Numerical Algorithms (2005) **40:** 201–216
DOI 10.1007/s11075-005-1533-3

Analysis of the convergence of the minimal and the orthogonal residual methods

H. Sadok

Laboratoire de Mathématiques Appliquées, Centre Universitaire de la Mi-voix, Batiment H. Poincaré, 50 rue F. Buisson, B.P. 699, F-62228 Calais Cedex, France
E-mail: sadok@lmpa.univ-littoral.fr

Received 1 June 2003; accepted 22 January 2005
Communicated by C. Brezinski

We consider two Krylov subspace methods for solving linear systems, which are the minimal residual method and the orthogonal residual method. These two methods are studied without referring to any particular implementations. By using the Petrov–Galerkin condition, we describe the residual norms of these two methods in terms of Krylov vectors, and the relationship between there two norms. We define the Ritz singular values, and prove that the convergence of these two methods is governed by the convergence of the Ritz singular values.

Keywords: GMRES, Krylov subspace methods, convergence analysis

AMS subject classification: 65F10

1. Introduction

In this paper we study the convergence behavior of the minimal residual method (MR) and the orthogonal residual method (OR) for solving linear systems of equations.

The MR method is widely used for solving large sparse linear systems and it can be efficiently implemented by the MINRES algorithm of Paige and Saunders [14], when the matrix is symmetric, and by the GMRES algorithm of Saad and Schultz [18] when the matrix is nonsymmetric.

The well known conjugate gradient method of Hestenes and Stiefel [7] is a powerful method for solving symmetric linear system see, for example, [8,14]. The residuals of this method are orthogonal. One of the generalization of this method to the nonsymmetric system which verifies this orthogonality property is Arnoldi's method, which can be efficiently implemented by the Full orthogonalization method of Saad [16].

A theoretical comparison of these two methods is given in [14] for the symmetric case and in [1,2,6,22] for the nonsymmetric case.

But most of the theoretical results about these methods are obtained from the algorithms themselves.

By using the Petrov–Galerkin condition [17], we give and prove some well known theoretical results as well as some new results about these two methods and show that their convergence behavior depends on the convergence of the singular Ritz values.

In this paper we present theoretical results on the MR and the OR methods. The next section establishes notations, and the methods are defined by using the Petrov–Galerkin conditions. In section 3 we begin by giving the norms of the MR and the OR methods in terms of determinants of Krylov matrices, then we analyze the QR factorization of these type of matrices. We show the relationship between the orthogonal and upper triangular matrix used in this factorization, and the orthogonal and the upper triangular Hessenberg matrices which characterize the algorithm of Arnoldi [16].

We define the Ritz singular values as the singular values of the upper Hessenberg matrix and we introduce the Ritz adjacent singular values. We give some inequalities relating these singular values. From these inequalities we deduce the important property, that is the smallest Ritz singular value is not bounded below in general, and we give two classes of matrices for which this fundamental property is satisfied. In section 4, the rate of convergence of the two methods is given in terms of Ritz singular values and Ritz adjacent singular values. Some examples are given for illustrating the importance of these results, for analyzing the convergence of these two methods. Finally, the last section is devoted to the symmetric case.

Throughout the paper, we assume exact arithmetic. For a vector v, $\|v\|$ denotes the Euclidean norm $\|v\| = \sqrt{(v^{\mathrm{T}}v)}$. For a matrix A, $\|A\|$ denotes the 2-norm and $\kappa(A)$ denotes the 2-condition number of A.

Moreover, we denote by I_n the $n \times n$ identity matrix and by $e_k^{(n)}$ its kth column.

2. Definitions and preliminaries

Consider the linear system of equations

$$Ax = b \tag{2.1}$$

with a nonsingular matrix $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$.

Let $x_0 \in \mathbb{R}^n$ be a given vector and r_0 its residual $r_0 = b - Ax_0$. The Krylov subspace $\mathcal{K}_k(A, r_0)$ is defined by

$$\mathcal{K}_k(A, r_0) = \mathrm{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}.$$

For both methods MR and OR, the iterates $\{x_k\}$ are defined by a so called Petrov–Galerkin condition

$$x_k - x_0 \in \mathcal{K}_k(A, r_0), \tag{2.2}$$

and

$$\left(ZA^i\, r_0, r_k\right) = 0 \quad \text{for } i = 0, \ldots, k-1, \tag{2.3}$$

where $r_k = b - Ax_k$, $Z = I$ for the OR method and $Z = A$ for the MR method.

Since $x_k - x_0 \in \mathcal{K}_k(A, r_0)$, we can write it as

$$x_0 - x_k = \sum_{j=1}^{k} a_j A^{j-1} r_0. \tag{2.4}$$

Thus, it follows that

$$r_k = r_0 + \sum_{j=1}^{k} a_j A^j r_0. \tag{2.5}$$

From the orthogonality condition (2.3), we obtain the system of linear equations

$$-\left(ZA^i r_0, r_0\right) = \sum_{j=1}^{k} a_j \left(ZA^i r_0, A^j r_0\right) \quad \text{for } i = 0, \ldots, k-1. \tag{2.6}$$

Let K_k be the Krylov matrix whose columns are $r_0, Ar_0, \ldots, A^{k-1}r_0$ and $W_k = AK_k$. The determinant of the system (2.6) is equal to $\det(K_k^{\mathrm{T}} Z^{\mathrm{T}} A K_k) = \det(K_k^{\mathrm{T}} Z^{\mathrm{T}} W_k)$.

If this determinant is different from zero, then x_k exists, and formulas (2.4) and (2.5) would allow us to obtain x_k and r_k.

3. Basic theory

3.1. Relationship between the two residual norms

Let r_k^{MR} and r_k^{OR} be the kth residuals for the MR method and the OR method, respectively. The norms of these residuals are given in the following:

Theorem 1.

(1) $$\left\|r_k^{\mathrm{MR}}\right\|^2 = \frac{\det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(W_k^{\mathrm{T}} W_k)},$$

(2) $$\left\|r_k^{\mathrm{OR}}\right\|^2 = \frac{\det(K_k^{\mathrm{T}} K_k)\det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(K_k^{\mathrm{T}} A K_k)^2} \quad \text{if} \quad \det\left(K_k^{\mathrm{T}} A K_k\right) \neq 0.$$

Proof. (1) Since $Z = A$, the orthogonality property (2.3) can be written as $(A^i r_0, r_k) = 0$ for $i = 1, \ldots, k$, then we get from (2.5)

$$\left(r_k^{\mathrm{MR}}, r_k^{\mathrm{MR}}\right) = \left(r_0 + \sum_{j=1}^{k} a_j A^j r_0, r_k\right) = (r_0, r_k) = \left(r_0, r_0 + \sum_{j=1}^{k} a_j A^j r_0\right).$$

By using Cramer rule for the system (2.6) we get

$$\left(r_k^{\mathrm{MR}}, r_k^{\mathrm{MR}}\right) = \frac{\det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(W_k^{\mathrm{T}} W_k)}.$$

(2) Similarly by using (2.3), we get

$$\left(r_k^{\mathrm{OR}}, r_k^{\mathrm{OR}}\right) = \left(r_0 + \sum_{j=1}^{k} a_j A^j r_0, r_k\right) = a_k\left(A^k r_0, r_k\right).$$

Using Cramer rule we have

$$a_k = \frac{(-1)^k \det(K_k^{\mathrm{T}} K_k)}{\det(K_k^{\mathrm{T}} A K_k)} \quad \text{and} \quad \left(A^k r_0, r_k\right) = \frac{(-1)^k \det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(K_k^{\mathrm{T}} A K_k)},$$

which completes the proof. □

Lemma 1.

$$(1) \quad \left\|r_k^{\mathrm{MR}}\right\|^2 = \frac{1}{e_1^H (K_{k+1}^{\mathrm{T}} K_{k+1})^{-1} e_1},$$

$$(2) \quad \left\|r_k^{\mathrm{OR}}\right\|^2 = \frac{e_{k+1}^H (K_{k+1}^{\mathrm{T}} K_{k+1})^{-1} e_{k+1}}{|e_{k+1}^H (K_{k+1}^{\mathrm{T}} K_{k+1})^{-1} e_1|^2}.$$

Now, since $r_0 = K_{k+1} e_1$ we have $\|r_0\|^2 = (r_0, r_0) = e_1^H K_{k+1}^H K_{k+1} e_1$. Consequently by using Kantorovich inequality we obtain:

Theorem 2.

$$1 \geqslant \frac{\|r_k^{\mathrm{MR}}\|}{\|r_0^{\mathrm{MR}}\|} \geqslant \frac{2\kappa(K_{k+1})}{1 + \kappa(K_{k+1})^2}.$$

This means there is no convergence as long as Krylov basis is well-conditioned.

Example 1. We consider the system with the $n \times n$ matrix

$$A_n = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 & 1 \\ 1 & \ddots & \ddots & \ddots & \vdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & \dots & 0 & 1 & 0 \end{pmatrix}$$

and $b = A(1, 0, \dots, 0)^{\mathrm{T}}$.

For this example, $\kappa(A_n) = 1$. Now, if $x_0 = 0$ then for $k = 1, \dots, n-1$, we have $K_k = [e_1, e_2, \dots, e_k]$, $\kappa(K_k) = 1$, and $\|r_k^{\mathrm{MR}}\| = 1$. Hence we obtain the solution at the nth iteration.

If now, $b = A(1, -1, 0, \dots, 0)^{\mathrm{T}}$.

Then

$$K_{k+1}^H K_{k+1} = \begin{pmatrix} 2 & -1 & \dots & \dots & 0 & 0 \\ -1 & 2 & -1 & \ddots & \vdots & 0 \\ 0 & -1 & 2 & -1 & \vdots & \vdots \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots \\ \vdots & \ddots & \ddots & -1 & 2 & -1 \\ & \dots & \dots & 0 & -1 & 2 \end{pmatrix},$$

$$\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_0^{\mathrm{MR}}\|^2} = \frac{k+2}{2(k+1)},$$

$$\frac{\|r_k^{\mathrm{OR}}\|^2}{\|r_0^{\mathrm{OR}}\|^2} = \frac{(k+2)(k+1)}{2}.$$

Example 2.

$$S = \begin{pmatrix} t & 1 & & & \\ & t & 1 & 0 & \\ & & \ddots & \ddots & \\ & 0 & & \ddots & 1 \\ s & & & & t \end{pmatrix}.$$

If $t = 0$, $b = (1, \dots, 1)^{\mathrm{T}}$ and $x_0 = 0$, we obtain:

$$\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_0^{\mathrm{MR}}\|^2} = \frac{(n-k)(s-1)^2}{n(n-k+s^2)},$$

$$\frac{\|r_k^{\mathrm{OR}}\|^2}{\|r_0^{\mathrm{OR}}\|^2} = \frac{(n-k+1)(n-k)(s-1)^2}{ns^2}.$$

If $s = 0$, $b = (1, 0, \dots, 0, 1)^{\mathrm{T}}$ and $x_0 = 0$, we obtain:

$$\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_0^{\mathrm{MR}}\|^2} = \frac{1}{-1 + 2(1 - t^{2(k+1)})/(1-t^2)}.$$

The relationship between these two residual norms is given in the following theorem.

Theorem 3.

$$\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_{k-1}^{\mathrm{MR}}\|^2} = s_k^2 \quad \text{and} \quad \left\|r_k^{\mathrm{MR}}\right\|^2 = c_k^2 \left\|r_k^{\mathrm{OR}}\right\|^2,$$

where

$$c_k^2 = 1 - s_k^2 = \frac{\det(K_k^{\mathrm{T}} A K_k)^2}{\det(W_k^{\mathrm{T}} W_k)\det(K_k^{\mathrm{T}} K_k)}.$$

Proof. By applying the Sylvester's identity [5] to the determinant $\det(K_{k+1}^{\mathrm{T}} K_{k+1})$ we obtain

$$\det\bigl(K_{k+1}^{\mathrm{T}} K_{k+1}\bigr) \det\bigl(W_{k-1}^{\mathrm{T}} W_{k-1}\bigr) = \det\bigl(K_k^{\mathrm{T}} K_k\bigr) \det\bigl(W_k^{\mathrm{T}} W_k\bigr) - \det\bigl(K_k^{\mathrm{T}} A K_k\bigr)^2.$$

Consequently

$$\begin{aligned}
&\frac{\det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(W_k^{\mathrm{T}} W_k)} \\
&\quad = \frac{\det(K_k^{\mathrm{T}} K_k)}{\det(W_{k-1}^{\mathrm{T}} W_{k-1})} - \frac{\det(K_k^{\mathrm{T}} A K_k)^2}{\det(W_k^{\mathrm{T}} W_k) \det(W_{k-1}^{\mathrm{T}} W_{k-1})} \\
&\quad = \frac{\det(K_k^{\mathrm{T}} K_k)}{\det(W_{k-1}^{\mathrm{T}} W_{k-1})} \left\{ 1 - \frac{\det(K_k^{\mathrm{T}} A K_k)^2}{\det(W_k^{\mathrm{T}} W_k) \det(K_k^{\mathrm{T}} K_k)} \frac{\det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(K_{k+1}^{\mathrm{T}} K_{k+1})} \right\}.
\end{aligned}$$

Now from theorem 1, we obtain

$$\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_{k-1}^{\mathrm{MR}}\|^2} = \begin{cases} 1 - \dfrac{\|r_k^{\mathrm{MR}}\|^2}{\|r_k^{\mathrm{OR}}\|^2} & \text{if } \det(K_k^{\mathrm{T}} A K_k) \neq 0, \\ 1 & \text{if } \det(K_k^{\mathrm{T}} A K_k) = 0 \end{cases}$$

which completes the proof. □

3.2. *The QR factorization of the Krylov matrix*

We consider now the QR factorization [8] of the Krylov matrix K_k. Let Q_k be an orthogonal matrix, i.e., $Q_k^{\mathrm{T}} Q_k = I_k$ and R_k be an upper triangular matrix of order k such that $K_k = Q_k R_k$.

From the fact that

$$K_{k+1} \begin{pmatrix} 0^{\mathrm{T}} \\ I_k \end{pmatrix} = A K_k$$

and by using the QR factorizations of the matrices K_k and K_{k+1} together, we get

$$\overline{H}_k \equiv Q_{k+1}^{\mathrm{T}} A Q_k = R_{k+1} \begin{pmatrix} 0^{\mathrm{T}} \\ I_k \end{pmatrix} R_k^{-1}. \tag{3.1}$$

Moreover we have

$$A Q_k = Q_{k+1} \overline{H}_k. \tag{3.2}$$

Since the matrices R_{k+1} and R_k^{-1} are upper triangular, $\overline{H}_k$ is a $(k+1) \times k$ upper Hessenberg matrix.

Now, the matrix H_k defined by

$$H_k \equiv Q_k^{\mathrm{T}} A Q_k, \tag{3.3}$$

is a square upper Hessenberg of order k. Moreover $\overline{H}_k$ has the form

$$\overline{H}_k = \begin{pmatrix} H_k \\ q_{k+1}^{\mathrm{T}} A Q_k \end{pmatrix}. \tag{3.4}$$

Hence by (3.1)

$$q_{k+1}^{\mathrm{T}} A Q_k = \frac{r_{k+1,k+1}}{r_{k,k}} e_k^{(k)\mathrm{T}}, \tag{3.5}$$

where $r_{i,j}$ denote the elements of the matrix R_{k+1}. We have

$$\frac{\det(K_k^{\mathrm{T}} K_k)}{\det(K_{k-1}^{\mathrm{T}} K_{k-1})} = \frac{\det(R_k^{\mathrm{T}} R_k)}{\det(R_{k-1}^{\mathrm{T}} R_{k-1})} = r_{k,k}^2.$$

If we set $h_{k+1,k} \equiv r_{k+1,k+1}/r_{k,k}$, and use the preceding relation, we obtain

$$h_{k+1,k}^2 = \frac{\det(K_{k+1}^{\mathrm{T}} K_{k+1}) \det(K_{k-1}^{\mathrm{T}} K_{k-1})}{\det(K_k^{\mathrm{T}} K_k)^2}. \tag{3.6}$$

Therefore from (3.4)–(3.6), the relation (3.2) can be written as

$$A Q_k = Q_k H_k + h_{k+1,k} q_{k+1} \big(e_k^{(k)}\big)^{\mathrm{T}}. \tag{3.7}$$

We shall now give some properties of the singular values of the matrices H_k and $\overline{H}_k$.

3.3. *Properties of the Ritz singular values*

Let $(\alpha_i^{(k)})$ and $(\beta_i^{(k)})$ for $i = 1, \dots, k$ be the singular values of the matrices H_k and $\overline{H}_k$, respectively. We order them by

$$\alpha_k^{(k)} \leqslant \alpha_{k-1}^{(k)} \leqslant \cdots \leqslant \alpha_1^{(k)} \quad \text{and} \quad \beta_k^{(k)} \leqslant \beta_{k-1}^{(k)} \leqslant \cdots \leqslant \beta_1^{(k)}.$$

These values $(\alpha_i^{(k)})$ and $(\beta_i^{(k)})$ will be called the Ritz singular values, and the Ritz adjacent singular values, respectively. The singular values of the matrix A will be denoted as $\sigma_1, \dots, \sigma_n$ and ordered as

$$\sigma_n \leqslant \sigma_{n-1} \leqslant \cdots \leqslant \sigma_1.$$

In the following theorem we gathered some interlacing properties relating the singular values of the three matrices H_k, $\overline{H}_k$ and A.

Theorem 4.

1. $\beta_i^{(k)} \in [\alpha_i^{(k)}, \alpha_{i-1}^{(k)}]$ for $i = 2, \dots, k$, and $\beta_1^{(k)} \geqslant \alpha_1^{(k)}$.
2. $\alpha_i^{(k)} \in [\beta_{i+1}^{(k)}, \beta_i^{(k)}]$ for $i = 1, \dots, k-1$, and $\beta_k^{(k)} \geqslant \alpha_k^{(k)}$.
3. $\alpha_i^{(k-1)} \in [\alpha_{i+1}^{(k)}, \alpha_i^{(k)}]$ for $i = 1, \dots, k-2$, and $\alpha_{k-1}^{(k)} \geqslant \alpha_{k-1}^{(k-1)}$.

4. $\beta_i^{(k-1)} \in [\beta_{i+1}^{(k)}, \beta_i^{(k)}]$ for $i = 1, \ldots, k-1$.
5. $\beta_i^{(k)} \in [\sigma_{n-k+i}, \sigma_i]$ for $i = 1, \ldots, k$.

Proof. 1.,2. From (3.2), (3.4) and (3.5) we deduce that

$$\overline{H}_k^{\mathrm{T}} \overline{H}_k = H_k^{\mathrm{T}} H_k + h_{k+1,k}^2 e_k^{(k)} e_k^{(k)\mathrm{T}}. \tag{3.8}$$

Hence the matrix $\overline{H}_k^{\mathrm{T}} \overline{H}_k$ is obtained by perturbing the matrix $H_k^{\mathrm{T}} H_k$ by a rank-one matrix. From [8, theorem 8.1.5, p. 412], we get the first two statements.

3. The matrix H_{k-1} is a $(k-1)$-square submatrix of H_k obtained by deleting the last row and the last column. Hence by using the interlacing theorem for singular values [20], we get the inequality given in 3.

4. The matrix $\overline{H}_k$ is obtained by adding one column to the matrix $\binom{\overline{H}_{k-1}}{0^{\mathrm{T}}}$, which has $\beta_1^{(k-1)}, \ldots, \beta_{k-1}^{(k-1)}$ as singular values. By using [20, theorem 1], we obtain the desired inequalities.

5. Finally, for the last part, we use the relation (3.2) which implies that

$$\overline{H}_k^{\mathrm{T}} \overline{H}_k = Q_k^{\mathrm{T}} A^{\mathrm{T}} A Q_k. \tag{3.9}$$

From this and by using [19, corollary 4.4, p. 198], we deduce that

$$\beta_i^{(k)} \in [\sigma_{n-k+i}, \sigma_i] \quad \text{for } i = 1, \ldots, k, \tag{3.10}$$

which completes the proof. □

Remarks.

1. From parts 2 and 5 of this theorem, we deduce that the smallest Ritz adjacent singular values $\beta_k^{(k)}$ not increase and are bounded below ($\sigma_n \leqslant \beta_k^{(k)} \leqslant \beta_{k-1}^{(k-1)}$). However the smallest Ritz singular values $\alpha_k^{(k)}$ are not bounded below. We only have the inequalities

$$0 \leqslant \alpha_k^{(k)} \leqslant \beta_k^{(k)} \quad \text{and} \quad 0 < \sigma_n \leqslant \beta_k^{(k)}. \tag{3.11}$$

2. The condition number of the matrix $\overline{H}_k$ is defined by $\kappa(\overline{H}_k) = \beta_1^{(k)}/\beta_k^{(k)}$. We have proved that $\beta_1^{(k)} \leqslant \sigma_1$ and $\beta_k^{(k)} \geqslant \sigma_n$. Hence $\kappa(\overline{H}_k) \leqslant \sigma_1/\sigma_n = \kappa(A)$.
3. The kth iterate of the OR method exists if and only if $\alpha_k^{(k)} \neq 0$.

As we said before the smallest Ritz singular values are not bounded below in general. We will give now two classes of matrices for which we have this fundamental property.

Theorem 5. If the matrix $A_s = (1/2)(A + A^{\mathrm{T}})$ is positive definite then

$$\alpha_k^{(k)} \geqslant \lambda_{\min}(A_s) > 0 \quad \forall k, \tag{3.12}$$

where $\lambda_{\min}(A_s)$ is the smallest eigenvalue of the matrix A_s.

Proof. From the Courant–Fischer Minimax Theorem [8], we have

$$0 < \lambda_{\min}(A_s) \leqslant (x, A_s x) = (x, Ax) \quad \forall x \in \mathbb{R}^n$$

such that $\|x\| = 1$.

It follows that $\forall y \in \mathbb{R}^k$ such that $\|y\| = 1$

$$0 < \lambda_{\min}(A_s) \leqslant (Q_k y, AQ_k y) \leqslant (y, H_k y) \leqslant \|H_k y\|.$$

From this we conclude that

$$0 < \lambda_{\min}(A_s) \leqslant \min_{\|y\|=1} \|H_k y\| = \alpha_k^{(k)},$$

which completes the proof. □

We consider now the skew-symmetric matrices ($A^{\mathrm{T}} = -A$). For such matrices we have the following result.

Theorem 6. If the matrix A is skew-symmetric, then

$$\alpha_{2k+1}^{(2k+1)} = 0 \quad \text{and} \quad \alpha_{2k}^{(2k)} \geqslant \sigma_n.$$

Proof. The matrix $H_k = Q_k^{\mathrm{T}} A Q_k$ is skew symmetric, which implies that $\det(H_k)$ vanishes when k is odd. Consequently $\alpha_k^{(k)} = 0$ if k is odd.

The matrix $H_{2k} = Q_{2k}^{\mathrm{T}} A Q_{2k}$ is a nonsingular skew symmetric matrix. All eigenvalues of H_{2k} are purely imaginary.

Moreover the eigenvalues of a real skew symmetric matrix come in complex conjugate pairs. So we can write these eigenvalues as $\pm \mathrm{i}\xi_j$ for $j = 1, \ldots, k$ with $\xi_j \in \mathbb{R}$ and $|\xi_k| \leqslant |\xi_{k-1}| \leqslant \cdots \leqslant |\xi_1|$. But the singular values of the normal matrix H_{2k} are $|\pm \mathrm{i}\xi_j| = |\xi_j|$.

If follows that $\alpha_{2k}^{(2k)} = \alpha_{2k-1}^{(2k)} = |\xi_k|$.

If we combine the first and the last part of theorem 3, we get

$$\alpha_{2k}^{(2k)} = \alpha_{2k-1}^{(2k)} = \beta_{2k}^{(2k)} \geqslant \sigma_n,$$

which proves the theorem. □

We are now in position to describe the rate of the convergence of these two methods in terms of the Ritz singular values.

4. The rate of convergence of the two methods

We will express the rate of convergence of the two methods in terms of singular values.

Theorem 7.

1. $$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} = \frac{\prod_{i=1}^{k-1}\beta_i^{(k-1)}}{\prod_{i=1}^{k}\beta_i^{(k)}}\sqrt{\sum_{i=1}^{k}\bigl(\beta_i^{(k)^2} - \alpha_i^{(k)^2}\bigr)} = \sqrt{1 - \prod_{i=1}^{k}\frac{\alpha_i^{(k)^2}}{\beta_i^{(k)^2}}}.$$

2. $$\frac{\|r_k^{\mathrm{OR}}\|}{\|r_{k-1}^{\mathrm{OR}}\|} = \frac{\prod_{i=1}^{k-1}\alpha_i^{(k-1)}}{\prod_{i=1}^{k}\alpha_i^{(k)}}\sqrt{\sum_{i=1}^{k}\bigl(\beta_i^{(k)^2} - \alpha_i^{(k)^2}\bigr)} \quad \text{if } \alpha_{k-1}^{(k-1)} \neq 0 \text{ and } \alpha_k^{(k)} \neq 0.$$

3. $$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_k^{\mathrm{OR}}\|} = \prod_{i=1}^{k}\frac{\alpha_i^{(k)}}{\beta_i^{(k)}} \qquad \text{if} \quad \alpha_k^{(k)} \neq 0.$$

Proof. 1. From theorem 1, we deduce that

$$\begin{aligned}\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_{k-1}^{\mathrm{MR}}\|^2} &= \frac{\det(K_{k+1}^{\mathrm{T}}K_{k+1})\det(W_{k-1}^{\mathrm{T}}W_{k-1})}{\det(W_k^{\mathrm{T}}W_k)\det(K_k^{\mathrm{T}}K_k)}\\ &= \frac{\det(K_k^{\mathrm{T}}K_k)}{\det(W_k^{\mathrm{T}}W_k)}\,\frac{\det(W_{k-1}^{\mathrm{T}}W_{k-1})}{\det(K_{k-1}^{\mathrm{T}}K_{k-1})}\,\frac{\det(K_{k+1}^{\mathrm{T}}K_{k+1})\det(K_{k-1}^{\mathrm{T}}K_{k-1})}{\det(K_k^{\mathrm{T}}K_k)^2}.\end{aligned}$$

Since $W_k^{\mathrm{T}}W_k = K_k^{\mathrm{T}}A^{\mathrm{T}}AK_k = R_k^{\mathrm{T}}\overline{H}_k^{\mathrm{T}}\overline{H}_kR_k$, we have

$$\frac{\det(W_k^{\mathrm{T}}W_k)}{\det(K_k^{\mathrm{T}}K_k)} = \det\bigl(\overline{H}_k^{\mathrm{T}}\overline{H}_k\bigr).$$

Consequently by using (3.6) we obtain

$$\frac{\|r_k^{\mathrm{MR}}\|^2}{\|r_{k-1}^{\mathrm{MR}}\|^2} = \frac{\det(\overline{H}_{k-1}^{\mathrm{T}}\overline{H}_{k-1})}{\det(\overline{H}_k^{\mathrm{T}}\overline{H}_k)}\,h_{k+1,k}^2. \tag{4.1}$$

On the other hand, from (3.8) we deduce that

$$\mathit{trace}\bigl(\overline{H}_k^{\mathrm{T}}\overline{H}_k\bigr) = \mathit{trace}\bigl(H_k^{\mathrm{T}}H_k\bigr) + h_{k+1,k}^2,$$

and by using the fact that $\mathit{trace}(\overline{H}_k^{\mathrm{T}}\overline{H}_k) = \sum_{i=1}^{k}\beta_i^{(k)^2}$, we get

$$h_{k+1,k}^2 = \sum_{i=1}^{k}\bigl(\beta_i^{(k)^2} - \alpha_i^{(k)^2}\bigr). \tag{4.2}$$

From this and by using the relation $\det(\overline{H}_k^{\mathrm{T}}\overline{H}_k) = \prod_{i=1}^{k}\beta_i^{(k)^2}$, we obtain the first equality.

For proving the second equality, we have just to express s_k from the singular values and to use theorem 2. We have

$$s_k^2 = 1 - c_k^2 = 1 - \frac{\det(K_k^{\mathrm{T}}AK_k)^2}{\det(W_k^{\mathrm{T}}W_k)\det(K_k^{\mathrm{T}}K_k)} = 1 - \frac{\det(H_k)^2}{\det(\overline{H}_k^{\mathrm{T}}\overline{H}_k)},$$

Consequently

$$s_k^2 = 1 - \prod_{i=1}^{k} \frac{\alpha_i^{(k)2}}{\beta_i^{(k)2}}$$

since $\det(H_k{}^2) = \prod_{i=1}^{k} \alpha_i^{(k)2}$.

2. The proof of the last equality is similar to the one given for the first part. In fact, from theorem 1 we obtain

$$\begin{aligned}\frac{\|r_k^{\mathrm{OR}}\|^2}{\|r_{k-1}^{\mathrm{OR}}\|^2} &= \frac{\det(K_{k-1}^{\mathrm{T}} A K_{k-1})^2}{\det(K_{k-1}^{\mathrm{T}} K_{k-1}) \det(K_k^{\mathrm{T}} K_k)} \frac{\det(K_k^{\mathrm{T}} K_k) \det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(K_k^{\mathrm{T}} A K_k)^2} \\ &= \frac{\det(K_{k-1}^{\mathrm{T}} A K_{k-1})^2}{\det(K_{k-1}^{\mathrm{T}} K_{k-1})^2} \frac{\det(K_k^{\mathrm{T}} K_k)^2}{\det(K_k^{\mathrm{T}} A K_k)^2} \frac{\det(K_{k-1}^{\mathrm{T}} K_{k-1}) \det(K_{k+1}^{\mathrm{T}} K_{k+1})}{\det(K_k^{\mathrm{T}} K_k)^2}.\end{aligned}$$

But $\det(K_k^{\mathrm{T}} A K_k)^2 / \det(K_k^{\mathrm{T}} K_k) = \det(H_k^{\mathrm{T}} H_k)$, hence

$$\frac{\|r_k^{\mathrm{OR}}\|^2}{\|r_{k-1}^{\mathrm{OR}}\|^2} = \frac{\det(H_{k-1}^{\mathrm{T}} H_{k-1})}{\det(H_k^{\mathrm{T}} H_k)} h_{k+1,k}^2. \tag{4.3}$$

The second equality now follows by using (4.2).

3. The last equality is deduced from theorem 2 and the second equality of 1. □

Remarks.

1. In this theorem we have shown that $c_k = \prod_{i=1}^{k} (\alpha_i^{(k)} / \beta_i^{(k)})$, and from theorem 2 we get

$$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} = \sqrt{1 - c_k^2} \quad \text{and} \quad \frac{\|r_k^{\mathrm{OR}}\|}{\|r_{k-1}^{\mathrm{OR}}\|} = \frac{c_{k-1}}{c_k} \sqrt{1 - c_k^2}. \tag{4.4}$$

   The first equality of (4.4) is well known and was given by Saad and Schultz in [18]. The second equality can be derived from formulas (5.4) and (4.4) of Brown [1].

2. If $c_k \geqslant \sqrt{2}/2$, then $\|r_k^{\mathrm{OR}}\| \leqslant \|r_{k-1}^{\mathrm{OR}}\|$, hence the residual norms decreases.

3. Since $\alpha_i^{(k)} / \beta_i^{(k)}$ converges to 1 when k tends to n, c_k tends to 1. In this case both methods (MR and OR) will have the same performance. If $\alpha_k^{(k)}$ is close to the origin, then c_k can be very small, and the residual norms for the OR method can grow without bound, which corresponds to the stagnation of the MR method, see also [1].

Example 3 (Perfectly conditioned matrices). Let us apply theorem 6, to the class of matrices perfectly conditioned. In this case all singulars value are equals, hence $\sigma_1 = \cdots = \sigma_n = \sigma$. Consequently the matrix A is a multiple of an orthogonal matrix. Therefore from theorem 3, we have $\beta_i^{(k)} = \sigma$ for $i = 1, \ldots, k$ and $\alpha_i^{(k)} = \sigma$ for $i = 1, \ldots, k-1$. Hence of the convergence of the two methods depends only of $\alpha_k^{(k)}$ and theorem 6 becomes:

If the matrix A is perfectly conditioned ($\sigma_i = \sigma$ for $i = 1, \ldots, n$), then

1. $$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} = \sqrt{1 - \frac{{\alpha_k^{(k)}}^2}{\sigma^2}}.$$
2. $$\frac{\|r_k^{\mathrm{OR}}\|}{\|r_{k-1}^{\mathrm{OR}}\|} = \frac{\alpha_{(k-1)}^{(k-1)}}{\alpha_k^{(k)}}\sqrt{1 - \frac{{\alpha_k^{(k)}}^2}{\sigma^2}} \quad \text{if } \alpha_{k-1}^{(k-1)} \neq 0 \text{and } \alpha_k^{(k)} \neq 0.$$
3. $$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_k^{\mathrm{OR}}\|} = \frac{\alpha_k^{(k)}}{\sigma} \quad \text{if } \alpha_k^{(k)} \neq 0.$$

We consider the system (2.1) with the $n \times n$ perfectly conditioned matrix

$$A_n = \begin{pmatrix} 0 & \ldots & \ldots & 0 & 1 \\ 1 & \ddots & & \vdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}$$

and $b = A(1, 0, \ldots, 0)^{\mathrm{T}}$. This example was also given by Brown [1].

For this example, $\kappa(A_n)$=1, hence from theorem 3, we deduce that

$$\beta_i^{(j)} = \alpha_i^{(j)} = 1 \quad \text{for } 1 \leqslant i < j \leqslant n, \quad \text{and} \quad \beta_j^{(j)} = 1 \quad \text{for } j = 1, \ldots, n.$$

Now, if $x_0 = 0$ then $\alpha_j^{(j)} = 0$, for $j = 1, \ldots, n-1$ and $\alpha_n^{(n)} = 1$. Consequently $\|r_j^{\mathrm{MR}}\| = 1$ and the jth iterates do not exist for $j = 1, \ldots, n-1$. Hence we obtain the solution at the nth iteration and the smallest Ritz singular value vanishes until the nth iteration.

For the class of perfectly conditioned matrix, the convergence of the two methods depends only of the smallest Ritz singular value. Now, from (4.1) we deduce that

$$\left(\alpha_k^{(k)}\right)^2 = \sigma^2\left(1 - h_{k+1,k}^2\right).$$

Since we have the solution of the system (2.1) if and only if $h_{k+1,k} = 0$, then $\alpha_k^{(k)}$ tends to σ.

If the matrix A is normal, it was proved in [10] that the matrix H_m is also normal, if m is the degree of the minimal polynomial of A for r_0. Hence for normal matrices $(\alpha_k^{(k)})$ and $(\beta_k^{(k)})$ converge to a singular value of the matrix A. Now, we shall give an example for which the smallest Ritz singular value does not converges to any singular value of the matrix A.

Example 4. This example is taken from [9]. We consider the $n \times n$ matrix $A = SBS^{-1}$, where

$$S = \begin{pmatrix} 1 & \beta & & & \\ & 1 & \beta & 0 & \\ & & \ddots & \ddots & \\ & 0 & & \ddots & \beta \\ & & & & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & & & & \\ & 2 & & 0 & \\ & & 3 & & \\ & 0 & & \ddots & \\ & & & & n \end{pmatrix}.$$

With the variable β, we can control the condition number of the similarity transformation S and make A more non-normal; see [9].

We used $b = A(1, 0, \ldots, 0, 1)^{\mathrm{T}}$, $x_0 = 0$, $\beta = 0.9$ and $n = 10$. For this matrix we have $\sigma_{10} = 4.6141\ 10^{-1}$ and $\sigma_9 = 1$, but $\alpha_9^{(9)} = \beta_9^{(9)} = 8.4797\ 10^{-1}$ and the two methods converge in 9 iterations.

From this example, we deduce that the smallest singular Ritz values and the smallest adjacent singular Ritz value do not converge to any singular value. But if the degree of the minimal polynomial of A for r_0 is n, then $\alpha_k^{(k)}$ and $\beta_k^{(k)}$ tend to σ_n.

These examples show that the convergence behavior of the smallest singular Ritz value can be used to explain the convergence of the MR and the OR method. Moreover we have pointed out the importance of the convergence of the smallest singular Ritz value. We give now a bound for the rate of convergence of theorem 6.

Theorem 8.

1. $$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} \leqslant \sqrt{1 - \frac{{\alpha_k^{(k)}}^2}{{\beta_1^{(k)}}^2}} \leqslant \sqrt{1 - \frac{{\alpha_k^{(k)}}^2}{{\sigma_1}^2}}.$$

2. $$\frac{\|r_k^{\mathrm{OR}}\|}{\|r_{k-1}^{\mathrm{OR}}\|} \leqslant \frac{\sqrt{\sum_{i=1}^k ({\beta_i^{(k)}}^2 - {\alpha_i^{(k)}}^2)}}{\alpha_k^{(k)}} \quad \text{if } \alpha_{k-1}^{(k-1)} \neq 0 \text{ and } \alpha_k^{(k)} \neq 0.$$

3. $$\frac{\|r_k^{\mathrm{OR}}\|}{\|r_k^{\mathrm{MR}}\|} \leqslant \frac{\beta_1^{(k)}}{\alpha_k^{(k)}} \leqslant \frac{\sigma_1}{\alpha_k^{(k)}} \quad \text{if } \alpha_k^{(k)} \neq 0.$$

4. $$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} \leqslant \frac{\sqrt{\sum_{i=1}^k ({\beta_i^{(k)}}^2 - {\alpha_i^{(k)}}^2)}}{\beta_k^{(k)}}.$$

Theses inequalities are deduced from theorems 6 and 3.

From parts 1 and 3 of theorem 9, we see that if a smallest singular value is bounded below, i.e., $\alpha_k^{(k)} \geqslant \eta$, then we have

$$\frac{\|r_k^{\mathrm{MR}}\|}{\|r_{k-1}^{\mathrm{MR}}\|} \leqslant \sqrt{1 - \frac{\eta^2}{{\sigma_1}^2}} \quad \text{and} \quad \frac{\|r_k^{\mathrm{OR}}\|}{\|r_k^{\mathrm{MR}}\|} \leqslant \frac{\sigma_1}{\eta}. \tag{4.5}$$

From this first inequality we can deduce that the restarted GMRES(k) that is GMRES restarted every k steps, is guaranteed to converge. Hence if the symmetric part of the matrix is positive definite, then from theorem 4, we have $\eta = \lambda_{\min}(A_s)$, and we get the result given in [3,4]. Similarly if the matrix is skew-symmetric, we deduce from theorem 5 that $\alpha_{2k}^{(2k)} \geqslant \sigma_n$, consequently GMRES($k$) converges, if $k \geqslant 2$, we obtain a result given in [11].

5. Symmetric case

This section will be devoted to symmetric matrices. The convergence of MinRes, which is the MR method when the matrix is symmetric, and CG (conjugate gradient), which is the OR method in the symmetric case, have been analyzed in several papers [16,21]. Recently Paige, Parlett and Van der Vorst gave new insight about the convergence of MinRes and CG. We will give now another explanations of the behavior of these two methods. In this section, we assume that the matrix A is symmetric.

We have the following result.

Theorem 9. If $rank(K_k) = k \leqslant m$, where m is the degree of the minimal polynomial of the matrix A for the vector r_0, then

1. $\alpha_{k-1}^{(k-1)} \neq 0$ or $\alpha_k^{(k)} \neq 0$.
2. $\alpha_m^{(m)} \neq 0$.
3. If the matrix A is positive definite, then $\alpha_k^{(k)} \geqslant \lambda_{\min}(A) > 0$.

Proof. 1. By applying Sylvester's identity to the determinant $\det(K_k^{\mathrm{T}} A K_k)$, we get

$$\det\bigl(K_k^{\mathrm{T}} A K_k\bigr) X_k = \det\bigl(K_{k-1}^{\mathrm{T}} A K_{k-1}\bigr) X_{k+1} - \det\bigl(W_{k-1}^{\mathrm{T}} W_{k-1}\bigr)^2, \tag{5.1}$$

where $X_k = \det(W_{k-2}^{\mathrm{T}} A W_{k-2})$.

Hence if we assume that $\alpha_{k-1}^{(k-1)} = 0$ and $\alpha_k^{(k)} = 0$, we have $\det(K_{k-1}^{\mathrm{T}} A K_{k-1}) = \det(K_k^{\mathrm{T}} A K_k) = 0$, and consequently from (5.1) we deduce that $\det(W_{k-1}^{\mathrm{T}} W_{k-1}) = 0$, which is not possible since $k \leqslant m$. Hence $\alpha_{k-1}^{(k-1)} \neq 0$ or $\alpha_k^{(k)} \neq 0$.

2. Since m is the degree of the minimal polynomial of A for r_0, we have $rank(K_{m+1}) = rank(K_m) = m$, which implies that $\det(K_{m+1}^{\mathrm{T}} A K_{m+1}) = 0$.

On the other hand, if we write (5.1) with $k = m + 1$, we get

$$\det\bigl(K_m^{\mathrm{T}} A K_m\bigr) X_{m+1} = \det\bigl(W_m^{\mathrm{T}} W_m\bigr)^2 \neq 0.$$

Hence $\det(K_m^{\mathrm{T}} A K_m) \neq 0$ and $\alpha_m^{(m)} \neq 0$.

3. If A is symmetric, then $A_s = A$ and the result follows by using theorem 5.

Since now, the matrix H_k is symmetric, we deduce that this matrix is tridiagonal, so we will give the ratio of the norm of the cg method only in terms of Ritz eigenvalues, which will be denoted by $\theta_1^{(k)}, \dots, \theta_k^{(k)}$ and ordered as $\theta_k^{(k)} \leqslant \theta_{k-1}^{(k)} \leqslant \cdots \leqslant \theta_1^{(k)}$. Let us

Table 1

k	1	$\cdots$	14	15	16	17	$\cdots$	24	25	26	27	$\cdots$	69
$\alpha_k^{(k)}$	165	$\cdots$	4.6	1.5	**0.71**	2.04	$\cdots$	3.5	2.51	**0.58**	1.02	$\cdots$	3.
$\beta_k^{(k)}$	168	$\cdots$	9.7	7.8	6.22	5.01	$\cdots$	3.52	3.52	3.51	3.48	$\cdots$	3.
$\|\|r_k^{cg}\|\|$	1.5e+4	$\cdots$	5.5e+1	1.4e+2	**2.3e+2**	6.2e+1	$\cdots$	1.5	2.4e+1	**9.3e+1**	4.5e+1	$\cdots$	6.5e−8
$\|\|r_k^{cr}\|\|$	1.5e+4	$\cdots$	1.76	1.75e+1	1.75e+1	1.68e+1	$\cdots$	5.62	5.62	5.61	5.56	$\cdots$	5.8e−8

remark that $\alpha_k^{(k)} = \min |\theta_i^{(k)}|$ and if the matrix A is symmetric positive definite we have $\alpha_i^{(k)} = \theta_i^{(k)}$ for $i = 1, \ldots, k$. $\square$

Let P_k be the characteristic polynomial of H_k, we have

$$P_{k+1}(t) = (h_{k,k} - t)P_k(t) - h_{k,k+1}^2 \; P_{k-1}(t). \tag{5.2}$$

Now, $P_k(t) = \prod_{i=1}^k (\theta_i^{(k)} - t)$, thus we have for $j \in \{1, \ldots, k\}$

$$h_{k,k+1}^2 = -\frac{P_{k+1}(\theta_j^{(k)})}{P_{k-1}(\theta_j^{(k)})} = -\frac{\prod_{i=1}^{k+1} (\theta_i^{(k+1)} - \theta_j^{(k)})}{\prod_{i=1}^{k-1} (\theta_i^{(k-1)} - \theta_j^{(k)})}. \tag{5.3}$$

By using (4.3), we get the following result.

Theorem 10. If $\alpha_i^{(i)} \neq 0$ for $i \in \{k-1, k\}$, then

$$\frac{||r_k^{cg}||^2}{||r_{k-1}^{cg}||^2} = -\frac{\prod_{i=1}^{k-1} (\theta_i^{(k-1)})^2}{\prod_{i=1}^{k} (\theta_i^{(k)})^2} \frac{\prod_{i=1}^{k+1} (\theta_i^{(k+1)} - \theta_j^{(k)})}{\prod_{i=1}^{k-1} (\theta_i^{(k-1)} - \theta_j^{(k)})} \quad \text{for } j \in \{1, \ldots, k\}.$$

In the following example, we show how to quantify peaks in the convergence behavior of CG method.

Example 5. Let us consider the matrix $A = diag(-4, -3, 4, 6, 8, \ldots, 198)$ of dimension 100. The right hand side has been chosen so that the solution of (2.1) is $x = (1, 2, \ldots, 100)^{\mathrm{T}}$, and $x_0 = 0$. We obtain the results shown in table 1.

We have two peaks, the first one at the step 16 and the second at the step 26. As we can see, we have peaks even when the Ritz smallest value are not close to zero. Peaks occurs when $\alpha_k^{(k)}$ is a local minimum. We also note a relationships between peaks and plateaus.

Acknowledgements

I wish to thank Gene Golub for his help in the proof of the last part of theorem 3. I would also like to thank Martin Gutknecht, Mickael Eiermann and Henk Van Der Vorst for their helpful discussions on various topics in this paper.

References

[1] P.N. Brown, A theoretical comparison of the Arnoldi and the GMRES algorithms, SIAM J. Sci. Statist. Comput. 12 (1991) 58–78.
[2] J. Cullum and A. Greenbaum, Relations between Galerkin and norm-minimizing iterative methods for solving linear systems, SIAM J. Matrix Anal. Appl. 17 (1996) 223–248.
[3] S.C. Eisenstat, H.C. Elman and M.H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, SIAM J. Numer. Anal. 20 (1983) 345–357.
[4] H.C. Elman, Iterative methods for large sparse nonsymmetric systems of linear equations, Ph.D. thesis, Computer Science Department, Yale University, New Haven, CT (1982).
[5] F.R. Gantmacher, *The Theory of Matrices*, Vol. 1 (Chelsea, New York, 1959).
[6] A. Greenbaum and L.N. Trefethen, GMRES/CR and ARNOLDI/LANCZOS as matrix approximation problems, SIAM J. Sci. Statist. Comput. 15 (1994) 357–368.
[7] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards 49 (1952) 409–436.
[8] G. Golub and C.F. van Loan, *Matrix Computations*, 2nd ed. (Johns Hopkins Univ. Press, Baltimore, MD, 1989).
[9] Y. Huang and H.A. van der Vorst, Some observations on the convergence behaviour of GMRES, Delft University of Technology, Report 89–09 (1989).
[10] T. Huckle, The Arnoldi method for normal matrices, SIAM J. Matrix Anal. Appl. 15 (1994) 479–489.
[11] K. Jbilou and H. Sadok, Analysis of some vector extrapolation methods for solving systems of linear equations, Numer. Math. 70 (1995) 73–89.
[12] N.M. Nachtigal, S.C. Reddy and L.N. Trefethen, How fast are nonsymmetric matrix iterations, SIAM J. Matrix Anal. Appl. 13 (1992) 778–795.
[13] C.C. Paige, B.N. Parlett and H.A. van der Vorst, Approximate solutions and eigenvalue bounds from Krylov subspaces, Numer. Linear Algebra Appl. 2 (1995) 115–134.
[14] C.C. Paige and M.A. Saunders, Solution of sparse indefinite systems of linear equations, SIAM J. Numer. Anal. 12 (1975) 617–629.
[15] Y. Saad, Variations on Arnoldi's method for computing eigenelements of large unsymmetric, Linear Algebra Appl. 34 (1980) 269–295.
[16] Y. Saad, Krylov subspace methods for solving large unsymmetric linear systems, Math. Comp. 37 (1981) 105–126.
[17] Y. Saad and M.H. Schultz, Conjugate gradient-like algorithms for solving nonsymmetric linear systems, Math. Comp. 44 (1985) 417–424.
[18] Y. Saad and M.H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Statist. Comput. 7 (1986) 856–869.
[19] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory* (Academic Press, New York, 1990).
[20] R.C. Thompson, Principal submatrices IX: interlacing inequalities for singular values of submatrices, Linear Algebra Appl. 5 (1972) 1–12.
[21] A. van der Sluis and H.A. van der Vorst, The rate of convergence of conjugate gradients, Numer. Math. 48 (1986) 543–560.
[22] H.A. van der Vorst and C. Vuik, The rate of convergence of the GMRES method, J. Comput. Appl. Math. 48 (1993) 327–341.