ISSN 1346-5597

National Institute of Informatics

NII Technical Report

Convergence of Inner-Iteration GMRES Methods for Least Squares Problems

Keiichi Morikuni and Ken Hayami

NII-2012-005E
Aug. 2012

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 for least squares problems preconditioned with inner iterations. 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 is improved particularly in the rank-deficient case. We analyse the spectrum of the preconditioned coefficient matrix, and characterize it by the spectral radius of the iteration matrix for the inner iterations. The analysis is supported by numerical experiments.

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 of A. The least squares problem (1.1) is equivalent to the normal equations

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

In addition, applying $B \in \mathbf{R}^{n\times m}$, we may transform the problem (1.1) to equivalent problems [6, Theorems 3.1, 3.11].

THEOREM 1.1. *$\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2 = \min_{\boldsymbol{z}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\|_2$ holds for all $\boldsymbol{b} \in \mathbf{R}^m$ if and only if $\mathcal{R}(AB) = \mathcal{R}(A)$.*

THEOREM 1.2. *$\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ 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 $\mathcal{R}(B^\mathsf{T} BA) = \mathcal{R}(A)$.*

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) [12] was extended to deal with least squares problems (1.1) in [6]. 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 without breakdown for arbitrary $\boldsymbol{b}$ for overdetermined, underdetermined, and rank-deficient problems, were shown.

In [9], these methods were preconditioned with 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. The

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

†Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies, Sokendai, 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).

Cimmino-NE and Cimmino-NR methods are mathematically equivalent to JOR applied to $AA^\mathsf{T}\boldsymbol{u} = \boldsymbol{b}$ with $\boldsymbol{x} = A^\mathsf{T}\boldsymbol{u}$ and $A^\mathsf{T}A\boldsymbol{x} = A^\mathsf{T}\boldsymbol{b}$, respectively. The normal-error (NE-)SOR and normal-residual (NR-)SOR methods are mathematically equivalent to SOR applied to $AA^\mathsf{T}\boldsymbol{u} = \boldsymbol{b}$ with $\boldsymbol{x} = A^\mathsf{T}\boldsymbol{u}$ and $A^\mathsf{T}A\boldsymbol{x} = A^\mathsf{T}\boldsymbol{b}$, respectively [9], [11], [1].

Krylov subspace methods preconditioned with inner iterations for solving linear systems of equations were described in [9] and references therein.

We assumed that A should be of full-column rank for the convergence theory for BA-GMRES with the Cimmino-NR and NR-SOR inner iterations in [9], but numerical experiments in [9] 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.

The outline of the paper is as follows. In Section 2, we introduce AB- and BA-GMRES, give their new convergence theory, and analyse the spectrum of the preconditioned matrix. In Section 3, we give a framework of inner-iteration preconditioning for these methods, and main results on sufficient conditions in terms of the inner iterations for the convergence. In Section 4, we analyse the spectrum of the the preconditioned matrix with inner iterations. In Section 5, we give the main conclusions of this paper.

Throughout this paper, we use bold letters for column vectors. $\boldsymbol{e}_j$ denotes the jth column of an identity matrix. We denote quantities related to the kth inner iteration with a superscript with brackets, e.g., $\boldsymbol{x}^{(k)}$, and for outer iterations with 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 $\boldsymbol{b}$.

2. GMRES methods for least squares problems. We first give an explanation about AB-GMRES and BA-GMRES. AB-GMRES applies GMRES to $\min_{\boldsymbol{u}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{u}\|_2$ with $\boldsymbol{x} = B\boldsymbol{u}$, whereas BA-GMRES applies GMRES to $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$.

Concerning the convergence of AB-GMRES and BA-GMRES, we have the following [6, Corollaries 3.8, 3.19].

THEOREM 2.1. *If* $\mathcal{R}(B^\mathsf{T}) = \mathcal{R}(A)$ *and* $R(B) = \mathcal{R}(A^\mathsf{T})$, *then AB-GMRES 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 all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown.*

Here we say 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 $\dim \mathcal{K}_k(AB, \boldsymbol{r}_0) < k$, where

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

is the Krylov subspace of dimension k [2]. The breakdown causes a division by 0 in the algorithm.

THEOREM 2.2. *If* $\mathcal{R}(B^\mathsf{T}) = \mathcal{R}(A)$ *and* $R(B) = \mathcal{R}(A^\mathsf{T})$, *then BA-GMRES 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 all* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown.*

We say 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 $\dim \mathcal{K}_k(BA, B\boldsymbol{r}_0) < k$ [2].

Note that $\mathcal{R}(B) = \mathcal{R}(A^\mathsf{T})$ gives $\mathcal{R}(AB) = \mathcal{R}(A)$ (cf. Theorem 1.1), and $\mathcal{R}(B^\mathsf{T}) = \mathcal{R}(A)$ gives $\mathcal{R}(B^\mathsf{T}BA) = \mathcal{R}(A)$ (cf. Theorem 1.2).

On the other hand, the convergence of the standard GMRES method for least squares problems $\min_{\tilde{\boldsymbol{x}}\in\mathbf{R}^N} \|\tilde{\boldsymbol{b}} - \tilde{A}\tilde{\boldsymbol{x}}\|_2$ with $\tilde{A} \in \mathcal{R}^{N\times N}$, is given as follows [5, Theorem 2.8].

THEOREM 2.3. *GMRES determines a solution of* $\min_{\tilde{\boldsymbol{x}}\in\mathbf{R}^N} \|\tilde{\boldsymbol{b}} - \tilde{A}\tilde{\boldsymbol{x}}\|_2$ *for all* $\tilde{\boldsymbol{b}} \in \mathcal{R}(\tilde{A})$, $\tilde{\boldsymbol{x}}_0 \in \mathbf{R}^N$ *if and only if* $\mathcal{R}(\tilde{A}) \cap \mathcal{N}(\tilde{A}) = \{\mathbf{0}\}$.

Here, $\mathcal{N}(\tilde{A})$ is the null space of $\tilde{A}$ and $\tilde{\boldsymbol{x}}_0$ is the initial approximate solution for GMRES. Applying this theorem to the case of AB- and BA-GMRES, we obtain the following.

THEOREM 2.4. *Suppose that* $R(B) = \mathcal{R}(A^\mathsf{T})$. *Then, AB-GMRES determines a solution of* $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ *without breakdown for all* $\boldsymbol{b} \in \mathcal{R}(A)$ *and* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *if and only if* $\mathcal{R}(A) \cap \mathcal{N}(B) = \{\mathbf{0}\}$.

Proof. Substitute AB, $\boldsymbol{u}$, and $\boldsymbol{b}$ into $\tilde{A}$, $\tilde{\boldsymbol{x}}$, and $\tilde{\boldsymbol{b}}$, respectively, in Theorem 2.3. $R(B) = \mathcal{R}(A^\mathsf{T})$ gives $\mathcal{R}(AB) = \mathcal{R}(AA^\mathsf{T}) = \mathcal{R}(A)$ and $\mathcal{N}(AB) = \mathcal{R}(B^\mathsf{T}A^\mathsf{T})^\perp = \mathcal{R}(B^\mathsf{T}B)^\perp = \mathcal{R}(B^\mathsf{T})^\perp = \mathcal{N}(B)$, where $S^\perp$ is the orthogonal complement of a subspace S. Theorem 1.1 completes the proof. ∎

We remark that this theorem is restricted to the consistent case $\boldsymbol{b} \in \mathcal{R}(A)$.

A similar theorem holds for BA-GMRES.

THEOREM 2.5. *Suppose that* $R(B^\mathsf{T}) = \mathcal{R}(A)$. *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* $\boldsymbol{x}_0 \in \mathbf{R}^n$ *if and only if* $\mathcal{N}(A) \cap \mathcal{R}(B) = \{\mathbf{0}\}$.

Proof. Substitute BA, $\boldsymbol{x}$, and $B\boldsymbol{b}$ into $\tilde{A}$, $\tilde{\boldsymbol{x}}$, and $\tilde{\boldsymbol{b}}$, respectively, in Theorem 2.3. $R(B^\mathsf{T}) = \mathcal{R}(A)$ 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 $\boldsymbol{b} \in \mathbf{R}^m$". Therefore, Theorem 1.2 completes the proof. ∎

In contrast to BA-GMRES, the condition $\boldsymbol{b} \in \mathcal{R}(AB) = \mathcal{R}(A)$ is required for AB-GMRES since the preconditioned system $\min_{\boldsymbol{u}\in\mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{u}\|_2$ is inconsistent for $\boldsymbol{b} \notin \mathcal{R}(AB)$.

Since the BA-GMRES algorithm will be treated in Section 3, it is given in the following. For the AB-GMRES algorithm, see [6].

ALGORITHM 2.6. *BA-GMRES method.*

1. *Let* $\boldsymbol{x}_0$ *be the initial approximate solution.*
2. $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$, $\tilde{\boldsymbol{r}}_0 = B\boldsymbol{r}_0$, $\beta = \|\tilde{\boldsymbol{r}}_0\|_2$, $\boldsymbol{v}_1 = \tilde{\boldsymbol{r}}_0/\beta$
3. *For* $k = 1, 2, \ldots$ *until convergence, Do*
4. $\quad \boldsymbol{w}_k = BA\boldsymbol{v}_k$
5. $\quad$ *For* $i = 1, 2, \ldots, k$, *Do*
6. $\quad\quad h_{i,k} = (\boldsymbol{w}_k, \boldsymbol{v}_i)$, $\boldsymbol{w}_k = \boldsymbol{w}_k - h_{i,k}\boldsymbol{v}_i$
7. $\quad$ *EndDo*
8. $\quad h_{k+1,k} = \|\boldsymbol{w}_k\|_2$, $\boldsymbol{v}_{k+1} = \boldsymbol{w}_k/h_{k+1,k}$
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, $\bar{H}_k = \{h_{i,j}\} \in \mathbf{R}^{(k+1)\times k}$.

2.1. Spectrum of the preconditioned matrix. We describe how BA-GMRES depends on the spectrum of the preconditioned matrix. Assume $\mathcal{R}(B^\mathsf{T}) = \mathcal{R}(A)$. Then, $B\boldsymbol{b} \in \mathcal{R}(BA) = \mathcal{R}(B)$ holds, and $\min_{\boldsymbol{x}\in\mathcal{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$ is equivalent to $BA\boldsymbol{x} = B\boldsymbol{b}$. Let $r = \text{rank } A$, $Q_1 \in \mathbf{R}^{n\times r}$ 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 $\boldsymbol{b}^2 = {Q_2}^\mathsf{T}B\boldsymbol{b} = \mathbf{0}$ since $B\boldsymbol{b} \in \mathcal{R}(BA)$. As shown in [5], 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 2.7. *A_{11} is nonsingular if and only if $\mathcal{R}(BA) \cap \mathcal{N}(BA)$.*

Proof. See [5, Theorem 2.3]. □

THEOREM 2.8. *Assume $\mathcal{R}(BA) \cap \mathcal{N}(BA)$. 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,

$$\begin{aligned} BA\boldsymbol{u} = \lambda\boldsymbol{u},\ \boldsymbol{u} \neq \mathbf{0} \quad &\Longleftrightarrow \quad Q^\mathsf{T}(BA)QQ^\mathsf{T}\boldsymbol{u} = \lambda Q^\mathsf{T}\boldsymbol{u},\ \boldsymbol{u} \neq \mathbf{0} \\ &\Longleftrightarrow \quad \begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}^1 \\ \boldsymbol{u}^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}^1 \\ \boldsymbol{u}^2 \end{bmatrix},\ \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}^1 \\ \boldsymbol{u}^2 \end{bmatrix} \neq \mathbf{0} \\ &\Longleftrightarrow \quad \begin{cases} A_{11}\boldsymbol{u}^1 + A_{12}\boldsymbol{u}^2 = \lambda\boldsymbol{u}^1, \\ \mathbf{0} = \lambda\boldsymbol{u}^2, \end{cases} \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}^1 \\ \boldsymbol{u}^2 \end{bmatrix} \neq \mathbf{0}. \end{aligned}$$

Hence, if $\lambda \neq 0$ is an eigenvalue of BA, then we have $\boldsymbol{u}^2 = \mathbf{0}$, $A_{11}\boldsymbol{u}^1 = \lambda\boldsymbol{u}^1$, and $\boldsymbol{u}^1 \neq \mathbf{0}$ so that $\lambda \neq 0$ is an eigenvalue of A_{11}. On the other hand, if $\lambda \neq 0$ is an eigenvalue of A_{11}, then by setting $\boldsymbol{u}^2 = \mathbf{0}$, we can show that $\lambda \neq 0$ is an eigenvalue of BA. □

Assume $\mathcal{R}(BA) \cap \mathcal{N}(BA) = \{\mathbf{0}\}$, equivalently $\mathcal{R}(B) \cap \mathcal{N}(A) = \{\mathbf{0}\}$. Then, A_{11} is nonsingular, and its eigenvalues are all nonzero and correspond to the nonzero eigenvalues of BA. Therefore, the convergence behavior of GMRES applied to $BA\boldsymbol{x} = B\boldsymbol{b}$ may be explained by the (nonzero) eigenvalues of A_{11} (or BA).

3. BA-GMRES preconditioned by stationary iterative methods as inner iterations. Instead of applying B explicitly as in Algorithm 2.6, consider using inner iterations as follows [9].

ALGORITHM 3.1. *BA-GMRES method with the inner-iteration preconditioning.*

1. *Let $\boldsymbol{x}_0$ be the initial approximate solution.*
2. $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
3. *Apply ℓ steps of a stationary iterative method to $A^\mathsf{T}A\boldsymbol{z} = A^\mathsf{T}\boldsymbol{r}_0$ to obtain* $\boldsymbol{z}_0^{(\ell)} = B^{(\ell)}\boldsymbol{r}_0$.
4. $\beta = \|\boldsymbol{z}_0^{(\ell)}\|_2$, $\boldsymbol{v}_1 = \boldsymbol{z}_0^{(\ell)}/\beta$
5. *For $k = 1, 2, \dots$ until convergence, Do*
6. $\quad \boldsymbol{u}_k = A\boldsymbol{v}_k$
7. $\quad$ *Apply ℓ steps of a stationary iterative method to $A^\mathsf{T}A\boldsymbol{y} = A^\mathsf{T}\boldsymbol{u}_k$ to obtain* $\boldsymbol{z}_k^{(\ell)} = B^{(\ell)}\boldsymbol{u}_k$.
8. $\quad$ *For $i = 1, 2, \dots, k$, Do*
9. $\quad\quad h_{i,k} = (\boldsymbol{z}_k^{(\ell)}, \boldsymbol{v}_i)$, $\boldsymbol{z}_k^{(\ell)} = \boldsymbol{z}_k^{(\ell)} - h_{i,k}\boldsymbol{v}_i$

10.   *EndDo*
11. $h_{k+1,k} = \|\boldsymbol{z}_k^{(\ell)}\|_2$, $\boldsymbol{v}_{k+1} = \boldsymbol{z}_k^{(\ell)}/h_{k+1,k}$
12. EndDo
13. $\boldsymbol{y}_k = \arg\min_{\boldsymbol{y}\in\mathbf{R}^k} \|\beta\boldsymbol{e}_1 - \bar{H}_k\boldsymbol{y}\|_2$, $\boldsymbol{x}_k = \boldsymbol{x}_0 + [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k]\,\boldsymbol{y}_k$

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

Similarly we can obtain AB-GMRES with stationary inner iterations.

In lines 3 and 7 in Algorithm 3.1, 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 $H = M^{-1}N = I - M^{-1}A^{\mathsf{T}}A$ 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 Cimmino-NR and NR-SOR methods [11] in Section 3.1.

Here, we define the following, e.g., [8].

DEFINITION 3.2. *A matrix C is called semi-convergent if $\lim_{i\to\infty} C^i$ exists.*

Hensel [7], Oldenburger [10], and Tanabe [13] showed the following.

THEOREM 3.3. *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 $\operatorname{index}(I - C) = 1$

   holds.

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

3.1. Convergence theory. 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

$$\boldsymbol{z}^{(\ell)} = H\boldsymbol{z}^{(\ell-1)} + M^{-1}A^{\mathsf{T}}\boldsymbol{c} = \sum_{j=0}^{\ell-1} H^j M^{-1}A^{\mathsf{T}}\boldsymbol{c}. \tag{3.1}$$

Hence, if we define the preconditioning matrix by

$$B^{(\ell)} = \sum_{j=0}^{\ell-1} H^j M^{-1}A^{\mathsf{T}}, \tag{3.2}$$

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

We obtain the following.

LEMMA 3.4. *Let $A \in \mathbf{R}^{m\times n}$, $A^{\mathsf{T}}A = M - N$, where M is nonsingular, $H = M^{-1}N$, and $B^{(\ell)}$ be given by* (3.2)*. Assume that H is semi-convergent. Then, $\operatorname{index}(B^{(\ell)}A) \le 1$ for all $\ell \ge 1$.*

Proof. Let $J = S^{-1}(I - H)S$ be the Jordan canonical form of $(I - H)$. Assume that H is semi-convergent. Then, from Theorem 3.3, $\mathrm{index}(I - H) = \mathrm{index}(J) \le 1$. Without loss of generality, we denote J as follows:

$$J = \begin{bmatrix} \tilde{J} & \\ & J_s \end{bmatrix} \in \mathbf{C}^{n \times n}, \quad \tilde{J} = \mathrm{diag}\,(J_1, J_2, \dots, J_{s-1}) \in \mathbf{C}^{r \times r}, \quad J_s = 0_{n-r}, \tag{3.3}$$

$$J_i = \mathrm{diag}\left(J_{i_1}, J_{i_2}, \dots, J_{i_{s_i}}\right) \in \mathbf{C}^{n_i \times n_i} \quad (i = 1, 2, \dots, s-1), \quad \sum_{i=1}^{s-1} n_i = r, \tag{3.4}$$

$$J_{i_j} = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & 1 & \quad 0 \\ & & \ddots & \ddots \\ & 0 & & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbf{C}^{n_{i_j} \times n_{i_j}}, \quad \sum_{j=1}^{s_i} n_{i_j} = n_i, \tag{3.5}$$

where $r = \mathrm{rank}\, A$, 0_{n-r} is the zero matrix of size $n - r$, $\tilde{J}$ has no eigenvalues equal to zero, s is the number of distinct eigenvalues of $I - H$, and λ_i is a nonzero eigenvalue of $I - H$. Since $\tilde{J}$ is nonsingular,

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

Since $\lambda_i = 1 - \nu(H)$, where $\nu(H)$ is an eigenvalue of H and $|\nu(H)| < 1$, $|1 - \lambda_i| = |\nu(H)| < 1$. The eigenvalue of $I_r - (I_r - \tilde{J})^\ell$ has the form $\mu_i = 1 - (1 - \lambda_i)^\ell$. If $\mu_i = 0$, then $(1 - \lambda_i)^\ell = 1$, which contradicts $|1 - \lambda_i| < 1$. Hence, $I_r - (I_r - \tilde{J})^\ell$ is nonsingular. Therefore, $\mathrm{index}(B^{(\ell)} A) \le 1$ for all $\ell \ge 1$. □

Lemma 3.5. *Using the notations and the assumption of Lemma 3.4, $\mathcal{R}\left({B^{(\ell)}}^\mathsf{T}\right) = \mathcal{R}(A)$ holds for all $\ell \ge 1$.*

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

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

$$C^{(\ell)} = \sum_{j=0}^{\ell-1} (I - SJS^{-1})^j M^{-1} = S \begin{bmatrix} \tilde{J}^{-1} \left[I_r - (I_r - \tilde{J})^\ell\right] & 0 \\ 0 & \ell I_{n-r} \end{bmatrix} S^{-1} M^{-1}.$$

As in Lemma 3.4, $I_r - (I_r - \tilde{J})^\ell$ is nonsingular. Hence, $C^{(\ell)}$ is nonsingular for all $\ell \ge 1$. Therefore, we have $\mathcal{R}\left({B^{(\ell)}}^\mathsf{T}\right) = \mathcal{R}(A)$ for all $\ell \ge 1$. □

Hence, we obtain the main result.

Theorem 3.6. *Assume that H is semi-convergent. Then, BA-GMRES with the inner-iteration preconditioning of the form* (3.1) *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 all $\boldsymbol{x}_0 \in \mathbf{R}^n$.*

Proof. Assume that H is semi-convergent. Then, from Lemma 3.4, we have $\mathrm{index}(B^{(\ell)} A) \le 1$, 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 3.5, 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 2.5 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. As for the convergence of AB-GMRES with inner iterations, it follows from Theorem 2.4 that a similar convergence theorem holds, which may be shown via lemmas similar to Lemmas 3.4 and 3.5.

Now, we consider applying Theorem 3.6 for BA-GMRES preconditioned with specific inner-iteration methods as follows. The inner-iteration preconditioning matrices for Cimmino-NR, NR-SOR, and NR-SSOR [9] are respectively obtained from (3.2) by setting

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

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. The following are the algorithms for these methods for inner iterations.

ALGORITHM 3.7. *Cimmino-NR method.*

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

ALGORITHM 3.8. *NR-SOR method.*

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

ALGORITHM 3.9. *NR-SSOR method.*

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

Here, $\boldsymbol{a}_j$ is the jth column of A.

According to Dax [4], the iteration matrix H for

- Cimmino-NR with $0 < \omega < 2/\rho(D^{-1/2}A^{\mathsf{T}}AD^{-1/2})$
- NR-SOR with $0 < \omega < 2$,
- NR-SSOR with $0 < \omega < 2$,

is semi-convergent, where $\rho(C)$ is the spectral radius of C. Here, we assume A has no zero columns. Hence, from Theorem 3.6, these methods can serve as the inner iterations for BA-GMRES.

THEOREM 3.10. *Assume that the relaxation parameter for Cimmino-NR satisfies* $0 < \omega < 2/\rho(D^{-1/2}A^{\mathsf{T}}AD^{-1/2})$. *Then, BA-GMRES with the Cimmino-NR inner-iteration preconditioning 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 all* $\boldsymbol{x}_0 \in \mathbf{R}^n$.

THEOREM 3.11. *Assume that the relaxation parameter for NR-SOR satisfies* $0 < \omega < 2$. *Then, BA-GMRES with the NR-SOR inner-iteration preconditioning 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 all*

$\boldsymbol{x}_0 \in \mathbf{R}^n$.

THEOREM 3.12. *Assume that the relaxation parameter for NR-SSOR satisfies $0 < \omega < 2$. Then, BA-GMRES with the NR-SSOR inner-iteration preconditioning 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 all* $\boldsymbol{x}_0 \in \mathbf{R}^n$.

We omit similar convergence theorems and their proofs for AB-GMRES preconditioned with the Cimmino-NE, NE-SOR, and NE-SSOR inner iterations [9].

4. Spectrum of the preconditioned matrix with inner iterations. Next, we analyse the spectrum of the preconditioned matrix for BA-GMRES with ℓ inner iterations. The preconditioned matrix may be expressed as $B^{(\ell)}A = I - H^\ell$.

Let ν be an eigenvalue of H. Then, there exists $\boldsymbol{v} \neq \boldsymbol{0}$ such that $H\boldsymbol{v} = \nu\boldsymbol{v}$ or $\left(I - H^\ell\right)\boldsymbol{v} = (1 - \nu^\ell)\boldsymbol{v}$. Hence, $B^{(\ell)}A$ has an eigenvalue $\mu = 1 - \nu^\ell$.

Assume that H is semi-convergent. Let $r = \operatorname{rank} A$. Then, from Theorem 3.3, 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

$$|\mu - 1| = |\nu|^\ell \leq \rho(H)^\ell < 1.$$

This means that r eigenvalues of $B^{(\ell)}A$ lie inside the circle of radius $\rho(H)^\ell$ with center at 1, and these eigenvalues approaches 1 as ℓ increases. The remaining $n - r$ eigenvalues are zero.

We demonstrate the above observation for a test matrix called Maragal_3 [3] of size $1{,}690 \times 860$ with 18,391 nonzero elements (nonzero density 1.27 %), and rank 613. 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 $\omega = 1$. The computations were done using MATLAB 2011b. As the number of inner iterations ℓ increased, the eigenvalues of $B^{(\ell)}A$ approached 1.

From the discussion in Section 2.1, we see that if H is semi-convergent, then $\|B^{(\ell)}\boldsymbol{r}\|_2$ depends on the eigenvalues of H not equal to 1, but not on the eigenvalues equal to 1.

5. Conclusions. We considered applying 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 analysed and characterized using the spectral radius of the iteration matrix for the inner iterations. Numerical experiments were done to examine the analysis.

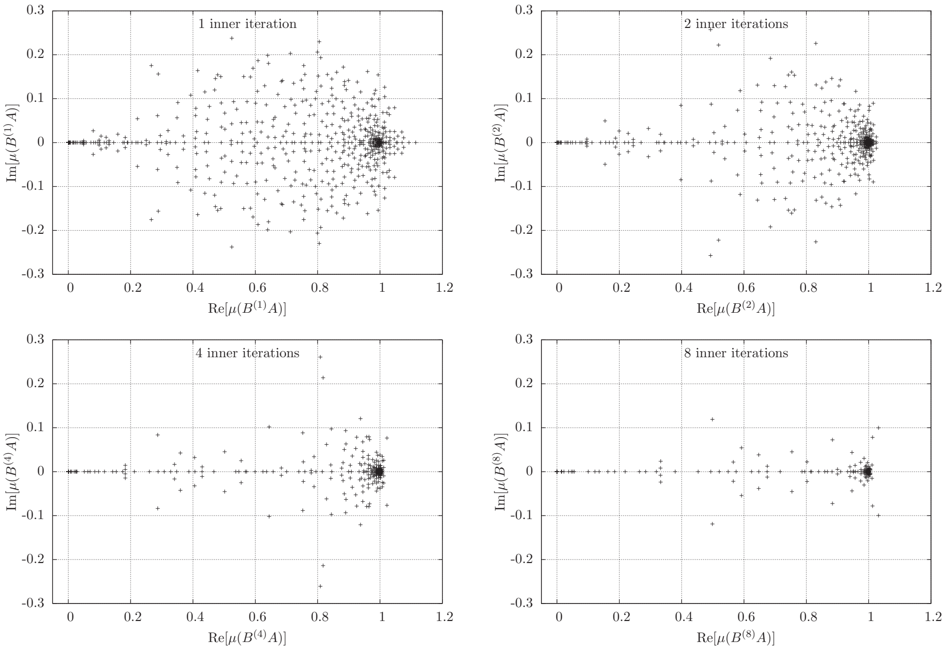

FIG. 4.1. *Spectrum of the preconditioned matrix $B^{(\ell)}A$ with NR-SOR inner iterations for Maragal_3. Upper left: $\ell = 1$, right upper: $\ell = 2$, left lower: $\ell = 4$, right lower: $\ell = 8$.*

REFERENCES

[1] Å. Björck and T. Elfving, *Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations*, BIT, 19 (1979), pp. 145–163.

[2] P. N. Brown and H. F. Walker, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.

[3] T. Davis, *The University of Florida Sparse Matrix Collection*, available online at `http://www.cise.ufl.edu/research/sparse/matrices/`, The University of Florida.

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

[5] K. Hayami and M. Sugihara, *A geometric view of Krylov subspace methods on singular systems*, Numer. Linear Algebra Appl., 18 (2011), pp. 449–469.

[6] K. Hayami, J.-F. Yin, and T. Ito, *GMRES methods for least squares problems*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2400–2430.

[7] V. K. Hensel, *Über potenzreihen von matrizen*, J. Reine Angew. Math., 155 (1926), pp. 107–110 (in German).

[8] 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.

[9] K. Morikuni and K. Hayami, *Inner-iteration Krylov subspace methods for least squares problems*, NII Technical Report, NII-2011-001E (2011), pp. 1–27 (revised, submitted).

[10] R. Oldenburger, *Infinite powers of matrices and characteristic roots*, Duke Math., 6 (1940), pp. 357–361.

[11] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.

[12] 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.

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