ISSN 1346-5597

National Institute of Informatics

NII Technical Report

On the Behaviour of the Conjugate Residual Method for Singular Systems

Ken HAYAMI

NII-2001-002E
July 2001

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

On the Behaviour of the Conjugate Residual Method for Singular Systems *

by Ken HAYAMI

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

Abstract

Consider applying the Conjugate Residual (CR) method, which is a Krylov subspace type iterative solver, to systems of linear equations $A\mathbf{x} = \mathbf{b}$ or least squares problems $\min_{\mathbf{x}\in\mathbf{R}^n} \|\mathbf{b} - A\mathbf{x}\|_2$, where A is singular and nonsymmetric.

We will show that when $R(A)^\perp = \ker A$, the CR method can be decomposed into the $R(A)$ and $\ker A$ components, and the necessary and sufficient condition for the CR method to converge to the least squares solution without breaking down for arbitrary $\mathbf{b}$ and initial approximate solution $\mathbf{x}_0$, is that the symmetric part $M(A)$ of A is semi-definite and $\mathrm{rank}\, M(A) = \mathrm{rank} A$. Furthermore, when $\mathbf{x}_0 \in R(A)$, the approximate solution converges to the pseudo inverse solution.

Next, we will also derive the necessary and sufficient condition for the CR method to converge to the least squares solution without breaking down for arbitrary initial approximate solutions, for the case when $R(A) \oplus \ker A = \mathbf{R}^n$ and $\mathbf{b} \in R(A)$.

Keywords: system of linear equations, iterative solver, singular system, Krylov subspace method, conjugate residual method

AMS Subject classification: Primary 65F10, Secondary 65N06

*Presented as an invited talk at the Fifth China-Japan Joint Seminar on Numerical Mathematics, Shanghai Jiaotong University, August 21-25, 2000.

1 Introduction

Consider the system of simultaneous linear equations

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{1.1}$$

where $A \in \mathbf{R}^{n\times n}$, $\boldsymbol{x}, \boldsymbol{b} \in \mathbf{R}^n$, which arises, for instance, in the discrete approximation of partial differential equations.

When A is nonsymmetric, there are Krylov subspace type iterative solvers for (1.1) based on biorthogonality, such as the Bi-CG[4] and Bi-CGSTAB[8] methods. There are also methods based on minimizing the residual $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$, such as the Conjugate Residual (CR) method[3], Generalized Conjugate Residual (GCR) method[3] and Generalized Minimum Residual (GMRES) method[7]. When the coefficient matrix A is regular, the behaviour of these methods is well understood[3, 6, 7].

On the other hand, in the discrete approximation of partial differential equations, the coefficient matrix of the resulting system of linear equations may be singular, depending on the boundary condition, for instance, when Neumann boundary conditions are imposed on the whole boundary. The computation of stationary probability vectors of stochastic matrices in the analysis of queuing networks also gives rise to singular systems[5]. For such singular systems, the system (1.1) may not always have a solution, so it is generally more appropriate to consider the least squares problem $\min_{\boldsymbol{x}\in\mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$.

When the system is singular, methods based on biorthogonality may diverge, and one has to modify the system in order to guarantee convergence. On the other hand, for methods based on minimizing the residual, by principle, the residual is expected to decrease monotonically without such modifications[1].

There are studies on the behaviour of methods based on residual minimization for singular systems, such as [2] for the GMRES method, [9] for the Orthomin(k) method, and [1] for the CR method. In this paper, we will modify and extend the analysis of the CR method for singular systems done in [1].

The following notations will be used.

$V^\perp$: Orthogonal complement of subspace V of $\mathbf{R}^n$.

$V \oplus W$: The direct sum of subspace V and subspace W.

For $X \in \mathbf{R}^{n\times n}$,

$R(X)$: the range space of X, i.e. the subspace spanned by the

column vectors of X,
$\ker X$: the kernel of X, i.e. the subspace of vectors $\boldsymbol{v} \in \mathbf{R}^n$ such that $X\boldsymbol{v} = \boldsymbol{0}$,
$M(X) := \dfrac{X + X^{\mathrm{T}}}{2}$: the symmetric part of X,
$\lambda_{\min}(X)$: the eigenvalue of X with minimum absolute value,
$\lambda^{+}_{\min}(X)$: the nonzero eigenvalue of X with minimum absolute value,
$\lambda_{\max}(X)$: the eigenvalue of X with maximum absolute value.

2 The CR method and its convergence for regular systems

For the system of linear equations

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{2.1}$$

where A is a real and regular (but not necessarily symmetric) $n \times n$ matrix, $\boldsymbol{b} \in \mathbf{R}^n$ is the right hand side, and $\boldsymbol{x} \in \mathbf{R}^n$ is the solution, the CR method[3] is given as follows.

The CR method

Choose $\boldsymbol{x}_0$.
$\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
$\boldsymbol{p}_0 = \boldsymbol{r}_0$

For $i = 0, 1, \ldots$ until the residual ($\boldsymbol{r}$) converges, do

$$\begin{aligned}
&\alpha_i = \frac{(\boldsymbol{r}_i, A\boldsymbol{p}_i)}{(A\boldsymbol{p}_i, A\boldsymbol{p}_i)} \\
&\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i\, \boldsymbol{p}_i \\
&\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \alpha_i A\boldsymbol{p}_i \\
&\beta_i = -\frac{(A\boldsymbol{r}_{i+1}, A\boldsymbol{p}_i)}{(A\boldsymbol{p}_i, A\boldsymbol{p}_i)} \\
&\boldsymbol{p}_{i+1} = \boldsymbol{r}_{i+1} + \beta_i\, \boldsymbol{p}_i\,.
\end{aligned} \tag{2.2}$$

First, note the following.

Lemma 2.1 If the symmetric part $M(A)$ of a matrix A is definite, then the matrix A is regular. □

Theorem 2.2 [3, 6] If the symmetric part $M(A)$ of A is definite, either of the following holds.

1. There exists $l \geq 0$, such that $\boldsymbol{p}_i \neq \boldsymbol{0}\,(0 \leq i < l)$ and $\boldsymbol{r}_l = \boldsymbol{0}$. Further, for $0 \leq i < l$,
$$\frac{\|\boldsymbol{r}_{i+1}\|_2{}^2}{\|\boldsymbol{r}_i\|_2{}^2} \leq 1 - \frac{\{\lambda_{\min}(M(A))\}^2}{\lambda_{\max}(A^{\mathrm{T}}A)} \tag{2.3}$$
holds.
2. For all $i \geq 0$, $\boldsymbol{p}_i \neq \boldsymbol{0}$, $\boldsymbol{r}_i \neq \boldsymbol{0}$ and (2.3) hold. □

Lemma 2.3 $M(A) := \dfrac{A + A^{\mathrm{T}}}{2}$ is not definite $\Longrightarrow {}^{\exists}\boldsymbol{v} \neq \boldsymbol{0}\,;\ (\boldsymbol{v}, A\boldsymbol{v}) = 0$. □

From Theorem 2.2 and Lemma 2.3, we obtain the following theorem.

Theorem 2.4 [1] If A is regular, C1–C3 are equivalent.

(C1) For arbitrary $\boldsymbol{x}_0$, the CR method converges without breakdown.

(C2) For arbitrary $\boldsymbol{x}_0$, the CR method does not break down.

(C3) The symmetric part $M(A)$ of A is definite. □

Here, 'breakdown' refers to the situation where the denominator $(A\boldsymbol{p}_i, A\boldsymbol{p}_i)$ of α_i in the CR method becomes 0 and it becomes impossible to continue the computation.

3 The convergence of the CR method for singular systems

3.1 Decomposition of the CR method

Next, modifying the arguments in [1], we will consider the convergence of the CR method for the least squares problem

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2\,, \tag{3.1}$$

where A is an $n \times n$ matrix which is not necessarily regular and $\boldsymbol{b} \in \mathbf{R}^n$.

In the following, let $\mathrm{rank} A = \dim R(A) = r > 0$, and

$\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r$: the orthonormal basis of $R(A)$,

$\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$: the orthonormal basis of $R(A)^\perp$,

$Q_1 := [\,\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r] : n \times r$ matrix,

$Q_2 := [\,\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n] : n \times (n-r)$ matrix,

$Q := [\,Q_1, Q_2] : n \times n$ orthogonal matrix.

Hence, $Q^{\mathrm{T}} Q = Q Q^{\mathrm{T}} = \mathrm{I}_n$ (I_n : identity matrix of order n).

(3.2)

Then, the orthogonal transformation of A gives

$$Q^{\mathrm{T}} A Q = \begin{bmatrix} {Q_1}^{\mathrm{T}} A Q_1 & {Q_1}^{\mathrm{T}} A Q_2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix},$$

and the following hold.

Theorem 3.1 $A_{11} = {Q_1}^{\mathrm{T}} A Q_1$: regular $\Longleftrightarrow$ $R(A)^\perp \oplus \ker A = \mathbf{R}^n$. □

Theorem 3.2 $A_{12} = {Q_1}^{\mathrm{T}} A Q_2 = 0 \Longleftrightarrow R(A)^\perp = \ker A$. □

Corollary 3.3 $R(A)^\perp = \ker A \Longrightarrow A_{11}$: regular . □

Lemma 3.4 $R(A)^\perp = \ker A^{\mathrm{T}}$. □

Lemma 3.5 The following (1), (2), (3), (4) are equivalent.

(1) $R(A)^\perp = \ker A$,

(2) $A_{12} = 0$,

(3) $\ker A^{\mathrm{T}} = \ker A$,

(4) $R(A^{\mathrm{T}}) = R(A)$. □

Hence, if and only if $R(A)^\perp = \ker A$ holds,

$$\tilde{A} = Q^{\mathrm{T}} A Q = \begin{bmatrix} A_{11} & 0 \\ 0 & 0 \end{bmatrix} \tag{3.3}$$

holds. Then, $A_{11} = {Q_1}^{\mathrm{T}} A Q_1$ is regular.

Using (3.2), the vectors $\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{b}, \boldsymbol{r}$ (the subscripts are abbreviated) used in the CR method (2.2) can be decomposed in to the $R(A)$ component and the $R(A)^{\perp} = \ker A$ component as

$$\tilde{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} {Q_1}^{\mathrm{T}}\boldsymbol{x} \\ {Q_2}^{\mathrm{T}}\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}^1 \\ \boldsymbol{x}^2 \end{bmatrix}, \qquad \tilde{\boldsymbol{p}} = Q^{\mathrm{T}}\boldsymbol{p} = \begin{bmatrix} {Q_1}^{\mathrm{T}}\boldsymbol{p} \\ {Q_2}^{\mathrm{T}}\boldsymbol{p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}^1 \\ \boldsymbol{p}^2 \end{bmatrix}, \tag{3.4}$$

$$\tilde{\boldsymbol{b}} = Q^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} {Q_1}^{\mathrm{T}}\boldsymbol{b} \\ {Q_2}^{\mathrm{T}}\boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}^1 \\ \boldsymbol{b}^2 \end{bmatrix}, \qquad \tilde{\boldsymbol{r}} = Q^{\mathrm{T}}\boldsymbol{r} = \begin{bmatrix} {Q_1}^{\mathrm{T}}\boldsymbol{r} \\ {Q_2}^{\mathrm{T}}\boldsymbol{r} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}^1 \\ \boldsymbol{r}^2 \end{bmatrix}. \tag{3.5}$$

Then, under the condition: $R(A)^{\perp} = \ker A$, the CR method can be decomposed into the $R(A)$ component and the $\ker A$ component as follows.

The decomposed CR method

Choose$\boldsymbol{x}_0$

$R(A)$ component	$\ker A$ component
$\boldsymbol{b}^1 = {Q_1}^{\mathrm{T}}\boldsymbol{b}$	$\boldsymbol{b}^2 = {Q_2}^{\mathrm{T}}\boldsymbol{b}$
$\boldsymbol{x}_0^1 = {Q_1}^{\mathrm{T}}\boldsymbol{x}_0$	$\boldsymbol{x}_0^2 = {Q_2}^{\mathrm{T}}\boldsymbol{x}_0$
$\boldsymbol{r}_0^1 = \boldsymbol{b}^1 - A_{11}\boldsymbol{x}_0^1$	$\boldsymbol{r}_0^2 = \boldsymbol{b}^2$
$\boldsymbol{p}_0^1 = \boldsymbol{r}_0^1$	$\boldsymbol{p}_0^2 = \boldsymbol{b}^2$

For $i = 0, 1, \ldots$ until the $R(A)$ component $(\boldsymbol{r}^1)$ of the residual converges, do

$\alpha_i = \dfrac{(\boldsymbol{r}_i^1,\, A_{11}\boldsymbol{p}_i^1)}{(A_{11}\boldsymbol{p}_i^1,\, A_{11}\boldsymbol{p}_i^1)}$	
$\boldsymbol{x}_{i+1}^1 = \boldsymbol{x}_i^1 + \alpha_i \boldsymbol{p}_i^1$	$\boldsymbol{x}_{i+1}^2 = \boldsymbol{x}_i^2 + \alpha_i \boldsymbol{p}_i^2$
$\boldsymbol{r}_{i+1}^1 = \boldsymbol{r}_i^1 - \alpha_i A_{11}\boldsymbol{p}_i^1$	$\boldsymbol{r}_{i+1}^2 = \boldsymbol{b}^2$
$\beta_i = -\dfrac{(A_{11}\boldsymbol{r}_{i+1}^1,\, A_{11}\boldsymbol{p}_i^1)}{(A_{11}\boldsymbol{p}_i^1,\, A_{11}\boldsymbol{p}_i^1)}$	
$\boldsymbol{p}_{i+1}^1 = \boldsymbol{r}_{i+1}^1 + \beta_i \boldsymbol{p}_i^1$	$\boldsymbol{p}_{i+1}^2 = \boldsymbol{b}^2 + \beta_i \boldsymbol{p}_i^2$.

(3.6)

The $R(A)$ component of the above algorithm can be regarded as the CR method applied to the system of linear equations $A_{11}\boldsymbol{x}^1 = \boldsymbol{b}^1$. Hence, from Theorem 2.2, we obtain the following lemma.

Lemma 3.6 If the symmetric part $M(A_{11})$ of $A_{11} = {Q_1}^{\mathrm{T}} A Q_1$ is definite, either of the following holds.

1. There exists $l \geq 0$, such that $\boldsymbol{p}_i^1 \neq \boldsymbol{0}\,(0 \leq i < l)$, $\boldsymbol{r}_l^1 = \boldsymbol{0}$. Further, for $0 \leq i < l$,
$$\frac{\|\boldsymbol{r}_{i+1}^1\|_2{}^2}{\|\boldsymbol{r}_i^1\|_2{}^2} \leq 1 - \frac{\{\lambda_{\min}(M(A_{11}))\}^2}{\lambda_{\max}(A_{11}^{\mathrm{T}} A_{11})} \tag{3.7}$$
holds.

2. For all $i \geq 0$, $\boldsymbol{p}_i^1 \neq \boldsymbol{0}$, $\boldsymbol{r}_i^1 \neq \boldsymbol{0}$ and (3.7) hold. □

3.2 Convergence theorem for the case $R(A)^{\perp} = \ker A$

Next, we extend Lemma 3.6 to obtain the convergence theorem for the CR method for singular systems, similarly to [1], but with modifications. First note the following.

Lemma 3.7 If $R(A)^{\perp} = \ker A$, then $\boldsymbol{v}^1 := {Q_1}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{0} \Longleftrightarrow A\boldsymbol{v} = \boldsymbol{0}$. □

Lemma 3.8 If $R(A)^{\perp} = \ker A$, and $M(A_{11})$ is regular,
then $\lambda_{\min}(M(A_{11})) = \lambda_{\min}^{+}(M(A))$. □

Lemma 3.9 If $R(A)^{\perp} = \ker A$, then
$M(A_{11})$: definite $\Longleftrightarrow$ '$M(A)$: semi-definite, $\operatorname{rank} M(A) = \operatorname{rank} A$ '. □

Lemma 3.10 If $R(A)^{\perp} = \ker A$, then $\lambda_{\max}({A_{11}}^{\mathrm{T}} A_{11}) = \lambda_{\max}(A^{\mathrm{T}} A)$. □

From Lemmas 3.6, 3.7, 3.8 and 3.10, we obtain the following theorem.

Theorem 3.11 If $R(A)^{\perp} = \ker A$, and the symmetric part $M(A_{11})$ of $A_{11} := {Q_1}^{\mathrm{T}} A Q_1$ is definite, either of the following holds.

1. There exists $l \geq 0$ such that, $A\boldsymbol{p}_i \neq \boldsymbol{0}\,(0 \leq i < l)$ and $A\boldsymbol{r}_l = \boldsymbol{0}$. This means that the least squares solution has been obtained. Further, for $0 \leq i < l$,
$$\frac{\|\boldsymbol{r}_{i+1}\|_2{}^2 - \min\limits_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2{}^2}{\|\boldsymbol{r}_i\|_2{}^2 - \min\limits_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2{}^2} \leq 1 - \frac{\{\lambda_{\min}^{+}(M(A))\}^2}{\lambda_{\max}(A^{\mathrm{T}} A)} \tag{3.8}$$
holds.

2. For all $i \geq 0$, $A\boldsymbol{p}_i \neq \boldsymbol{0}$, $A\boldsymbol{r}_i \neq \boldsymbol{0}$ and (3.8) hold. □

Further, from Lemma 2.3 and Theorem 3.11, we obtain the following theorem.

Theorem 3.12 If $R(A)^{\perp} = \ker A$, C1–C4 are equivalent.

(C1) For arbitrary $\boldsymbol{x}_0$, the CR method does not break down, and the $R(A)$ component of the residual converges to $\mathbf{0}$.

(C2) For arbitrary $\boldsymbol{x}_0$, the CR method does not break down.

(C3) The symmetric part $M(A_{11})$ of $A_{11} := {Q_1}^{\mathrm{T}} A Q_1$ is definite.

(C4) $M(A)$ is semi-definite and $\operatorname{rank} M(A) = \operatorname{rank} A$. □

As to where the approximate solution of the CR method converges, the following theorem holds.

Theorem 3.13 If $R(A)^{\perp} = \ker A$ and the $R(A)$ component of the residual converges to $\mathbf{0}$ (least squares solution), the $R(A)$ component $\boldsymbol{x}_i^1$ of the approximate solution $\boldsymbol{x}_i$ converges to ${A_{11}}^{-1}\boldsymbol{b}^1$.

Moreover, if $\boldsymbol{b} \in R(A)$, the $\ker A$ component $\boldsymbol{x}_i^2$ of $\boldsymbol{x}_i$ is always equal to $\boldsymbol{x}_0^2$. Then, the approximate solution $\boldsymbol{x}_i$ converges to
$Q_1 {A_{11}}^{-1} Q_1^{\mathrm{T}} \boldsymbol{b} + Q_2 Q_2^{\mathrm{T}} \boldsymbol{x}_0$.

Further, if the $\ker A$ component of $\boldsymbol{x}_0$: $\boldsymbol{x}_0^2 = \mathbf{0}$ (i.e. $\boldsymbol{x}_0 \in R(A)$), $\boldsymbol{x}_i$ converges to the least squares solution with minimum norm (pseudo inverse solution) $Q_1 A_{11}^{-1} Q_1^{\mathrm{T}} \boldsymbol{b}$.

(Here, $A_{11} := Q_1^{\mathrm{T}} A Q_1$, $\boldsymbol{b}^1 := Q_1^{\mathrm{T}} \boldsymbol{b}$, $\boldsymbol{x}_0^2 := Q_2^{\mathrm{T}} \boldsymbol{x}_0$, where $\boldsymbol{x}_0$ is the initial approximate solution.) □

As an example, consider the ordinary differential equation

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + \beta \frac{\mathrm{d}u}{\mathrm{d}x} = f(x) \qquad (\,0 < x < 1\,) \tag{3.9}$$

with periodic boundary condition: $u(0) = u(1)$. The singular system of linear equations obtained by discretizing this problem using an equi-distant central finite difference scheme satisfies $R(A)^{\perp} = \ker A$ and the condition (C4) in the above theorem. Further, if the source term f is appropriately chosen, $\boldsymbol{b} \in R(A)$ will also be satisfied.

3.3 Convergence theorem for the case $R(A)^\perp \neq \ker A$

Next, as an extension of [1], we consider the case when $R(A)^\perp \neq \ker A$. In this case, from Theorem 3.2, $A_{12} = 0$ does not hold even when the orthogonal transformation $Q^{\mathrm{T}}AQ$ is used. Thus, the CR method cannot be decomposed into the $R(A)$ and $R(A)^\perp$ components even if the transformation $Q^{\mathrm{T}}AQ$ is used.

Hence, for the general case when $R(A)^\perp = \ker A$ does not necessarily hold, besides $Q = [Q_1, Q_2]$, let

$\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r$: orthonormal basis of $(\ker A)^\perp$,
$\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_n$: orthonormal basis of $\ker A$,
$U_1 := [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r] : n \times r$ matrix,
$U_2 := [\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_n] : n \times (n-r)$ matrix,
$U := [U_1, U_2] : n \times n$ orthogonal matrix.

(Note that $\dim(\ker A)^\perp = \dim R(A) = r$, from the dimension theorem.)

Then, we have

Lemma 3.14

$$\tilde{A}' := Q^{\mathrm{T}}AU = \begin{bmatrix} {Q_1}^{\mathrm{T}}AU_1 & 0 \\ 0 & 0 \end{bmatrix}. \quad \square$$

Lemma 3.15 $A'_{11} := {Q_1}^{\mathrm{T}}AU_1$ is a $r \times r$ regular matrix. $\square$

Thus, let the vectors $\boldsymbol{b}$ and $\boldsymbol{r}$ used in the CR method (2.2) be decomposed into the $R(A)$ and $R(A)^\perp$ components as in (3.5). On the other hand, let $\boldsymbol{x}$ and $\boldsymbol{p}$ be decomposed into the $(\ker A)^\perp$ and $\ker A$ components as

$$\tilde{\boldsymbol{x}} = U^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} {U_1}^{\mathrm{T}}\boldsymbol{x} \\ {U_2}^{\mathrm{T}}\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}^1 \\ \boldsymbol{x}^2 \end{bmatrix}, \qquad \tilde{\boldsymbol{p}} = U^{\mathrm{T}}\boldsymbol{p} = \begin{bmatrix} {U_1}^{\mathrm{T}}\boldsymbol{p} \\ {U_2}^{\mathrm{T}}\boldsymbol{p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}^1 \\ \boldsymbol{p}^2 \end{bmatrix}. \tag{3.10}$$

Then, the CR method can be partially decomposed as follows.

Partially decomposed CR method

(For the general case including $R(A)^{\perp} \neq \ker A$.)

Choose $\boldsymbol{x}_0$.

$$\begin{aligned}
&\boldsymbol{b}^1 = {Q_1}^{\mathrm{T}}\boldsymbol{b} && \boldsymbol{b}^2 = {Q_2}^{\mathrm{T}}\boldsymbol{b} \\
&\boldsymbol{x}_0^1 = {U_1}^{\mathrm{T}}\boldsymbol{x}_0 && \boldsymbol{x}_0^2 = {U_2}^{\mathrm{T}}\boldsymbol{x}_0 \\
&\boldsymbol{r}_0^1 = \boldsymbol{b}^1 - A'_{11}\boldsymbol{x}_0^1 && \boldsymbol{r}_0^2 = \boldsymbol{b}^2 \\
&\boldsymbol{p}_0^1 = {U_1}^{\mathrm{T}}(Q_1\boldsymbol{r}_0^1 + Q_2\boldsymbol{b}^2) && \boldsymbol{p}_0^2 = {U_2}^{\mathrm{T}}(Q_1\boldsymbol{r}_0^1 + Q_2\boldsymbol{b}^2)
\end{aligned}$$

For $i = 0, 1, \ldots$ until the $R(A)$ component $(\boldsymbol{r}^1)$ of the residual converges, do

$$\alpha_i = \frac{(\boldsymbol{r}_i^1,\, A'_{11}\boldsymbol{p}_i^1)}{(A'_{11}\boldsymbol{p}_i^1,\, A'_{11}\boldsymbol{p}_i^1)}$$

$$\begin{aligned}
&\boldsymbol{x}_{i+1}^1 = \boldsymbol{x}_i^1 + \alpha_i\boldsymbol{p}_i^1 && \boldsymbol{x}_{i+1}^2 = \boldsymbol{x}_i^2 + \alpha_i\boldsymbol{p}_i^2 \\
&\boldsymbol{r}_{i+1}^1 = \boldsymbol{r}_i^1 - \alpha_i A'_{11}\boldsymbol{p}_i^1 && \boldsymbol{r}_{i+1}^2 = \boldsymbol{b}^2
\end{aligned}$$

$$\beta_i = -\frac{(A'_{11}{U_1}^{\mathrm{T}}(Q_1\boldsymbol{r}_{i+1}^1 + Q_2\boldsymbol{b}^2),\, A'_{11}\boldsymbol{p}_i^1)}{(A'_{11}\boldsymbol{p}_i^1,\, A'_{11}\boldsymbol{p}_i^1)}$$

$$\boldsymbol{p}_{i+1}^1 = {U_1}^{\mathrm{T}}(Q_1\boldsymbol{r}_{i+1}^1 + Q_2\boldsymbol{b}^2) + \beta_i\boldsymbol{p}_i^1 \qquad \boldsymbol{p}_{i+1}^2 = {U_2}^{\mathrm{T}}(Q_1\boldsymbol{r}_{i+1}^1 + Q_2\boldsymbol{b}^2) + \beta_i\boldsymbol{p}_i^2 \tag{3.11}$$

In the left side of the above algorithm, there are terms involving $\boldsymbol{b}^2$ in the computation of β_i and $\boldsymbol{p}_{i+1}^1$, which makes the algorithm complex and not completely decomposed. However, if we consider the case when $\boldsymbol{b} \in R(A)$, then we have $\boldsymbol{b}^2 = {Q_2}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{0}$, and the algorithm becomes more decomposed and simple.

Now note the following Lemma.

Lemma 3.16 ${U_1}^{\mathrm{T}}Q_1$: regular $\iff R(A)^{\perp} \oplus \ker A = \mathbf{R}^n$. $\square$

Further, in order to analyze the convergence of the decomposed algorithm, let us concentrate on the left side of the above algorithm, and for simplicity redefine $A := A'_{11} = {Q_1}^{\mathrm{T}}AU_1$, $B := {U_1}^{\mathrm{T}}Q_1 \in \mathbf{R}^{r\times r}$ and $\boldsymbol{x} := \boldsymbol{x}^1$, $\boldsymbol{p} := \boldsymbol{p}^1$, $\boldsymbol{r} := \boldsymbol{r}^1 \in \mathbf{R}^r$. Then, we have the following.

Lemma 3.17

(1) $(A\boldsymbol{p}_i, A\boldsymbol{p}_{i-1}) = 0 \quad (i \geq 1)$,

(2) $(A\boldsymbol{p}_i, A\boldsymbol{p}_i) \leq (AB\boldsymbol{r}_i, AB\boldsymbol{r}_i) \quad (i \geq 0)$,
(3) $(\boldsymbol{r}_i, A\boldsymbol{p}_{i-1}) = 0 \quad (i \geq 1)$,
(4) $(\boldsymbol{r}_i, A\boldsymbol{p}_i) = (\boldsymbol{r}_i, AB\boldsymbol{r}_i) \quad (i \geq 0)$. □

Lemma 3.18

$$\frac{\|\boldsymbol{r}_{i+1}\|_2{}^2}{\|\boldsymbol{r}_i\|_2{}^2} \leq 1 - \left\{ \frac{(\boldsymbol{r}_i, C\boldsymbol{r}_i)}{(\boldsymbol{r}_i, \boldsymbol{r}_i)} \right\}^2 \frac{(\boldsymbol{r}_i, \boldsymbol{r}_i)}{(C\boldsymbol{r}_i, C\boldsymbol{r}_i)},$$

where $C := AB$. □

Lemma 3.19

$$\frac{\|\boldsymbol{r}_{i+1}\|_2{}^2}{\|\boldsymbol{r}_i\|_2{}^2} \leq 1 - \frac{\lambda_{\min}(M)^2}{\lambda_{\max}(C^{\mathrm{T}}C)},$$

where $M := \dfrac{C + C^{\mathrm{T}}}{2}$. □

From the above, we obtain the following convergence theorem of the CR method for the case $R(A) \oplus \ker A = \mathbf{R}^n$, $\boldsymbol{b} \in R(A)$. (Note that, $M(C)$: definite $\Longrightarrow C$: regular $\Longrightarrow B$: regular $\Longleftrightarrow R(A) \oplus \ker A = \mathbf{R}^n$.)

Theorem 3.20 Let $A'_{11} = {Q_1}^{\mathrm{T}} A U_1$, $B := {U_1}^{\mathrm{T}} Q_1$, $C := A'_{11} B$, $M(C) := \dfrac{C + C^{\mathrm{T}}}{2}$. Then, if $M(C)$ is definite and $\boldsymbol{b} \in R(A)$, either of the following holds for the partially decomposed CR method (3.11).

1. There exists $l \geq 0$ such that, $\boldsymbol{p}_i^1 \neq \mathbf{0}$ $(0 \leq i < l)$ and $\boldsymbol{r}_l^1 = \mathbf{0}$. Further, for $0 \leq i < l$,
$$\frac{\|\boldsymbol{r}_{i+1}^1\|_2{}^2}{\|\boldsymbol{r}_i^1\|_2{}^2} \leq 1 - \frac{\{\lambda_{\min}(M(C))\}^2}{\lambda_{\max}(C^{\mathrm{T}}C)} \tag{3.12}$$
holds.
2. For all $i \geq 0$, $\boldsymbol{p}_i^1 \neq \mathbf{0}$, $\boldsymbol{r}_i^1 \neq \mathbf{0}$ and (3.12) hold. □

Finally, we obtain the following convergence theorem for singular systems.

Theorem 3.21 If $R(A) \oplus \ker A = \mathbf{R}^n$ and $\boldsymbol{b} \in R(A)$ hold, C1–C3 are equivalent.

(C1) For arbitrary $\boldsymbol{x}_0$, the CR method does not break down, and the $R(A)$ component of the residual converges to $\mathbf{0}$.

(C2) For arbitrary $\boldsymbol{x}_0$, the CR method does not break down.

(C3) $M(C)$ is definite,

where $M(C) := \dfrac{C + C^{\mathrm{T}}}{2}$, $C := A'_{11}B$, $A'_{11} = {Q_1}^{\mathrm{T}} A U_1$, $B := {U_1}^{\mathrm{T}} Q_1$. □

[**Remark**] When $R(A)^{\perp} = \ker A$, we can take $U_1 = Q_1$. Then, $B = U_1^{\mathrm{T}} Q_1 = Q_1^{\mathrm{T}} Q_1 = \mathrm{I}_r$, where I_r is the $r \times r$ identity matrix. Moreover, $A'_{11} = U_1^{\mathrm{T}} A Q_1 = Q_1^{\mathrm{T}} A Q_1 = A_{11}$, $C = A'_{11} B = A_{11}$, $M(C) = M(A_{11})$.

Note that if $R(A) \oplus \ker A = \mathbf{R}^n$ and $\boldsymbol{b} \in R(A)$ hold, and the $R(A)$ component of the residual converges to $\mathbf{0}$ in the CR method, according to the partially decomposed CR method (3.11) and Theorems 3.20, 3.21, the $(\ker A)^{\perp}$ component $\boldsymbol{x}_i^1$ of the approximate solution converges to ${A'_{11}}^{-1} \boldsymbol{b}^1$.

As an example, consider the ordinary differential equation (3.9) with Neumann boundary conditions at both ends: $\left.\dfrac{\mathrm{d}u}{\mathrm{d}x}\right|_{x=0} = \left.\dfrac{\mathrm{d}u}{\mathrm{d}x}\right|_{x=1} = 0$. Unlike the case for the periodic boundary condition, the singular system of equations obtained by equi-spaced central finite difference discretization of the problem does not satisfy the condition $R(A)^{\perp} = \ker A$. However, the system does satisfy the condition $R(A) \oplus \ker A = \mathbf{R}^n$ in the above theorem, and when the source term f is appropriately chosen, the condition $\boldsymbol{b} \in R(A)$ can also be satisfied. It is not easy to check whether the condition (C3) is satisfied in the general case, but for instance when the number of finite difference nodes is 3, it can be shown that $M(C)$ is negative-definite.

3.4 Summary of the convergence theorems

Finally, a summary of the theorems obtained on the convergence of the conjugate residual method for singular systems is given.

First, the inclusion relationship of the set of $n \times n$ real matrices A is given by

$$\mathbf{R}^{n\times n} \supset \{A \in \mathbf{R}^{n\times n} | R(A) \oplus \ker(A) = \mathbf{R}^n\} \supset \{A \in \mathbf{R}^{n\times n} | R(A)^{\perp} = \ker A\},$$

and the following hold.

$$R(A) \oplus \ker(A) = \mathbf{R}^n \iff B := U_1^{\mathrm{T}} Q_1 : \text{regular} \iff A_{11} := Q_1^{\mathrm{T}} A Q_1 : \text{regular}.$$

$$R(A)^{\perp} = \ker A \Longleftrightarrow A_{12} = 0.$$

Next, since $M(C)$: definite $\Longrightarrow$ C : regular $\Longrightarrow$ B : regular, we have $M(C)$: definite $\Longrightarrow$ $R(A) \oplus \ker(A) = \mathbf{R}^n$. Hence,

$$\{A \in \mathbf{R}^{n\times n} | M(C) : \text{definite}\} \subset \{A \in \mathbf{R}^{n\times n} | R(A) \oplus \ker(A) = \mathbf{R}^n\}.$$

Regarding convergence, if $R(A)^{\perp} = \ker A$, the symmetric part $M(A)$ of A is semi-definite, and $\operatorname{rank} M(A) = \operatorname{rank} A$, then, the CR method does not break down for arbitrary right hand side $\boldsymbol{b}$ and initial approximate solution $\boldsymbol{x}_0$, the $R(A)$ component of the residual converges to $\mathbf{0}$, and the least squares solution is obtained. Further, if $\boldsymbol{b} \in R(A)$, the $R(A)^{\perp} = \ker A$ component of the approximate solution remains the same. Moreover, if $\boldsymbol{x}_0 \in R(A)$, the approximate solution converges to the least squares solution with the minimum norm (pseudo inverse solution).

If $R(A) \oplus \ker A = \mathbf{R}^n$, $M(C)$: definite and $\boldsymbol{b} \in R(A)$, the CR method does not break down for arbitrary initial approximate solution $\boldsymbol{x}_0$, and the $R(A)$ component of the residual converges to $\mathbf{0}$, giving the least squares solution.

Acknowledgements

The author would like to thank Mr. Eiji Yatagawa and Professor Shao-Liang Zhang of the University of Tokyo, and Professor Kunio Tanabe of the Institute of Statistical Mathematics for valuable remarks.

This research was partly supported by the Grant-in-Aid for Scientific Research (Fundamental Research (C)) of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] K. Abe, H. Ogata, M. Sugihara, S.-L. Zhang and T. Mitsui, Convergence theory of the CR method for linear singular systems, Trans. Japan Soc. Ind. Appl. Maths. Vol.9 No. 1 (1999), 1–13 (in Japanese).

[2] P. Brown and H.F. Walker, GMRES on (nearly) singular systems, SIAM J. Matrix Anal. Appl. Vol. 18 No. 1 (1997), 37–51.

[3] S.C. Eisenstat, H.C. Elman and M.H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, SIAM J. Numer. Anal. Vol. 20 No. 2 (1983), 345–357.

[4] R. Fletcher, Conjugate gradient methods for indefinite systems, in: *Numerical Analysis Dundee 1975*, ed. G. Watson, Lecture Notes in Mathematics 506, Springer-Verlag, Berlin, 1976, pp. 73–89.

[5] R.W. Freund, R. W. and M. Hochbruck, On the use of two QMR algorithms for solving singular systems and applications in Markov chain modeling, Num. Lin. Alg. Appl. Vol. 1 No. 4 (1994), 403–420.

[6] M. Mori, M. Sugihara and K. Murota, *Linear Computations*, Iwanami, Tokyo, 1994 (in Japanese).

[7] Y. Saad and M.H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. Vol. 7 No. 3 (1986), 856–869.

[8] H.A. van der Vorst, Bi-CGSTAB: A more smoothly converging variant of CG-S for the solution of nonsymmetric linear systems, SIAM J. Sci. Statist. Comput. 13 (1992), 631–644.

[9] S.-L. Zhang, Y. Oyanagi and M. Sugihara, Necessary and sufficient conditions for the convergence of Orthomin(k) on inconsistent linear systems, Numer. Math. 87 (2000), 391–405.