter	Time	ℓ	ω
CGLS	diag	†	†	-	-	15,654	121.5	-	-	†	†	-	-
	NE-SSOR	†	†			2,088	36.29	1	1.0	†	†		
LSMR	diag	†	†	-	-	9,381	72.32	-	-	†	†	-	-
	NE-SSOR	9,940	65.46	1	0.9	1028	31.64	2	1.2	†	†		
BA-GMRES	diag	2,708	101.8	-	-	2,491	230.6	-	-	6,834	2,041	-	-
	NR-SOR	515	*8.70	5	1.4	334	*16.21	7	1.5	1,543	*147.1	4	1.3
optimal	NR-SOR	430	8.50	7	1.4	359	16.08	6	1.4	952	97.01	10	1.3
BiCGSTAB	diag	†	†	-	-	†	†	-	-	†	†	-	-
-LS	NR-SOR	†	†			†	†			†	†		

preconditioned by the NR-SOR inner iterations. The fourth column gives the value of the relaxation parameter ω which was optimal among $0.1, 0.2, \ldots, 1.9$ in terms of the CPU time except for the proposed BA-GMRES (NR-SOR). For BA-GMRES (NR-SOR), ℓ and ω were determined using Procedure 6.1. Hence, these comparisons are advantageous for the optimized preconditioners compared to the proposed method. The * indicates the fastest method in terms of CPU time for each problem. The † indicates the case where the method did not satisfy the criterion (6.1) within iterations equal to the size n of the problem.

For the last four problems, the proposed BA-GMRES preconditioned by NR-SOR inner iterations with automatically tuned parameters was the fastest. The CPU time for tuning the parameters was marginal compared to the total CPU time. For example, the tuning time for Maragal_6 was 0.07 seconds out of the total CPU time of 8.70 seconds. The total CPU time for the proposed method with automatically tuned parameters was close to that with optimal parameters as reported in [20], except for the case lp_cre_aT. This indicates that fast convergence of inner iterations may not necessarily give fast convergence of outer iterations (see step 3 in Procedure 6.1). The proposed method gave much more accurate solutions than both the direct

solvers for Maragal_6 and Maragal_7. LSMR preconditioned by the NE-SSOR inner iterations broke down with many combinations of the NE-SSOR parameters trying to take the square root of a negative value corresponding to a theoretically nonnegative inner product, which was caused by rounding error (see step 8 in Algorithm C.2). This is not the case for LSMR with diagonal scaling since the preconditioning matrix can be applied symmetrically as $D^{-\frac{1}{2}}A^{\mathsf{T}}AD^{-\frac{1}{2}}$. The BiCGSTAB-LS methods did not converge for Maragal_6–8 for all the parameters that were tried. CGLS and LSMR with reorthogonalization [14], [10] combined with these preconditioners were also tested, but they were slow to converge.

Figure 6.1 shows the relative residual $\|A^{\mathsf{T}}\boldsymbol{r}_k\|_2/\|A^{\mathsf{T}}\boldsymbol{b}\|_2$ versus the CPU time for Maragal_6 for each method. The parameter values of the NE-SSOR inner-iteration preconditioning for CGLS were $\ell = 1$ and $\omega = 1.0$, and we used the preconditioning parameters given in Table 6.3 for the other methods. The convergence curve for the CGLS-type method are quite oscillatory as explained in [14]. The LSMR and BA-GMRES-type methods yield smoother convergence curves and the latter gives quicker convergence.

6.2. Underdetermined problems. Next, experiments were done for underdetermined problems ($m < n$). The proposed AB-GMRES preconditioned by the NE-SOR inner iterations was compared with previous methods: the preconditioned CGNE [5], MRNE, and BiCGSTAB-NE methods. We call the MINRES method applied to the normal equations of the second kind (1.4) the MRNE method. We call BiCGSTAB applied to the normal equations of the second kind (1.4) the BiCGSTAB-NE method. The initial dual (or so-called shadow) residual for BiCGSTAB-NE was set to the initial primal residual. Note that these methods work in m-dimensional space.

The preconditioned MRNE method can be implemented based on the preconditioned CG method as follows. Let $M \in \mathbf{R}^{m\times m}$ be symmetric and positive definite. Then, $AA^{\mathsf{T}}M$ is symmetric with respect to the M^{-1}-inner product, where the M^{-1}-inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is $(\boldsymbol{a},\boldsymbol{b})_{M^{-1}} = (M^{-1}\boldsymbol{a},\boldsymbol{b})$. Hence, MINRES applied to $AA^{\mathsf{T}}M^{-1}\boldsymbol{u} = \boldsymbol{b}$, $\boldsymbol{x} = A^{\mathsf{T}}M^{-1}\boldsymbol{u}$ with the M^{-1}-inner product is equivalent to CG applied to $A^{\mathsf{T}}M^{-1}A\boldsymbol{x} = A^{\mathsf{T}}M^{-1}\boldsymbol{b}$.

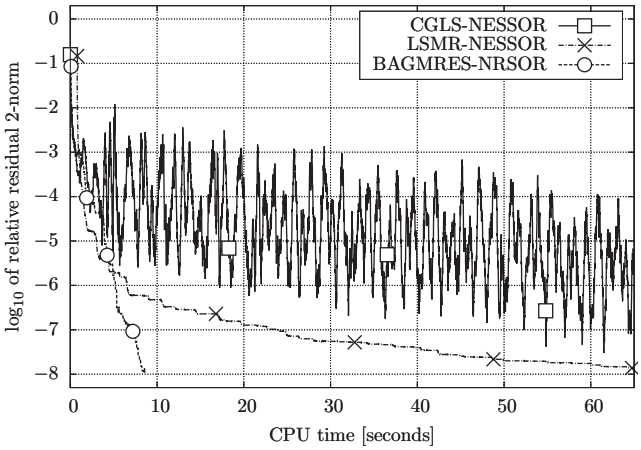


FIG. 6.1. *Relative residual vs. CPU time for Maragal_6.*

The diagonal scaling $D = \mathrm{diag}(AA^\mathsf{T})$ was applied to the CGNE, MRNE, AB-GMRES, and BiCGSTAB-NE methods, i.e., the conjugate gradient (CG) [16] and MINRES [22] methods were applied to $D^{-\frac{1}{2}}AA^\mathsf{T}D^{-\frac{1}{2}}\boldsymbol{u} = D^{-\frac{1}{2}}\boldsymbol{b}$, $\boldsymbol{x} = A^\mathsf{T}D^{-\frac{1}{2}}\boldsymbol{u}$ and GMRES and BiCGSTAB were applied to $AA^\mathsf{T}D^{-1}\boldsymbol{u} = \boldsymbol{b}$ $\boldsymbol{x} = A^\mathsf{T}D^{-1}\boldsymbol{u}$ with $B = A^\mathsf{T}D^{-1}$. Since CGNE and MRNE require symmetric preconditioning, the NE-SSOR inner-iteration preconditioning was applied to CGNE and MRNE. The NE-SOR inner-iteration preconditioning was applied to AB-GMRES and BiCGSTAB-NE. See Appendix C for their algorithms. BiCGSTAB-NE preconditioned by inner iterations can be implemented similarly to AB-GMRES preconditioned by inner iterations.

The vector $\boldsymbol{b}$ was given by $\boldsymbol{b} = A\,[1, 1, \ldots, 1]$. Therefore, the test problems were consistent, i.e., $\boldsymbol{b} \in \mathcal{R}(A)$.

In exact arithmetic, the CGNE, MRNE, AB-GMRES methods with diagonal scaling and preconditioned by the NE-(S)SOR inner iterations with $0 < \omega < 2$ determine the minimum-norm solution of $A\boldsymbol{x} = \boldsymbol{b}$ for all $\boldsymbol{x}_0 \in \mathbf{R}^n$ and for all $\boldsymbol{b} \in \mathcal{R}(A)$. However, this is not necessarily the case for the BiCGSTAB-type methods.

The stopping criterion for the kth (outer) iteration was

$$\|\boldsymbol{b} - A\boldsymbol{x}_k\|_2 < 10^{-8}\,\|\boldsymbol{b}\|_2\,. \tag{6.2}$$

Table 6.4 gives the CPU time in seconds taken by the direct method and the resulting relative residual 2-norm $\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$, similar to Table 6.2. (A sparse direct method for the minimum-norm solution is not implemented in Matlab.) The ‡ indicates that the direct method did not achieve the criterion (6.2).

Table 6.5 gives the CPU time in seconds for the iterative methods to achieve the stopping criterion (6.2) for each problem, similar to Table 6.3. The proposed method with automatically tuned parameters was the fastest for Maragal_6T and Maragal_8. The proposed method gave more accurate solutions than the direct solver landmarkT and Maragal_8. The BiCGSTAB-NE did not converge for the six problems for all the parameters that we tried. Although the residual $\|\boldsymbol{r}^{(k)}\|_2 = \|\boldsymbol{c} - A\boldsymbol{x}^{(k)}\|_2$ for the NE-SOR inner iterations was explicitly evaluated in Procedure 6.1, the CPU time for tuning the parameters was again marginal compared to the total CPU time. For example, the CPU time for tuning the parameters for Maragal_6T was 0.10 seconds out of the total CPU time of 7.81 seconds. The total CPU time for the proposed method with automatically tuned parameters was close to the optimal one except for lp_cre_a. CGNE and MRNE with reorthogonalization [14] combined with these preconditioners were also tested, but they were slow to converge.

Figure 6.2 shows the relative residual $\|\boldsymbol{r}_k\|_2/\|\boldsymbol{b}\|_2$ versus the CPU time for Maragal_6T for each method. The convergence curve for CGNE with NE-SSOR is oscillatory and slow to converge. The convergence curve for MRNE with NE-SSOR is smoother than the one for CGNE with NE-SSOR but slower to converge than AB-GMRES with NE-SOR.

TABLE 6.4
Results of the direct solver for underdetermined problems.

	landmarkT		lp_cre_a		lp_dfl001	
	Time	Rel. res.	Time	Rel. res.	Time	Rel. res.
SuiteSparseQR	0.556	$\ddagger 1.09{\cdot}10^{155}$	0.003	$0.003{\cdot}10^{-17}$	0.011	$2.34{\cdot}10^{-16}$

	Maragal_6T		Maragal_7T		Maragal_8	
	Time	Rel. res.	Time	Rel. res.	Time	Rel. res.
SuiteSparseQR	26.30	$1.73{\cdot}10^{-9}$	66.50	$1.12{\cdot}10^{-11}$	3,635	$\ddagger 1.25{\cdot}10^{-8}$

TABLE 6.5
Results for underdetermined problems.

Solver	Precon.	landmarkT				lp_cre_a				lp_dfl001			
		Iter	Time	ℓ	ω	Iter	Time	ℓ	ω	Iter	Time	ℓ	ω
CGNE	diag	185	1.26	-	-	2,122	0.39	-	-	406	0.19	-	-
	NE-SSOR	56	0.69	1	1.1	823	0.29	1	0.6	154	0.13	1	1.0
MRNE	diag	160	1.08	-	-	2,347	0.36	-	-	398	0.19	-	-
	NE-SSOR	49	*0.60	1	1.1	781	*0.23	1	0.6	154	*0.12	1	1.0
AB-GMRES	diag	158	1.12	-	-	1,019	4.61	-	-	384	1.27	-	-
	NE-SOR	26	2.20	11	0.7	1,502	10.68	6	1.5	92	0.18	4	0.9
	optimal	77	0.97	2	0.8	315	0.73	12	0.5	26	0.08	10	1.8
BiCGSTAB	diag	†	†	-	-	†	†	-	-	†	†	-	-
-NE	NE-SOR	†	†			†	†			†	†		

Solver	Precon.	Maragal_6T				Maragal_7T				Maragal_8			
		Iter	Time	ℓ	ω	Iter	Time	ℓ	ω	Iter	Time	ℓ	ω
CGNE	diag	†	†	-	-	12,916	100.4	-	-	†	†	-	-
	NE-SSOR	8,996	50.38	1	1.2	1,975	33.62	1	1.1	†	†		
MRNE	diag	†	†	-	-	6,449	50.93	-	-	†	†	-	-
	NE-SSOR	6,108	33.90	1	1.0	913	*15.81	1	0.9	40,659	829.7	1	1.1
AB-GMRES	diag	2,808	109.0	-	-	2,637	262.2	-	-	†	†	-	-
	NE-SOR	459	*7.81	5	1.3	429	16.85	4	1.3	1,371	*140.4	7	1.5
	optimal	402	7.48	6	1.4	303	13.77	6	1.5	990	101.6	10	1.3
BiCGSTAB	diag	†	†	-	-	†	†	-	-	†	†	-	-
-NE	NE-SOR	†	†			†	†			†	†		

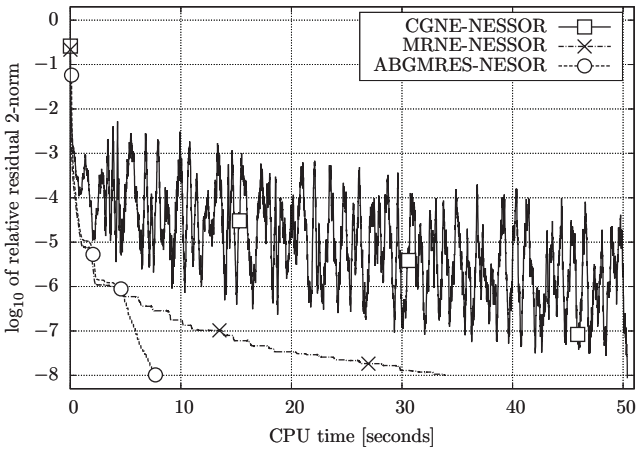


FIG. 6.2. *Relative residual vs. CPU time for Maragal_6T.*

7. Conclusions. We considered applying stationary inner-iteration preconditioning to GMRES methods for least squares problems and gave a general convergence theory for the methods. Theoretical justifications for the convergence were given also for specific inner-iteration methods like NR-SOR. We have reinforced the previous theory particularly for the rank-deficient case. The spectrum of the preconditioned matrix was analyzed and characterized using the spectral radius of the inner-iteration matrix. Based on this, a convergence bound was obtained for the proposed methods. Finally, numerical experiments on rank-deficient overdetermined and consistent underdetermined problems showed that the proposed methods BA-GMRES and AB-GMRES preconditioned by the NR- and NE-SOR inner iterations, respectively, are

more robust and efficient compared to previous methods for some problems.

Appendix A. Breakdown of GMRES. The generalized minimal residual (GMRES) method applied to the system of linear equations

$$A\boldsymbol{x} = \boldsymbol{b}, \quad A \in \mathbf{R}^{n\times n}, \quad \boldsymbol{b} \in \mathbf{R}^n,$$

determines the kth approximate solution $\boldsymbol{x}_k$ in the subspace $\boldsymbol{x}_0+\mathcal{K}_k(A, \boldsymbol{r}_0)$ minimizing $\|\boldsymbol{r}_k\|_2$, where $\boldsymbol{x}_0$ is the initial approximate solution, $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k$ is the residual at the kth step, and

$$\mathcal{K}_k(A, \boldsymbol{r}_0) = \mathrm{span}\{\boldsymbol{r}_0, A\boldsymbol{r}_0, \ldots, A^{k-1}\boldsymbol{r}_0\},$$

is the Krylov subspace of order k.

The convergence of GMRES including the singular case was analyzed in [4]. There, non-breakdown of GMRES was defined as follows.

DEFINITION A.1. *We say that GMRES does not break down at the kth step if*

$$\dim A\mathcal{K}_k(A, \boldsymbol{r}_0) = k.$$

This definition was intended to focus on essential breakdown of the method, as opposed to breakdown associated with any specific implementation of GMRES.

For instance, the algorithm of GMRES with Gram-Schmidt orthogonalization (GMRES-GS) is given as follows. (Note that, in exact arithmetic, GMRES with Gram-Schmidt orthogonalization is equivalent to GMRES with modified Gram-Schmidt orthogonalization (GMRES-mGS).)

ALGORITHM A.2. *GMRES method with Gram-Schmidt orthogonalization (GMRES-GS).*

1. Let $\boldsymbol{x}_0$ *be the initial approximate solution.*
2. $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$, $\beta := \|\boldsymbol{r}_0\|_2$, $\boldsymbol{v}_1 := \boldsymbol{r}_0/\beta$
3. For $k = 1, 2, \ldots$ *until convergence, Do*
4. $h_{i,k} := (\boldsymbol{v}_i, A\boldsymbol{v}_k), \quad i = 1, 2, \ldots, k$
5. $\boldsymbol{w}_k := A\boldsymbol{v}_k - \sum_{i=1}^{k} h_{i,k}\boldsymbol{v}_i$
6. $h_{k+1,k} := \|\boldsymbol{w}_k\|_2$. *If* $h_{k+1,k} = 0$, *then go to line 9.*
7. $\boldsymbol{v}_{k+1} := \boldsymbol{w}_k/h_{k+1,k}$
8. EndDo
9. $\boldsymbol{y}_k := \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}\|_2$, $\boldsymbol{x}_k := \boldsymbol{x}_0 + V_k\boldsymbol{y}_k$

Here, $\bar{H}_k = \{h_{i,j}\} \in \mathbf{R}^{(k+1)\times k}$ and $V_k = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k]$, where $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are orthonormal. Let $H_k = \{h_{ij}\} \in \mathbf{R}^{k\times k}$. Then, we have $AV_k = V_kH_k + \boldsymbol{w}_k\boldsymbol{e}_k^{\mathsf{T}}$.

The breakdown of GMRES-GS was defined as follows.

DEFINITION A.3 ([24], [23], [13]). *GMRES-GS is said to break down at the kth step when* $h_{k+1,k} = 0$.

This definition seems different from Definitions 2.1 or A.1.

The convergence condition for GMRES-GS including the singular case was given as follows.

THEOREM A.4 ([13, Theorems 2.8]). *GMRES-GS determines a solution of* $A\boldsymbol{x} = \boldsymbol{b}$ *for all* $\boldsymbol{b} \in \mathcal{R}(A)$ *and for all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *if and only if* $\mathcal{R}(A) \cap \mathcal{N}(A) = \{\boldsymbol{0}\}$.

Based on Theorem A.4, convergence conditions for AB- and BA-GMRES with the modified Gram-Schmidt orthogonalization were given in [14, Theorem 3.7, Corollary 3.8, Theorem 3.18, Corollary 3.19].

Now, the following question arises:

Are Definitions A.1, 2.1, and A.3 equivalent for GMRES-GS or not?
We consider this question in the following. Remark that trivially $\dim A\mathcal{K}_k \leq \dim\mathcal{K}_k \leq k$ holds for each k, where $\mathcal{K}_k = \mathcal{K}_k(A, \boldsymbol{r}_0)$. Therefore, Definitions 2.1 and A.1 are equivalent.

First, we use Definition 2.1. Assume that GMRES does not breakdown until the $(k-1)$st step. Then, $\dim A\mathcal{K}_{k-1} = \dim\mathcal{K}_{k-1} = k-1$. Thus, there exists V_{k-1}. Then, consider the two cases: $\dim A\mathcal{K}_k < \dim\mathcal{K}_k$ and $\dim\mathcal{K}_k < k$.

Case i. Assume that GMRES breaks down at the kth step due to $\dim A\mathcal{K}_k < \dim\mathcal{K}_k$. Then, $\dim\mathcal{K}_k = k$. ($\because$ Assume $\dim\mathcal{K}_k \leq k-1$. Then, $\dim A\mathcal{K}_k \leq k-2$. However, this contradicts with the assumption $\dim A\mathcal{K}_k = k-1$. Hence, $\dim\mathcal{K}_k = k$.) Moreover, we have

$$\begin{aligned} \dim A\mathcal{K}_k < \dim\mathcal{K}_k = k \quad &\Longleftrightarrow \quad \mathrm{rank}AV_k < \mathrm{rank}V_k = k, \\ &\Longrightarrow \quad \boldsymbol{v}_i \neq \boldsymbol{0}, \quad i = 1, 2, \ldots, k, \end{aligned}$$

$h_{i+1,i} = \|\boldsymbol{w}_i\|_2 \neq 0$ for $i = 1, 2, \ldots, k-1$ in Algorithm A.2, and

$$\begin{aligned} \mathrm{rank}AV_k &= \mathrm{rank}\left([V_k, \boldsymbol{w}_k]\begin{bmatrix} H_k \\ \boldsymbol{e}_k{}^{\mathsf{T}} \end{bmatrix}\right) = \dim\mathcal{R}\left(\begin{bmatrix} H_k{}^{\mathsf{T}}, \boldsymbol{e}_k \end{bmatrix}\begin{bmatrix} V_k{}^{\mathsf{T}} \\ \boldsymbol{w}_k{}^{\mathsf{T}} \end{bmatrix}\right) \\ &= \dim\mathcal{R}\left(\begin{bmatrix} H_k{}^{\mathsf{T}}, \boldsymbol{e}_k \end{bmatrix}\begin{bmatrix} V_k{}^{\mathsf{T}} \\ \boldsymbol{w}_k{}^{\mathsf{T}} \end{bmatrix}[V_k, \boldsymbol{w}_k]\right) = \mathrm{rank}\left(\begin{bmatrix} H_k \\ h_{k+1,k}{}^2\boldsymbol{e}_k{}^{\mathsf{T}} \end{bmatrix}\right) < k, \end{aligned}$$

where $h_{k+1,k} = \|\boldsymbol{w}_k\|_2$ and $\boldsymbol{w}_k = A\boldsymbol{v}_k - \sum_{i=1}^{k} h_{i,k}\boldsymbol{v}_i$. Hence, $h_{k+1,k} = 0$, since $h_{k+1,k} \neq 0$ implies $\mathrm{rank}\left(\begin{bmatrix} H_k{}^{\mathsf{T}}, h_{k+1,k}{}^2\boldsymbol{e}_k \end{bmatrix}\right) = k$. This is consistent with Definition A.3. Therefore, $\dim A\mathcal{K}_k < \dim\mathcal{K}_k \Longrightarrow \dim\mathcal{K}_k = k$ and $h_{k+1,k} = 0$. Note also that $\mathrm{rank}H_k = k-1$.

Moreover, since H_k is singular, the problem

$$\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_k} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 = \min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - H_k\boldsymbol{y}\|_2 \tag{A.1}$$

does not have a unique solution. ∎

Case ii. Assume that GMRES breaks down at the kth step due to $\dim\mathcal{K}_k < k$. Then, $\boldsymbol{w}_{k-1} = A\boldsymbol{v}_{k-1} - \sum_{i=1}^{k-1} h_{i,k-1}\boldsymbol{v}_i$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k-1}$ are linearly dependent, i.e., there exists $\boldsymbol{c} = [c_1, c_2, \ldots, c_k]^{\mathsf{T}} \neq \boldsymbol{0}$ such that $[V_{k-1}, \boldsymbol{w}_{k-1}]\,\boldsymbol{c} = \boldsymbol{0}$, or $[V_{k-1}, \boldsymbol{w}_{k-1}]^{\mathsf{T}}[V_{k-1}, \boldsymbol{w}_{k-1}]\,\boldsymbol{c} = \boldsymbol{0}$. This gives $c_i = 0$, $i = 1, 2, \ldots k-1$ and $c_k \neq 0$. Hence, $h_{k,k-1} = 0$. Note also that $\dim\mathcal{K}_{k-1} = \mathrm{rank}V_{k-1} = k-1$. Here, $\boldsymbol{v}_i \neq \boldsymbol{0}, i = 1, 2, \ldots, k-1$ and $h_{i+1,i} \neq 0$, $i = 1, \ldots, k-2$ in Algorithm A.2.

Moreover, we have $AV_{k-1} = V_{k-1}H_{k-1}$ and

$$\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_{k-1}} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 = \min_{\boldsymbol{y}\in\mathbf{R}^{k-1}} \|\beta\boldsymbol{e}_1 - H_{k-1}\boldsymbol{y}\|_2 = 0,$$

since the assumption $\dim A\mathcal{K}_{k-1} = k-1$ implies $k-1 = \mathrm{rank}(AV_{k-1}) = \mathrm{rank}H_{k-1}$, i.e., H_{k-1} is nonsingular. Finally, if $\boldsymbol{y}_{k-1} = \arg\min_{\boldsymbol{y}\in\mathbf{R}^{k-1}} \|\beta\boldsymbol{e}_1 - H_{k-1}\boldsymbol{y}\|_2$, then $\boldsymbol{x}_{k-1} = \boldsymbol{x}_0 + V_{k-1}\boldsymbol{y}_{k-1}$ is a solution of $A\boldsymbol{x} = \boldsymbol{b}$, i.e., GMRES determines a solution of $A\boldsymbol{x} = \boldsymbol{b}$ at the $(k-1)$st step. ∎

Conversely, we use Definition A.3. Assume that GMRES does not break down until the $(k-1)$st step, i.e., $h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-1$, and breaks down at the kth step due to $h_{k+1,k} = 0$. Then, we have $AV_k = V_k H_k$. Consider the two cases: $\mathrm{rank}H_k = k-1$ and k.

Case i'. Assume $\mathrm{rank}H_k = k-1$. Then, since $h_{i+1,i} \neq 0$ for $i = 1, 2, \ldots, k-1$,

$$\mathrm{rank}H_k = \dim\mathcal{R}\left(H_k{}^{\mathsf{T}} V_k{}^{\mathsf{T}} V_k\right) = \mathrm{rank}V_k H_k = \mathrm{rank}AV_k < \mathrm{rank}V_k = k.$$

This inequality is equivalent to $\dim A\mathcal{K}_k < \dim\mathcal{K}_k$. Note also that $\dim A\mathcal{K}_{k-1} = k-1$ holds, since $\dim A\mathcal{K}_{k-1} = \mathrm{rank}(AV_{k-1}) = \mathrm{rank}(V_k \bar{H}_{k-1}) = k-1$ for $h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-1$. ∎

Case ii'. Assume $\mathrm{rank}H_k = k$, i.e., H_k is nonsingular. Then, we have

$$\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_k} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 = \min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - H_k\boldsymbol{y}\|_2 = 0,$$

If $\boldsymbol{y}_k = \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - H_k\boldsymbol{y}\|_2$, then $\boldsymbol{x}_k = \boldsymbol{x}_0 + V_k\boldsymbol{y}_k$ is a solution of $A\boldsymbol{x} = \boldsymbol{b}$, i.e., GMRES determines a solution of $A\boldsymbol{x} = \boldsymbol{b}$ at the kth step. Moreover, since $h_{k+1,k} = \|\boldsymbol{w}_k\|_2 = 0$ and $\boldsymbol{v}_{k+1}$ is not defined, $\dim\mathcal{K}_{k+1} < k+1$. Note also that $\dim A\mathcal{K}_k = k$ holds, since $\dim A\mathcal{K}_k = \mathrm{rank}(AV_k) = \mathrm{rank}(V_{k+1}\bar{H}_k) = \mathrm{rank}(V_k H_k) = k$. ∎

We summarize the above discussion in the following.

Case i $\dim A\mathcal{K}_{k-1} = k-1$ and $\dim A\mathcal{K}_k < \dim\mathcal{K}_k \Longrightarrow h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-1$, $h_{k+1,k} = 0$, and $\mathrm{rank}H_k = k-1$,

Case ii $\dim A\mathcal{K}_{k-1} = k-1$ and $\dim\mathcal{K}_k < k \Longrightarrow h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-2$, $h_{k,k-1} = 0$, and $\mathrm{rank}H_{k-1} = k-1$ (GMRES determines a solution of $A\boldsymbol{x} = \boldsymbol{b}$ at the $(k-1)$st step.),

Case i' $h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-1$, $h_{k+1,k} = 0$, and $\mathrm{rank}H_k = k-1 \Longrightarrow \dim A\mathcal{K}_k = k-1$ and $\dim A\mathcal{K}_k < \dim\mathcal{K}_k$,

Case ii' $h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-1$, $h_{k+1,k} = 0$, and $\mathrm{rank}H_k = k \Longrightarrow \dim A\mathcal{K}_k = k$ and $\dim\mathcal{K}_{k+1} < k+1$ (GMRES determines a solution of $A\boldsymbol{x} = \boldsymbol{b}$ at the kth step.).

This further boils down to the following.

Case I $\dim A\mathcal{K}_{k-1} = k-1$ and $\dim A\mathcal{K}_k < \dim\mathcal{K}_k \Longleftrightarrow h_{i+1,i} \neq 0$, $i = 1, 2, \ldots, k-1$, $h_{k+1,k} = 0$, and $\mathrm{rank}H_k = k-1$.

Case II $\dim A\mathcal{K}_{k-1} = k-1$ and $\dim\mathcal{K}_k < k \Longleftrightarrow h_{i+1,i} \neq 0$ $i = 1, 2, \ldots, k-2$, $h_{k,k-1} = 0$, and $\mathrm{rank}H_{k-1} = k-1$. (GMRES determines a solution of $A\boldsymbol{x} = \boldsymbol{b}$ at the $(k-1)$st step.)

Therefore, we may say that Definitions 2.1 and A.1 classify two kinds of breakdown with in Definition A.3. See also [25, Propositions 4.2, 4.3] for similar results for the consistent case $\boldsymbol{b} \in \mathcal{R}(A)$.

Appendix B. Algorithms of inner-iteration methods. We give the algorithms of the methods for the inner-iteration preconditioning which works on the normal equations $A^{\mathsf{T}}A\boldsymbol{z} = A^{\mathsf{T}}\boldsymbol{c}$. Let $\boldsymbol{a}_j$ be the jth column of A.

ALGORITHM B.1. *Richardson-NR method.*

1. *Let* $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r}^{(0)} := \boldsymbol{c}$.
2. *For* $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad \boldsymbol{z}^{(k)} := \boldsymbol{z}^{(k-1)} + \omega A^{\mathsf{T}}\boldsymbol{r}^{(k-1)}$, $\boldsymbol{r}^{(k)} := \boldsymbol{c} - A\boldsymbol{z}^{(k)}$
4. *EndDo*

ALGORITHM B.2. *Cimmino-NR method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r}^{(0)} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad \boldsymbol{d}^{(k)} := D^{-1}A^{\mathsf{T}}\boldsymbol{r}^{(k-1)}$, $\boldsymbol{z}^{(k)} := \boldsymbol{z}^{(k-1)} + \omega\boldsymbol{d}^{(k)}$, $\boldsymbol{r}^{(k)} := \boldsymbol{r}^{(k-1)} - \omega A\boldsymbol{d}^{(k)}$
4. EndDo

ALGORITHM B.3. *NR-SOR method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad$ *For* $j := 1, 2, \ldots, n$, *Do*
4. $\quad\quad d_j^{(k)} := (\boldsymbol{r}, \boldsymbol{a}_j)/\|\boldsymbol{a}_j\|_2{}^2$, $z_j^{(k)} := z_j^{(k-1)} + \omega d_j^{(k)}$, $\boldsymbol{r} := \boldsymbol{r} - \omega d_j^{(k)}\boldsymbol{a}_j$
5. $\quad$ *EndDo*
6. EndDo

ALGORITHM B.4. *NR-SSOR method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad$ *For* $j = 1, 2, \ldots, n$, *Do*
4. $\quad\quad d_j^{(k-\frac{1}{2})} := (\boldsymbol{r}, \boldsymbol{a}_j)/\|\boldsymbol{a}_j\|_2{}^2$, $z_j^{(k-\frac{1}{2})} := z_j^{(k-1)} + \omega d_j^{(k-\frac{1}{2})}$, $\boldsymbol{r} := \boldsymbol{r} - \omega d_j^{(k-\frac{1}{2})}\boldsymbol{a}_j$
5. $\quad$ *EndDo*
6. $\quad$ *For* $j = n, n-1, \ldots, 1$, *Do*
7. $\quad\quad d_j^{(k)} := (\boldsymbol{r}, \boldsymbol{a}_j)/\|\boldsymbol{a}_j\|_2{}^2$, $z_j^{(k)} := z_j^{(k-\frac{1}{2})} + \omega d_j^{(k)}$, $\boldsymbol{r} := \boldsymbol{r} - \omega d_j^{(k)}\boldsymbol{a}_j$
8. $\quad$ *EndDo*
9. EndDo

Next, we give the algorithms of the methods for the inner-iteration preconditioning which works on the normal equations of the second kind $AA^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{c}$, $\boldsymbol{z} = A^{\mathsf{T}}\boldsymbol{y}$. Let $\boldsymbol{\alpha}_i$ be the ith row of A and c_i be the ith component of $\boldsymbol{c} \in \mathbf{R}^m$.

ALGORITHM B.5. *Richardson-NE method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r}^{(0)} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad \boldsymbol{z}^{(k)} := \boldsymbol{z}^{(k-1)} + \omega A^{\mathsf{T}}\boldsymbol{r}^{(k-1)}$, $\boldsymbol{r}^{(k)} := \boldsymbol{c} - A\boldsymbol{z}^{(k)}$
4. EndDo

ALGORITHM B.6. *Cimmino-NE method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r}^{(0)} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad \boldsymbol{d}^{(k)} := D^{-1}\boldsymbol{r}^{(k-1)}$, $\boldsymbol{z}^{(k)} := \boldsymbol{z}^{(k-1)} + \omega A^{\mathsf{T}}\boldsymbol{d}^{(k)}$, $\boldsymbol{r}^{(k)} := \boldsymbol{c} - A\boldsymbol{z}^{(k)}$
4. EndDo

ALGORITHM B.7. *NE-SOR method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad$ *For* $i := 1, 2, \ldots, m$, *Do*
4. $\quad\quad d_i^{(k)} := \left[c_i - (\boldsymbol{\alpha}_i, \boldsymbol{x}^{(k)})\right]/\|\boldsymbol{\alpha}_i\|_2{}^2$, $\boldsymbol{z}^{(k-1)} := \boldsymbol{z}^{(k-1)} + d_i^{(k)}\boldsymbol{\alpha}_i$
5. $\quad$ *EndDo*
6. $\quad \boldsymbol{z}^{(k)} := \boldsymbol{z}^{(k-1)}$
7. EndDo

ALGORITHM B.8. *NE-SSOR method.*

1. Let $\boldsymbol{z}^{(0)} := \boldsymbol{0}$ *and* $\boldsymbol{r} := \boldsymbol{c}$.
2. For $k = 1, 2, \ldots, \ell$, *Do*
3. $\quad$ *For* $i = 1, 2, \ldots, m$, *Do*

4. $\quad d_i^{(k-\frac{1}{2})} := \left[c_i - (\boldsymbol{\alpha}_i, \boldsymbol{x}^{(k)})\right] / {\|\boldsymbol{\alpha}_i\|_2}^2$, $\boldsymbol{z}^{(k-1)} := \boldsymbol{z}^{(k-1)} + d_i^{(k-\frac{1}{2})}\boldsymbol{\alpha}_i$
5. *EndDo*
6. *For* $i = m, m-1, \ldots, 1$, *Do*
7. $\quad d_i^{(k)} := \left[c_i - (\boldsymbol{\alpha}_i, \boldsymbol{x}^{(k)})\right] / {\|\boldsymbol{\alpha}_i\|_2}^2$, $\boldsymbol{z}^{(k-1)} := \boldsymbol{z}^{(k-1)} + d_i^{(k)}\boldsymbol{\alpha}_i$
8. *EndDo*
9. $\boldsymbol{z}^{(k)} := \boldsymbol{z}^{(k-1)}$
10. *EndDo*

Appendix C. Algorithms based on CG and MINRES.

We give an algorithm of CGLS preconditioned by inner iterations, which is improved to reduce a matrix vector multiplication per iteration. Note that the improved method is mathematically equivalent to the original method in [20].

ALGORITHM C.1. *CGLS method preconditioned by inner iterations.*

1. *Let* $\boldsymbol{x}_0$ *be the initial approximate solution,* $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$, *and* $\boldsymbol{s}_0 := A^{\mathsf{T}}\boldsymbol{r}_0$.
2. *Apply* ℓ *steps of a stationary iterative method to* $A^{\mathsf{T}}A\boldsymbol{z} = \boldsymbol{s}_0$, $\boldsymbol{y} = A\boldsymbol{z}$ *to obtain* $\boldsymbol{z}_0 := C^{(\ell)}\boldsymbol{s}_0$.
3. $\boldsymbol{p}_0 := \boldsymbol{z}_0$, $\gamma_0 := (\boldsymbol{s}_0, \boldsymbol{z}_0)$, $\boldsymbol{q}_0 := \boldsymbol{p}_0$
4. *For* $k = 0, 1, 2, \ldots$ *until convergence, Do*
5. $\quad \alpha_k := \gamma_k / {\|\boldsymbol{q}_k\|_2}^2$, $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k\boldsymbol{p}_k$, $\boldsymbol{r}_{k+1} := \boldsymbol{r}_k - \alpha_k\boldsymbol{q}_k$, $\boldsymbol{s}_{k+1} := A^{\mathsf{T}}\boldsymbol{r}_{k+1}$
6. $\quad$ *Apply* ℓ *steps of a stationary iterative method to* $A^{\mathsf{T}}A\boldsymbol{z} = \boldsymbol{s}_{k+1}$, $\boldsymbol{y} = A\boldsymbol{z}$ *to obtain* $\boldsymbol{z}_{k+1} := C^{(\ell)}\boldsymbol{s}_{k+1}$ *and* $\boldsymbol{y}_{k+1} := A\boldsymbol{z}_{k+1}$.
7. $\quad \gamma_{k+1} := (\boldsymbol{s}_{k+1}, \boldsymbol{z}_{k+1})$, $\beta_k := \gamma_{k+1}/\gamma_k$, $\boldsymbol{p}_{k+1} := \boldsymbol{z}_{k+1} + \beta_k\boldsymbol{p}_k$, $\boldsymbol{q}_{k+1} := \boldsymbol{y}_{k+1} + \beta_k\boldsymbol{q}_k$
8. *EndDo*

We give algorithms of other methods preconditioned by inner iterations, which were used in section 6.

ALGORITHM C.2. *LSMR method preconditioned by inner iterations.*

1. *Let* $\boldsymbol{x}_0$ *be the initial approximate solution and* $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.
2. $\beta_1\boldsymbol{u}_1 := \boldsymbol{r}_0$, $\bar{\boldsymbol{v}}_1 := A^{\mathsf{T}}\boldsymbol{u}_1$ *(shorthand for* $\beta_1 := \|\boldsymbol{r}_0\|_2$, $\boldsymbol{u}_1 := \boldsymbol{r}_0/\beta_1$*)*
3. *Apply* ℓ *steps of a stationary iterative method to* $A^{\mathsf{T}}A\boldsymbol{z} = \bar{\boldsymbol{v}}_1$, $\boldsymbol{w} = A\boldsymbol{z}$ *to obtain* $\boldsymbol{w}_1 := AC^{(\ell)}\bar{\boldsymbol{v}}_1$.
4. $\alpha_1 := (\boldsymbol{u}_1, \boldsymbol{w}_1)^{1/2}$, $\hat{\boldsymbol{v}}_1 := \bar{\boldsymbol{v}}_k/\alpha_k$, $\bar{\alpha}_1 := \alpha_1$, $\bar{\zeta}_1 := \alpha_1\beta_1$, $\rho_0 := 1$, $\bar{\rho}_0 := 1$, $\bar{c}_0 := 1$, $\bar{s}_0 := 0$, $\hat{\boldsymbol{h}}_1 := \hat{\boldsymbol{v}}_1$, $\hat{\bar{\boldsymbol{h}}}_0 := \boldsymbol{0}$
5. *For* $k = 1, 2, \ldots$ *until convergence, Do*
6. $\quad \beta_{k+1}\boldsymbol{u}_{k+1} := \boldsymbol{w}_k/\alpha_k - \alpha_k\boldsymbol{u}_k$, $\bar{\boldsymbol{v}}_{k+1} := A^{\mathsf{T}}\boldsymbol{u}_{k+1} - \beta_{k+1}\hat{\boldsymbol{v}}_k$.
7. $\quad$ *Apply* ℓ *steps of a stationary iterative method to* $A^{\mathsf{T}}A\boldsymbol{z} = \bar{\boldsymbol{v}}_{k+1}$, $\boldsymbol{w} = A\boldsymbol{z}$ *to obtain* $\boldsymbol{w}_{k+1} := AC^{(\ell)}\bar{\boldsymbol{v}}_{k+1}$.
8. $\quad \alpha_{k+1} := (\boldsymbol{u}_{k+1}, \boldsymbol{w}_{k+1})^{1/2}$, $\hat{\boldsymbol{v}}_{k+1} := \bar{\boldsymbol{v}}_{k+1}/\alpha_{k+1}$.
9. $\quad \rho_k := (\bar{\alpha}_k^2 + {\beta_{k+1}}^2)$, $c_k := \bar{\alpha}_k/\rho_k$, $s_k := \beta_{k+1}/\rho_k$, $\theta_{k+1} := s_k$, $\bar{\alpha}_{k+1} := c_k\alpha_{k+1}$
10. $\quad \bar{\theta}_k := \bar{s}_{k-1}\rho_k$, $\bar{\rho}_k := ((\bar{c}_{k-1}\rho_k) + {\theta_{k+1}}^2)^{\frac{1}{2}}$, $\bar{c}_k := \bar{c}_{k-1}\rho_k/\bar{\rho}_k$, $\bar{s}_k := \theta_{k+1}/\bar{\rho}_k$, $\zeta_k := \bar{c}_k\bar{\zeta}_k$, $\bar{\zeta}_{k+1} := -\bar{s}_k\bar{\zeta}_k$
11. $\quad \hat{\bar{\boldsymbol{h}}}_k := \hat{\boldsymbol{h}}_k - (\bar{\theta}_k\rho_k/(\rho_{k-1}\bar{\rho}_{k-1}))\hat{\bar{\boldsymbol{h}}}'_{k-1}$, $\boldsymbol{w}_k := \boldsymbol{w}_{k-1} + (\zeta_k/(\rho_k\bar{\rho}_k))\hat{\boldsymbol{h}}_k$, $\hat{\boldsymbol{h}}_{k+1} := \hat{\boldsymbol{v}}_{k+1} - (\theta_{k+1}/\rho_k)\hat{\boldsymbol{h}}_k$
12. *EndDo*
13. *Apply* ℓ *steps of a stationary iterative method to* $A^{\mathsf{T}}A\boldsymbol{x} = \boldsymbol{w}_k$ *to obtain* $\boldsymbol{x}_k := C^{(\ell)}\boldsymbol{w}_k$.

ALGORITHM C.3. *CGNE method preconditioned by inner iterations.*

1. *Let* $\boldsymbol{x}_0$ *be the initial approximate solution and* $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.

2. *Apply ℓ steps of a stationary iterative method to $AA^T\boldsymbol{z}_0 = \boldsymbol{r}_0$ to obtain $\tilde{\boldsymbol{z}}_0 := C^{(\ell)}\boldsymbol{r}_0$.*
3. $\boldsymbol{q}_0 := \tilde{\boldsymbol{z}}_0$, $\gamma_0 := (\boldsymbol{r}_0, \tilde{\boldsymbol{z}}_0)$
4. *For $k = 0, 1, 2, \ldots$ until convergence, Do*
5. $\quad \boldsymbol{s}_k := A^T\boldsymbol{q}_k$, $\alpha_k := \gamma_k/(\boldsymbol{s}_k, \boldsymbol{s}_k)$, $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k\boldsymbol{s}_k$, $\boldsymbol{r}_{k+1} := \boldsymbol{r}_k - \alpha_k A\boldsymbol{s}_k$
6. $\quad$ *Apply ℓ steps of a stationary iterative method to $AA^T\boldsymbol{z}_{k+1} = \boldsymbol{r}_{k+1}$ to obtain $\tilde{\boldsymbol{z}}_{k+1} := C^{(\ell)}\boldsymbol{r}_{k+1}$.*
7. $\quad \gamma_{k+1} := (\boldsymbol{r}_{k+1}, \tilde{\boldsymbol{z}}_{k+1})$, $\beta_k := \gamma_{k+1}/\gamma_k$, $\boldsymbol{q}_{k+1} := \tilde{\boldsymbol{z}}_{k+1} + \beta_k\boldsymbol{q}_k$
8. *EndDo*

ALGORITHM C.4. *MRNE method preconditioned by inner iterations.*

1. *Let $\boldsymbol{x}_0$ be the initial approximate solution and $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}_0$.*
2. *Apply ℓ steps of a stationary iterative method to $AA^{\mathsf{T}}\boldsymbol{u} = \boldsymbol{r}_0$, $\boldsymbol{s} = A^{\mathsf{T}}\boldsymbol{u}$ to obtain $\boldsymbol{s}_0 := A^{\mathsf{T}}C^{(\ell)}\boldsymbol{r}_0$.*
3. $\boldsymbol{p}_0 := \boldsymbol{s}_0$, $\gamma_0 := \|\boldsymbol{s}_0\|_2{}^2$
4. *For $k = 0, 1, 2, \ldots$ until convergence, Do*
5. $\quad \boldsymbol{t}_k := A\boldsymbol{p}_k$
6. $\quad$ *Apply ℓ steps of a stationary iterative method to $AA^{\mathsf{T}}\boldsymbol{u} = \boldsymbol{t}_k$, $\boldsymbol{v} = A^{\mathsf{T}}\boldsymbol{u}$ to obtain $\boldsymbol{v}_k := A^{\mathsf{T}}C^{(\ell)}\boldsymbol{t}_k$.*
7. $\quad \alpha_k := \gamma_k/(\boldsymbol{v}_k, \boldsymbol{p}_k)$, $\boldsymbol{x}_k := \boldsymbol{x}_k + \alpha_k\boldsymbol{p}_k$, $\boldsymbol{r}_{k+1} := \boldsymbol{r}_k - \alpha_k\boldsymbol{t}_k$, $\boldsymbol{s}_{k+1} := \boldsymbol{s}_k - \alpha_k\boldsymbol{v}_k$, $\gamma_k := \|\boldsymbol{s}_{k+1}\|_2{}^2$, $\beta_k := \gamma_{k+1}/\gamma_k$, $\boldsymbol{p}_{k+1} := \boldsymbol{s}_k + \beta_k\boldsymbol{p}_k$
8. *EndDo*

REFERENCES

[1] M. ARIOLI, V. PTÁK, AND Z. STRAKOŠ, *Krylov sequences of maximal length and convergece of GMRES*, BIT, 38 (1998), pp. 636–643.

[2] Z.-Z. BAI, *Sharp error bounds of some krylov subspace methods for non-hermitian linear systems*, Appl. Math. Comput., 109 (2000), pp. 273–285.

[3] Å. BJÖRCK AND T. ELFVING, *Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations*, BIT, 19 (1979), pp. 145–163.

[4] P. N. BROWN AND H. F. WALKER, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.

[5] E.J. CRAIG, *The n-step iteration procedures*, J. Math. Phys., 34 (1955), pp. 64–73.

[6] X. CUI, K. HAYAMI, AND J.-F. YIN, *Greville's method for preconditioning least squares problems*, Adv. Comput. Math., 35 (2011), pp. 243–269.

[7] T. DAVIS, *The University of Florida Sparse Matrix Collection*, `http://www.cise.ufl.edu/research/sparse/matrices/`.

[8] T. A. DAVIS, *Algorithm 915, suiteSparseQR: Multifrontal multithreaded rank-revealing sparse QR factorization*, ACM Trans. Math. Software, 38 (2011).

[9] A. DAX, *The convergece of linear stationary iterative processes for solving singular unstructured systems of linear equations*, SIAM Review, 32 (1990), pp. 611–635.

[10] D. C.-L. FONG AND M. SAUNDERS, *LSMR: An iterative algorithm for sparse least-squares problems*, SIAM J. Sci. Comput., 33 (2011), pp. 2950–2971.

[11] L. FOSTER, *San Jose State University Singular Matrix Database*, `http://www.math.sjsu.edu/singular/matrices/`.

[12] A. GREENBAUM, V. PTÁK, AND Z. STRAKOŠ, *Any nonincreasing convergence curve is possible for gmres*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.

[13] K. HAYAMI AND M. SUGIHARA, *A geometric view of Krylov subspace methods on singular systems*, Numer. Linear Algebra Appl., 18 (2011), pp. 449–469.

[14] K. HAYAMI, J.-F. YIN, AND T. ITO, *GMRES methods for least squares problems*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2400–2430.

[15] V. K. HENSEL, *Über Potenzreihen von Matrizen*, J. Reine Angew. Math., 155 (1926), pp. 107–110 (in German).

[16] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.

[17] H. B. Keller, *On the solution of singular and semidefinite linear systems by iteration*, J. SIAM Numer. Anal., 2 (1965), pp. 281–290.
[18] C. D. Meyer and R. J. Plemmons, *Convergent powers of a matrix with applications to iterative methods for singular linear sysmtes*, SIAM J. Numer. Anal., 14 (1977), pp. 699–705.
[19] K. Morikuni and K. Hayami, *Inner-iteration Krylov subspace methods for least squares problems*, NII Technical Report NII-2011-001E, `http://www.nii.ac.jp/TechReports/11-001E.html`, (2011), pp. 1–27.
[20] ———, *Inner-iteration Krylov subspace methods for least squares problems*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 1–22.
[21] R. Oldenburger, *Infinite powers of matrices and characteristic roots*, Duke Math., 6 (1940), pp. 357–361.
[22] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., (1975), pp. 617–629.
[23] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
[24] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
[25] L. Smoch, *Spectral behaviour of GMRES applied to singular systems*, Adv. Comput. Math., 27 (2007), pp. 151–166.
[26] K. Tanabe, *Characterization of linear stationary iterative processes for solving a singular system of linear equations*, Numer. Math., 22 (1974), pp. 349–359.
[27] J. D. Tebbens and G. Meurant, *Prescribing the behaviour of early terminating GMRES and Arnoldi iterations*, Numer. Algor., DOI 10.1007/s11075-013-9695-x, (2013).
[28] H. A. Van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems*, SIAM J. Sci. Stat. Comput., 13 (1992), pp. 631–644.