Numerical Linear Algebra with Applications, Vol. 1(5), 427–447 (1994)

On the Convergence Behavior of the Restarted GMRES Algorithm for Solving Nonsymmetric Linear Systems

Wayne Joubert

Computer Research and Applications, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

The solution of nonsymmetric systems of linear equations continues to be a difficult problem. A main algorithm for solving nonsymmetric problems is restarted GMRES. The algorithm is based on restarting full GMRES every s iterations, for some integer $s > 0$. This paper considers the impact of the restart frequency s on the convergence and work requirements of the method. It is shown that a good choice of this parameter can lead to reduced solution time, while an improper choice may hinder or preclude convergence. An adaptive procedure is also presented for determining automatically when to restart. The results of numerical experiments are presented.

KEY WORDS Nonsymmetric linear systems Conjugate gradient type iterative methods GMRES Restarting

1. Introduction

We consider linear systems of the form

$$Au = b \tag{1.1}$$

where $A \in \mathbb{C}^{N \times N}$ is nonsingular and possibly non-Hermitian.

A major class of methods for solving (1.1) is the class of *polynomial methods*, defined by

$$u^{(n)} = u^{(0)} + Q_{n-1}(A)r^{(0)}, \quad \deg Q_{n-1} \le n - 1 \tag{1.2}$$

where $u^{(n)}$, $n \ge 0$, is the nth iterate which approximates $u = A^{-1}b$, and $r^{(n)} = b - Au^{(n)}$ is the associated residual. Alternatively,

$$r^{(n)} = P_n(A)r^{(0)}, \quad \deg P_n \le n, \quad P_n(0) = 1 \tag{1.3}$$

CCC 1070–5325/94/050427–21

Received February 1992
Revised 17 May 1994

with $P_n(z) = 1 - zQ_{n-1}(z)$, or

$$r^{(n)} - r^{(0)} \in AK_n(r^{(0)}, A) \tag{1.4}$$

where $K_n(v, A) = \text{span}\{A^i v\}_{i=0}^{n-1}$ is the associated Krylov space. Examples of such methods include the conjugate gradient [13], biconjugate gradient [20] and GMRES [26] methods.

An orthogonality condition may be used to specify P_n in (1.2)–(1.4). For non- Hermitian problems, two approaches are often used. First, the condition

$$r^{(n)} \perp K_n(\tilde{r}^{(0)}, A^*) \tag{1.5}$$

may be imposed, for some vector $\tilde{r}^{(0)}$, typically $\tilde{r}^{(0)} = r^{(0)}$, and A^* the conjugate transpose of A, giving rise to the Lanczos or biconjugate gradient method. Specifically, (1.3,1.5) yield

$$r^{(n)} = r^{(0)} - A\mathbf{K}_n(r^{(0)}, A)[\mathbf{K}_n(\tilde{r}^{(0)}, A^*)^* A\mathbf{K}_n(r^{(0)}, A)]^{-1}\mathbf{K}_n(\tilde{r}^{(0)}, A^*)^* r^{(0)} \tag{1.6}$$

as long as the matrix $\mathbf{K}_n(\tilde{r}^{(0)}, A^*)^* A\mathbf{K}_n(r^{(0)}, A)$ is nonsingular. Here the standard Krylov basis matrix is defined by $\mathbf{K}_n(v, A) = [v, Av, \ldots, A^{n-1}v]$ and the degree of a vector $d(v, A) = \min\{\deg P : P \text{ monic}, P(A)v = 0\}$, and $n \leq d(r^{(0)}, A)$ is assumed, so that $\mathbf{K}_n(r^{(0)}, A)$ has full rank.

A second method is obtained by imposing

$$r^{(n)} \perp AK_n(r^{(0)}, A^*) \tag{1.7}$$

or equivalently

$$P_n \text{ chosen such that } ||r^{(n)}|| \text{ is minimized} \tag{1.8}$$

where $||\cdot||$ denotes the standard Euclidean norm. This defines the minimal residual method, which is implemented for example by the GMRES algorithm. The resulting residuals are

$$r^{(n)} = r^{(0)} - A\mathbf{K}_n(r^{(0)}, A)[\mathbf{K}_n(r^{(0)}, A)^* A^* A\mathbf{K}_n(r^{(0)}, A)]^{-1}\mathbf{K}_n(r^{(0)}, A)^* r^{(0)} \tag{1.9}$$

assuming $n \leq d(r^{(0)}, A)$, so $\mathbf{K}_n(r^{(0)}, A)^* A^* A\mathbf{K}_n(r^{(0)}, A)$ is nonsingular. Note $n \geq d(r^{(0)}, A)$ implies $r^{(n)} = 0$, so the method always converges; for a full discussion see [18].

The Lanczos approach has several advantages. Since $\mathbf{K}_n(\tilde{r}^{(0)}, A^*)^* A\mathbf{K}_n(r^{(0)}, A)$ is a Hankel matrix with only $2n - 1$ distinct entries, two-term recurrences may be used to implement (1.6). Also, practical experience suggests that the number of iterations required to converge to the stopping criterion $||r^{(n)}||/||b|| \leq \zeta$ by the biconjugate gradient method is not much more than that required by the minimal residual method (1.3,1.7), which is optimal [14]. On the other hand, the matrix $\mathbf{K}_n(\tilde{r}^{(0)}, A^*)^* A\mathbf{K}_n(r^{(0)}, A)$ may be singular or nearly so, leading to $r^{(n)}$ very large or undefined in situations which are difficult to predict beforehand. For recent work on improvements to the biconjugate gradient method, see for example [15], [12], [9], and [28].

In contrast, the minimal residual method guarantees $||r^{(n)}||$ is nonincreasing in n and is guaranteed to converge. However, for general A, $\mathbf{K}_n(r^{(0)}, A)^* A^* A\mathbf{K}_n(r^{(0)}, A)$ contains $n(n + 1)/2$ distinct entries, so algorithms to implement (1.9) must use long recurrences ([5], [6], [19], [14, sec. 2.5]). Nonetheless, methods of this type are the methods of choice in various situations, such as when A is the result of a very good preconditioner so that only a few iterations are required, or when reliable solution of (1.1) must be guaranteed and certain properties of A are known which can assure convergence of the method.

Owing to the long recurrences, it is common to weaken the condition (1.7)–(1.8). This can be done by restarting: for some $s \geq 1$, s steps of the minimal residual method are applied, and the resulting iterate is used as the start of another s step cycle, and so forth. The choice of such $s < \infty$ means that the method in some cases may not converge. The purpose of the current study is to examine the impact of s on the convergence behavior and computational work requirements of the method.

Section 2 gives a brief review of some common restarted algorithms and their properties. Then in section 3 a general convergence theory for the minimal residual method is set forth. This is used in section 4 to derive some results on the optimal choice of the restart frequency. In section 5 an adaptive algorithm is defined which determines automatically when to restart, and numerical results with this algorithm are given in section 6.

2. Definition of restarted algorithms

A restarted method is the repeated application of the minimal residual method (1.3,1.7), typically for a fixed number s of iterations per cycle. The minimal residual method itself may be implemented by any one of several algorithms; for a comprehensive survey, see [17].

For this study, two standard algorithms are considered: the Orthomin or GCR algorithm ([29], [4]), and the GMRES algorithm. The formulas for these algorithms are given below. In section 5 the appropriate modifications of these algorithms are given to implement adaptive restarting.

Orthomin is defined by

$$q^{(n)} = r^{(n)} + \sum_{i=0}^{n-1} \alpha_{n,i} q^{(i)}, \qquad \alpha_{n,i} = -\frac{(Aq^{(i)}, Ar^{(n)})}{(Aq^{(i)}, Aq^{(i)})}$$

$$u^{(n+1)} = u^{(n)} + \lambda_n q^{(n)}, \qquad r^{(n+1)} = r^{(n)} - \lambda_n A q^{(n)}, \qquad \lambda_n = \frac{(Ar^{(n)}, r^{(n)})}{(Aq^{(n)}, Aq^{(n)})}$$

where $(u, v) = u^* v$ defines the standard ℓ^2 inner product. The algorithm is guaranteed not to break down if and only if A is definite, in the sense that $v^* A v \neq 0$ for all $v \in \mathbb{C} \setminus \{0\}$. Note that a real matrix A is definite if and only if its Hermitian part $A_H = (A + A^*/2)$ or its negative $-A_H$ is Hermitian and positive definite (HPD), and a complex matrix A is definite if and only if $(e^{i\theta} A)_H$ is HPD for some θ [18].

The GMRES algorithm is defined as follows. For $Q_i = [q^{(0)} \ldots q^{(i-1)}]$ let

$$h_{1,0} q^{(0)} = r^{(0)}, \qquad h_{i+1,i} q^{(i)} = \hat{q}^{(i)} \equiv A q^{(i-1)} - \sum_{j=0}^{i-1} h_{j+1,i} q^{(j)}, \quad i \geq 1$$

$$u^{(n)} = u^{(0)} + Q_n y_n, \qquad r^{(n)} = r^{(0)} - A Q_n y_n = r^{(0)} - Q_{n+1} H_{n+1} y_n$$

where $h_{j+1,i} = (q^{(j)}, Aq^{(i-1)})$, $j < i$, and $h_{i+1,i} = \|\hat{q}^{(i)}\|$. Thus $AQ_n = Q_{n+1} H_{n+1}$, where $H_{n+1} = \{h_{i,j}\}$ is upper Hessenberg.

Table 1. Average work per iteration, restart frequency s (excluding matrix–vector products)

Method	Dot products: u^*v	SAXPYs: $y \leftarrow y + \alpha x$	Total
Orthomin	$\frac{(s+5)}{2}$	$s+1$	$\frac{3}{2}s + \frac{7}{2}$
GMRES	$\frac{(s+1)(s+2)}{(2s)}$	$\frac{(s^2+5s+2)}{(2s)}$	$s + 4 + \frac{2}{s}$

The matrix H_{n+1} is factored via a QR factorization using Givens rotations:

$$H_{n+1} = P_{n+1} \begin{bmatrix} I_n \\ 0 \end{bmatrix} U_n, \quad P_{n+1} = \prod_{i=1}^{n} R_{n+1,i}, \quad R_{n+1,i} = \begin{bmatrix} I_{i-1} & & \\ & \begin{bmatrix} c_i & -s_i^* \\ s_i & c_i^* \end{bmatrix} & \\ & & I_{n-i} \end{bmatrix}$$

where I_n is the identity and U_n is upper triangular. Then y_n may be obtained by solving $U_n y_n = [\, I_n \quad 0 \,] z_{n+1} \equiv \check{z}_n$, where

$$z_{n+1} \equiv P_{n+1}^* h_{1,0} e_1 = \begin{bmatrix} z_{n-1} \\ \hat{\zeta}_n \\ \tilde{\zeta}_n \end{bmatrix} = R_{n+1,n}^* \begin{bmatrix} z_{n-2} \\ \hat{\zeta}_{n-1} \\ \tilde{\zeta}_{n-1} \\ 0 \end{bmatrix} = \begin{bmatrix} z_{n-2} \\ \hat{\zeta}_{n-1} \\ c_n^* \tilde{\zeta}_{n-1} \\ -s_n \tilde{\zeta}_{n-1} \end{bmatrix}$$

from which the recurrences $\hat{\zeta}_n = c_n^* \tilde{\zeta}_{n-1}$, $\tilde{\zeta}_n = -s_n \tilde{\zeta}_{n-1}$ and $\tilde{\zeta}_0 = h_{1,0}$ are obtained. Furthermore,

$$H_{n+1} y_n = P_{n+1} \begin{bmatrix} U_n \\ 0 \end{bmatrix} U_n^{-1} \check{z}_n = P_{n+1} \begin{bmatrix} \check{z}_n \\ 0 \end{bmatrix}$$

The residual norm is calculated by $||r^{(n)}||^2 = ||r^{(n-1)}||^2 - |\hat{\zeta}_n|^2 = |\tilde{\zeta}_n|^2$.

For general A, the cost to perform n steps of a minimal residual algorithm is proportional to n^2. Table 1 summarizes the average work per iteration for the restarted versions of Orthomin and GMRES, where s denotes the restart frequency. For Orthomin, the cost of a residual norm computation $||r^{(n)}||$ for the stopping test at each iteration is included; for GMRES, this quantity is available without extra vector work. It is assumed that vector scalings are not performed explicitly but accumulated within a table when necessary. For both algorithms, one matrix–vector product is required per step.

The work per iteration for restarted algorithms may be approximated by $a_w s + b_w$, where $a_w, b_w > 0$ are constants. The values of a_w and b_w depend on the costs of the dot product, SAXPY and matrix–vector product operations on the given computer.

The theory and adaptive techniques of this study directly apply to other restarted algorithms as well, such as the GMRES / Chebyshev basis algorithm defined in [17] which is up to twice as fast as GMRES and gives the same iterates.

3. Basic convergence theory

The two fundamental quantities to be used to study the convergence behavior of the minimal residual method are defined as follows. First, for $\mathbb{K}$ denoting the reals $\mathbb{R}$ or the complex numbers $\mathbb{C}$, and $A \in \mathbb{K}^{N\times N}$ and $n \geq 0$, let

$$\psi_{n,\mathbb{K}}(A) = \sup_{\substack{r\in\mathbb{K}^N\\ r\neq 0}} \inf_{\substack{\deg P_n\leq n\\ P_n(0)=1}} \frac{||P_n(A)r||}{||r||}, \qquad \varphi_{n,\mathbb{K}}(A) = \inf_{\substack{\deg P_n\leq n\\ P_n(0)=1}} \sup_{\substack{r\in\mathbb{K}^N\\ r\neq 0}} \frac{||P_n(A)r||}{||r||}$$

where P_n is a polynomial over $\mathbb{C}$. The quantity $\varphi_{n,\mathbb{K}}(A)$ measures the effectiveness of the best fixed polynomial P_n which must be applied to all initial residuals $r = r^{(0)}$ (and implicitly the effectiveness of the polynomial preconditioner Q_{n-1} which in some sense is optimal, where $P_n(z) = 1 - zQ_{n-1}(z)$), whereas $\psi_{n,\mathbb{K}}(A)$ allows a different polynomial to be applied to each residual $r^{(0)}$. In either case, when $\mathbb{K}$ is suppressed in the notation, it is assumed to be $\mathbb{C}$.

This section focuses on $\psi_{n,\mathbb{K}}(A)$, which embodies the convergence information for the minimal residual method applied to A. Unfortunately, good estimates of $\psi_{n,\mathbb{K}}(A)$ are known only for certain cases. This section gives some results which are known.

We define the degree of a matrix as $d(A) = \min\{\deg P : P \text{ monic}, P(A) = 0\}$. Note for any v, $0 \leq d(v, A) \leq d(A) \leq N$.

It may be shown easily using (1.6) that $\psi_{n,\mathbb{K}}(A)^2 = \sup_{r\in\mathbb{K}^N} F_n(r)$ holds for $A \in \mathbb{K}^{N\times N}$, where

$$F_n(r) = \begin{cases} 0, & d(r, A) < n \\ 1 - r^*A\mathbf{K}_n(r, A)[\mathbf{K}_n(r, A)^*A^*A\mathbf{K}_n(r, A)]^{-1}\mathbf{K}_n(r, A)^*A^*r/r^*r, & \text{else} \end{cases}$$

Clearly, F_n considered as a map on $\mathbb{R}^{2N}$ has any number of continuous derivatives on the open set which is the complement of $\mathcal{S} = \{r \in \mathbb{C} : d(r, A) < n\}$. Furthermore,

Lemma 3.1. *$F_n : \mathbb{C}^N \to \mathbb{C}$ is continuous.*

Proof If $n \geq d(A)$, then $F_n(v) = 0$ for all v. Otherwise, F_n, being a rational function of the components of vectors in $\mathbb{R}^{2N}$, is continuous on the set $\{v : d(v, A) \geq n\}$, which is open and contains almost every vector and is thus dense in $\mathbb{C}^N$. Suppose that $v_i \to \check{v}$ and $d(v_i, A) \geq n > d(\check{v}, A)$ for all i. It suffices to show $F_n(v_i) \to 0 = F(\check{v})$. Let $\check{P}_n$ satisfy $\check{P}_n(0) = 1$, $\deg \check{P}_n = d(\check{v}, A)$ and $\check{P}_n(A)\check{v} = 0$. Then $0 \leq F_n(v_i) \leq ||\check{P}_n(A)v_i||^2/||v_i||^2 \to 0$ as $i \to \infty$. ■

The following theorem gives a sequence of simple results concerning $\psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(A)$.

Theorem 3.1. *Let $A \in \mathbb{K}^{N\times N}$.*

1. *$0 \leq \psi_{n,\mathbb{K}}(A) \leq \varphi_{n,\mathbb{K}}(A) \leq 1$ for all n.*
2. *$\psi_{0,\mathbb{K}}(A) = \varphi_{0,\mathbb{K}}(A) = 1$ and $\psi_{d(A),\mathbb{K}}(A) = \varphi_{d(A),\mathbb{K}}(A) = 0$.*
3. *For fixed A, $\psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(A)$ are each nonincreasing in n.*
4. *For any nonzero $c \in \mathbb{K}$, $\psi_{n,\mathbb{K}}(cA) = \psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(cA) = \varphi_{n,\mathbb{K}}(A)$*
5. *When $\mathbb{K} = \mathbb{R}$, $\psi_{n,\mathbb{R}}(A) \leq \psi_{n,\mathbb{C}}(A)$, while $\varphi_{n,\mathbb{R}}(A) = \varphi_{n,\mathbb{C}}(A)$.*
6. *For any m, n, $\psi_{mn,\mathbb{K}}(A) \leq [\psi_{m,\mathbb{K}}(A)]^n$ and $\varphi_{mn,\mathbb{K}}(A) \leq [\varphi_{m,\mathbb{K}}(A)]^n$.*
7. *For fixed n, $\psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(A)$ are each continuous in A.*
8. *GMRES(s) applied to $A \in \mathbb{K}^{N\times N}$ converges for every $r^{(0)} \in \mathbb{K}^N$ if and only if $\psi_{s,\mathbb{K}}(A) < 1$. Furthermore, $||r^{(ms)}||/||r^{(0)}|| \leq [\psi_{s,\mathbb{K}}(A)]^m$, where $r^{(ms)}$ denotes the GMRES(s) residual.*

Proof 1: Follows from

$$\psi_{n,\mathbb{K}}(A) = \sup_{\substack{r\in\mathbb{K}^N\\ r\neq 0}} \inf_{\substack{\deg P_n\le n\\ P_n(0)=1}} \frac{\|P_n(A)r\|}{\|r\|}$$

$$\le \sup_{\substack{r\in\mathbb{K}^N\\ r\neq 0}} \inf_{\substack{\deg P_n\le n\\ P_n(0)=1}} \sup_{\substack{r'\in\mathbb{K}^N\\ r'\neq 0}} \frac{\|P_n(A)r'\|}{\|r'\|} = \varphi_{n,\mathbb{K}}(A)$$

2–4: Follow trivially.

5: Clearly, $\psi_{n,\mathbb{R}}(A) \le \psi_{n,\mathbb{C}}(A)$. Now, for A real, let $P(z) = P_r(z) + \mathrm{i}P_i(z)$ have degree bounded by n and P_r, P_i real and $P(0) = 1$. Then

$$\inf_P \sup_{r\in\mathbb{R}^N} \frac{\|P_r(A)r\|}{\|r\|} \le \inf_P \sup_{r\in\mathbb{R}^N} \frac{\|(P_r(A) + \mathrm{i}P_i(A))r\|}{\|r\|} = \inf_P \sup_{r\in\mathbb{R}^N} \frac{\|P(A)r\|}{\|r\|}$$

$$= \varphi_{n,\mathbb{R}}(A) \le \inf_{P\text{ real}} \sup_{r\in\mathbb{R}^N} \frac{\|P(A)r\|}{\|r\|} = \inf_P \sup_{r\in\mathbb{R}^N} \frac{\|P_r(A)r\|}{\|r\|}$$

There exists a real polynomial $\tilde{P}$ which minimizes $\inf_{P\text{ real}} \sup_{r\in\mathbb{R}^N} \|P(A)r\|/\|r\| = \inf_{P\text{ real}} \|P(A)\|$. This can be seen by taking $\|P_i(A)\| \downarrow \inf_{P\text{ real}} \|P(A)\|$ and by boundedness taking a convergent subsequence $P_{i_j}(A) \to \tilde{P}(A)$. The above argument shows $\tilde{P}$ is also a minimizer for $\varphi_{n,\mathbb{R}}(A)$. Finally, since $\tilde{P}(A)$ is real,

$$\varphi_{n,\mathbb{C}}(A) = \inf_P \sup_r \frac{\|P(A)r\|}{\|r\|} \le \sup_r \frac{\|\tilde{P}(A)r\|}{\|r\|} = \sup_{r\in\mathbb{R}^N} \frac{\|\tilde{P}(A)r\|}{\|r\|} = \varphi_{n,\mathbb{R}}(A)$$

6: Using 5 and the definition of F_n and (1.9), $\inf_{P_n}$ in either case may be taken over polynomials in $\mathbb{K}$. To show the first result, note for any $r \in \mathbb{K}^N$ and $P_{m;i}$ a polynomial over $\mathbb{K}$ with $\deg P_{m;i} \le m$ and $P_{m;i}(0) = 1$,

$$\inf_{P_{mn}} \frac{\|P_{mn}(A)r\|}{\|r\|} \le \inf_{\{P_{m;i}\}_{i=1}^{n}} \frac{\|\prod_{i=1}^{n} P_{m;i}(A)r\|}{\|r\|}$$

$$= \inf_{\{P_{m;i}\}_{i=1}^{n-1}} \frac{\|\prod_{i=1}^{n-1} P_{m;i}(A)r\|}{\|r\|} \inf_{P_{m;n}} \frac{\|P_{m;n}(A)[\prod_{i=1}^{n-1} P_{m;i}(A)r]\|}{\|[\prod_{i=1}^{n-1} P_{m;i}(A)r]\|}$$

$$\le \inf_{\{P_{m;i}\}_{i=1}^{n-1}} \frac{\|\prod_{i=1}^{n-1} P_{m;i}(A)r\|}{\|r\|} \sup_{r'\in\mathbb{K}^N} \inf_{P_{m;n}} \frac{\|P_{m;n}(A)r'\|}{\|r'\|}$$

$$= \inf_{\{P_{m;i}\}_{i=1}^{n-1}} \frac{\|\prod_{i=1}^{n-1} P_{m;i}(A)r\|}{\|r\|} \psi_{m,\mathbb{K}}(A)$$

An induction argument completes the proof. The second result follows easily by taking P_m a polynomial over $\mathbb{K}$ and

$$\varphi_{mn,\mathbb{K}}(A) = \inf_{P_{mn}} \sup_{r\in\mathbb{K}^N} \frac{\|P_{mn}(A)r\|}{\|r\|} \le \inf_{P_m} \sup_{r\in\mathbb{K}^N} \frac{\|[P_m(A)]^n r\|}{\|r\|}$$

$$= \inf_{P_m} ||[P_m(A)]^n|| \leq \inf_{P_m} ||P_m(A)||^n = [\varphi_{m,\mathbb{K}}(A)]^n$$

7: The proof is straightforward but lengthy. A proof for $\mathbb{K} = \mathbb{C}$ is found in [14, sec. 3.2]; the result for $\mathbb{K} = \mathbb{R}$ can be proven similarly. 8: If $\psi_{s,\mathbb{K}}(A) < 1$, then clearly GMRES converges according to the bound shown. On the other hand, if $\psi_{s,\mathbb{K}}(A) = 1$, then by the Lemma, F_n attains its maximum of 1 for some unit vector $r' \in \mathbb{K}^N$, and setting $r^{(0)} = r'$ thus causes stagnation $r^{(ms)} = r^{(0)}$ for all m. ■

The following theorem gives some practical results which may be used to bound $\psi_n(A)$ in various cases.

Theorem 3.2. *Let $A \in \mathbb{K}^{N\times N}$.*

1. *Let $P^{-1}AP = A'$, for $P \in \mathbb{K}^{N\times N}$. Then $\psi_{n,\mathbb{K}}(A) \leq \mathrm{cond}(P)\psi_{n,\mathbb{K}}(A')$ and furthermore $\varphi_{n,\mathbb{K}}(A) \leq \mathrm{cond}(P)\varphi_{n,\mathbb{K}}(A')$, where $\mathrm{cond}(P) = ||P|| \cdot ||P^{-1}||$.*
2. *In particular, if $J(A) = \inf\{\mathrm{cond}(P) : \; AP = PJ \text{ Jordan decomposition}\}$ is defined as the Jordan condition number of A, then $\psi_n(A) \leq J(A)\psi_n(J)$ and $\varphi_n(A) \leq J(A)\varphi_n(J)$.*
3. *For unitary $U \in \mathbb{K}$, $\psi_{n,\mathbb{K}}(U^*AU) = \psi_{n,\mathbb{K}}(A)$ and $\varphi_{n,\mathbb{K}}(U^*AU) = \varphi_{n,\mathbb{K}}(A)$.*
4. *Let $A = PJQ^*$, $Q = P^{-*}$, be a Jordan decomposition, and define a partitioning $P = [P_1, \ldots, P_k]$, $Q = [Q_1, \ldots, Q_k]$, $J = \mathrm{diag}\{J_i\}_{i=1}^k$. Then $\varphi_n(A) \leq \inf_{P_n} \sum_{i=1}^k ||P_i|| \cdot ||Q_i|| \cdot ||P_n(J_i)||$.*
5. *Define $\hat{\varphi}_n(\Omega) = \inf_{P_n(0)=1} \sup_{z\in\Omega} |P_n(z)|$ for any $\Omega \subseteq \mathbb{C}$. Then $\varphi_n(J) = \hat{\varphi}_n(\sigma(J))$ for J diagonal and $\sigma(J)$ denoting the spectrum of J. Also, $\hat{\varphi}_n$ is monotone in the sense that $\Omega_1 \supseteq \Omega_2$ implies $\hat{\varphi}_n(\Omega_1) \geq \hat{\varphi}_n(\Omega_2)$ and $n_1 \leq n_2$ implies $\hat{\varphi}_{n_1}(\Omega) \geq \hat{\varphi}_{n_2}(\Omega)$.*
6. *For any $\epsilon > 0$, let $\Lambda_\epsilon = \cup_{||E||\leq\epsilon}\sigma(A + E)$, the ϵ- pseudospectrum of A. Then*

$$\hat{\varphi}_n(\sigma(A)) \leq \varphi_n(A) \leq \frac{\mathcal{L}(\partial\Lambda_\epsilon)}{2\pi\epsilon}\hat{\varphi}(\Lambda_\epsilon)$$

where $\mathcal{L}(\partial\Lambda_\epsilon)$ denotes the arc length of the boundary $\partial\Lambda_\epsilon$.

Proof The proofs of 1–5 are straightforward. For 6, see [24]. ■

The following bounds are valid when A is definite.

Theorem 3.3. *Suppose A is definite, and furthermore A_H is HPD (note if A is definite, there exists c such that $(cA)_H$ is HPD). Let $A_N = (A - A^*)/2$ denote the skew-Hermitian part of A, and $\rho_{min}(A) = \inf_v |v^*Av|/v^*v$. Then $\psi_1(A)^2$ is bounded by any of the following:*

$$1 - \rho_{min}(A)\rho_{min}(A^{-1}), \qquad 1 - \frac{\rho_{min}(A)^2}{||A||^2}, \qquad 1 - \frac{1}{||A_H^{-1}||^2||A||^2}$$

$$1 - \frac{1}{\mathrm{cond}(A_H)} \cdot \frac{1}{(1 + ||A_H^{-1}|| \cdot ||A_N||)^2}, \qquad 1 - \frac{1}{\mathrm{cond}(A_H)} \cdot \frac{1}{1 + ||A_N||^2||A_H^{-1}||/||A_H||}$$

Proof See [14, sec. 3.1], [18], and [4]. ■

The following new theorem gives some results on the limitations of polynomial methods. Specifically, a polynomial method may stagnate until the last step $n = d(A)$ owing to either a poor eigenvalue distribution or an ill-conditioned eigenvector matrix P ($AP = PJ$ denoting a Jordan decomposition).

Theorem 3.4. *Let $A \in \mathbb{K}^{N\times N}$.*

1. *For any nonsingular $P \in \mathbb{K}^{N\times N}$ and for any $n < N$ there exists nonsingular $A \in \mathbb{K}^{N\times N}$ and $\hat{A} = PAP^{-1}$ such that $\psi_{n,\mathbb{K}}(\hat{A}) = 1$.*
2. *For any nonsingular $A \in \mathbb{K}^{N\times N}$ and for any $n < d(A)$ there exists nonsingular $P \in \mathbb{K}^{N\times N}$ and $\hat{A} = PAP^{-1}$ such that $\psi_{n,\mathbb{K}}(\hat{A}) = 1$.*

Proof 1: Let $v \in \mathbb{K}^N$, $v \neq 0$. It is enough to construct A such that $d(v, \hat{A}) = N$ and $\hat{A}^i v \perp v$, $1 \leq i \leq N-1$, i.e. $A^i w \perp Mw$ where $w = P^{-1}v$, $M = P^*P$. Let $w_1 = w/\sqrt{w^*Mw}$, and let $\{w_i\}_{i=1}^N$ be an M-orthonormal basis, in the sense that $w_i^*Mw_j = \delta_{i,j}$. Let $A = w_1 w_N^* M + \sum_{i=1}^{n-1} w_{i+1}w_i^*M$. Note $w_i \in$ Range A for each i, so A is nonsingular. Also, $A^i w_1 = w_{i+1}$, $1 \leq i \leq N-1$, giving the desired result. 2: Given $v \in \mathbb{K}^N$ satisfying $d(v, A) > n$, it is sufficient to construct an HPD matrix $M \in \mathbb{K}^{N\times N}$ with $Mv \perp AK_n(v, A)$; then, letting $M = P^*P$ by Cholesky factorization, we have $P^*Pv \perp AK_n(v, A)$ or $(Pv) \perp \hat{A}K_n((Pv), \hat{A})$, and $d((Pv), \hat{A}) > n$. Here it is assumed that $K_n(v, A)$ is a linear space over $\mathbb{K}$.

There exists x such that $x \perp AK_n(v, A)$, $x \not\perp v$, that is, $x \in [AK_n(v, A)]^\perp \setminus \{v\}^\perp$. Such x exists unless $[AK_n(v, A)]^\perp \subseteq \{v\}^\perp$, i.e. $v \in AK_n(v, A)$, implying $d(v, A) \leq n$, which contradicts $n < d(v, A)$. It will suffice to find an HPD matrix M such that $Mv = x/v^*x$, or $vv^*x = M^{-1}x$.

Let $M' = vv^* + X^\perp X^{\perp *}$, where $X^\perp$ is defined to be some matrix whose columns form a basis for $\{x\}^\perp$ and which satisfies $X^{\perp *}X^\perp = I$. Clearly, M' is Hermitian, and $y^*M'y \geq 0$ for all y. We will now show that M' is nonsingular. Suppose $M'y = 0$. Then $vv^*y + X^\perp X^{\perp *}y = 0$. Since $v \notin \text{Range}(X^\perp) = \{x\}^\perp$, v together with the columns of $X^\perp$ form a basis for the entire space $\mathbb{K}^N$. Thus $v(v^*y) = 0$ and $X^\perp(X^{\perp *}y) = 0$; or, $v^*y = 0$ and $X^{\perp *}y = 0$. Since y is perpendicular to the entire space, it must be zero. Thus M' is nonsingular. We complete the proof by setting $M = (M')^{-1}$. ■

Various bounds for $\psi_n(A)$ are known which are based on properties of A or its spectrum; see for example [10], [27], [1], [8], [3] and [26].

If J the Jordan form of A is diagonal and $\sigma(A) \subseteq \Omega \subseteq \mathbb{C} \setminus \{0\}$ for some compact set Ω, then

$$\psi_n(A) \leq \varphi_n(A) \leq J(A)\varphi_n(J) \leq J(A)\hat{\varphi}_n(\Omega) \tag{3.1}$$

The links in this sequence of bounds may be loose in various cases. The first inequality is in fact an equality in some cases; it is not known whether $\psi_n(A) = \varphi_n(A)$ in general (see [16]). The second inequality may be very pessimistic, depending on the conditioning of the eigenvalues. The third inequality may be pessimistic if $\sigma(A)$ is not well distributed through Ω.

We conclude this section by giving bounds for $\hat{\varphi}_n(\Omega)$ for some special cases of Ω. In particular, let Ω be an ellipse in $\mathbb{C}$ with one of its axes contained within the real axis in $\mathbb{C}$. The ellipse is centered at d with horizontal axis of length $2a$ and vertical axis of length $2b$. It is assumed that $d - a > 0$, and the foci of the ellipse are given by $d \pm c$, where $c^2 = a^2 - b^2$. For $c \neq 0$ this elliptical region gives rise to the scaled and translated Chebyshev polynomials $\tilde{P}_n(\lambda) = T_n(\frac{\lambda-d}{c})/T_n(\frac{-d}{c})$ where $T_n(z) = \cosh(n \cosh^{-1}(z))$ (see [22]). These polynomials may be used to formulate bounds for $\hat{\varphi}_n(\Omega)$.

From [22] we see that the Chebyshev polynomials take on their maximum values at the intersections of the major axis with the boundary of the ellipse. This leads to the following results. When $c = 0$ (Ω a circle – see e.g. [21, p. 22]) and $c^2 > 0$ we obtain respectively

$$\hat{\varphi}_n(\Omega) = \left(\frac{a}{d}\right)^n, \qquad \hat{\varphi}_n(\Omega) \leq \left|\frac{T_n(\frac{-a}{c})}{T_n(\frac{-d}{c})}\right| = \frac{\cosh\ n\ \log\left(\frac{\sqrt{a+c}-\sqrt{a-c}}{\sqrt{a+c}+\sqrt{a-c}}\right)}{\cosh\ n\ \log\left(\frac{\sqrt{d+c}-\sqrt{d-c}}{\sqrt{d+c}+\sqrt{d-c}}\right)} \tag{3.2}$$

where $c > 0$. On the other hand, when $c^2 < 0$ we have the bound

$$\hat{\varphi}_n(\Omega) \leq \left|\frac{T_n(\frac{ib}{c})}{T_n(\frac{-d}{c})}\right| = \frac{\left(\frac{b}{\tilde{c}} + \sqrt{(\frac{b}{\tilde{c}})^2 - 1}\right)^n + \left(\frac{b}{\tilde{c}} + \sqrt{(\frac{b}{\tilde{c}})^2 - 1}\right)^{-n}}{\left(\frac{d}{\tilde{c}} + \sqrt{(\frac{d}{\tilde{c}})^2 + 1}\right)^n + \left(-\left(\frac{d}{\tilde{c}} + \sqrt{(\frac{d}{\tilde{c}})^2 + 1}\right)\right)^{-n}} \tag{3.3}$$

where $\tilde{c} = \mathrm{i}c > 0$. Situations in which these inequalities are in fact equalities are discussed in [7].

In the special case of $b = 0$, letting $\kappa \equiv (d + a)/(d - a)$ gives

$$\hat{\varphi}_n(\Omega) = 1/\cosh\left(n \log\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)\right) \tag{3.4}$$

4. Computational complexity of linear equation solution

In this section the theory of section 3 is applied to give estimates of the total work required to solve (1.1) by a restarted method. It is assumed that the stopping criterion

$$\frac{||r^{(n)}||}{||r^{(0)}||} \leq \zeta \tag{4.1}$$

is to be satisfied.

The work required to solve (1.1) is the average work per iteration multiplied by the number of iterations. A total of m restart cycles for convergence translates to $n = ms$ total iterations, assuming that the final restart cycle is brought to completion. The criterion (4.1) is satisfied if $\psi_s(A)^m \leq \zeta$ or

$$m \geq \frac{\log \zeta^{-1}}{-\log \psi_s(A)}$$

This combined with the results of section 2 yields

$$s(a_w s + b_w) \max\{1, \lceil \log \zeta^{-1}/(-\log \psi_s(A)) \rceil\} \tag{4.2}$$

as a bound on the work required to satisfy (4.1).

This gives some indication of the cost of using a restarted method rather than using a method with a short recurrence. If the polynomial (1.7)–(1.8) of degree n were applied directly to $r^{(0)}$ with only $c_w n$ total work for some constant c_w, then since $\psi_{ms}(A) \leq \psi_s(A)^m$, the work to calculate $u^{(ms)}$ from $u^{(0)}$ would be no more (and often less) than

$$s\ c_w \max\{1, \lceil \log \zeta^{-1}/(-\log \psi_s(A)) \rceil\} \tag{4.3}$$

Thus the cost of (4.2) relative to (4.3) is at least $(a_w s + b_w)/c_w$, which is large if the best value of s to make (4.2) small is large. However, as mentioned before, (4.3) is not practical since the polynomial (1.7)–(1.8) is usually not known, though some methods such as Chebyshev acceleration [23] and hybrid methods [25] attempt to satisfy (1.8) approximately using a short recurrence.

In this section we seek to determine the minimizer of (4.2) and the behavior of (4.2) at its minimizer, depending on the properties of A. Owing to the scarcity of analytic results on $\psi_s(A)$, it is necessary to focus on some special cases of A. These cases illustrate some of the possibilities for the behavior of (4.2), and also, owing to the continuity of the function ψ_s, give an idea of the general behavior of (4.2) near the special cases.

4.1. The case of A HPD

For this case the work function

$$w(s) = \frac{s(a_w s + b_w)\log\zeta^{-1}}{\log\cosh s \log\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}} \tag{4.4}$$

is used, based on the bound (3.4). Here, κ represents the condition number of A.

We prove a theorem to show $w(s)$ of (4.4) has a unique minimum value. It is necessary first to develop results on infinite series and discrete convolutions. For functions f and g defined on the nonnegative integers, we define the convolution $f * g$ by $(f * g)(n) = \sum_{i=0}^{n} f(i)g(n-i)$.

Lemma 4.1. *Define $\mathcal{H}(n) = 1/(n+1)$. Then the function $\mathcal{H} * \mathcal{H}$ is nonincreasing. Furthermore, $\mathcal{H} * \mathcal{H}$ restricted to the positive integers is strictly decreasing.*

Proof It is easily verified that $\mathcal{H} * \mathcal{H}(0) = \mathcal{H} * \mathcal{H}(1) = 1$. For $n > 1$,

$$\mathcal{H} * \mathcal{H}(n-1) - \mathcal{H} * \mathcal{H}(n) = -\frac{1}{n+1} + \sum_{i=0}^{n-1} \frac{1}{i+1}\frac{1}{(n-i)(n-i+1)}$$

$$> -\frac{1}{n+1} + \frac{1}{n}\sum_{i=0}^{n-1} \frac{1}{(n-i)(n-i+1)}$$

But

$$\sum_{i=0}^{n-1} \frac{1}{(n-i)(n-i+1)} = \frac{1}{(n+1)n} + \frac{1}{n(n-1)} + \ldots + \frac{1}{2\cdot 1} = \frac{n}{n+1}$$

by induction. Substituting gives $\mathcal{H} * \mathcal{H}(n-1) - \mathcal{H} * \mathcal{H}(n) > 0$. ■

Theorem 4.1. *The function $w(s)$ of (4.4) for $a_w > 0$, $b_w > 0$, $0 < \zeta < 1$ and $\kappa > 1$ has a unique minimizer s_{opt} on $(0, \infty)$ and is strictly monotone on either side of s_{opt}.*

Proof First let $\alpha = a_w \log\zeta^{-1}$, $\beta = b_w \log\zeta^{-1}$ and $\gamma = -\log\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$, $0 < \gamma < \infty$, so $w(s) = s(\alpha s + \beta)/\log\cosh s\gamma$. Clearly, $w(s) > 0$ for all s, and $w(s) \to \infty$ as s tends to 0 or ∞. Also, $w(s)$ has any number of derivatives on $(0, \infty)$. We show $w'(s) = 0$ has a unique solution.

The numerator of $w'(s)$ via the quotient rule is

$$(2\alpha s + \beta)\log(\cosh(s\gamma)) - s\gamma(\alpha s + \beta)\tanh(s\gamma) = 0 \tag{4.5}$$

We now show $(s\gamma)\tanh(s\gamma) - \log(\cosh(s\gamma)) > 0$. Setting $z = \tanh s\gamma, 0 < z < 1$, and multiplying by 2 yields the equivalent inequality

$$z(\log(1+z) - \log(1-z)) + \log(1+z) + \log(1-z) > 0$$

Expansion by power series of the form $\log(1-z) = -z\sum_{n=0}^{\infty}\mathcal{H}(n)z^n$, $\log(1+z) = z\sum_{n=0}^{\infty}\widetilde{\mathcal{H}}(n)z^n$, where $\mathcal{H}(n) = 1/(n+1)$ and $\widetilde{\mathcal{H}}(n) = (-1)^n\mathcal{H}(n)$, yields

$$z\sum_{n=1}^{\infty}(1-(-1)^n)\mathcal{H}(n-1)z^n - z\sum_{n=1}^{\infty}(1-(-1)^n)\mathcal{H}(n)z^n > 0$$

This is clearly true, due to the monotonicity of $\mathcal{H}$.

We may thus solve (4.5) for β/α, yielding

$$\frac{\beta}{\alpha} = f(s) \equiv -s\frac{2\log(\cosh(s\gamma)) - (s\gamma)\tanh(s\gamma)}{\log(\cosh(s\gamma)) - (s\gamma)\tanh(s\gamma)}$$

We show f is strictly increasing, and $\lim_{s\downarrow 0} f(s) = 0$ and $\lim_{s\uparrow\infty} f(s) = \infty$. By Taylor expansion around zero, $\lim_{s\downarrow 0} f(s) = 0$.

It suffices to show $\tilde{f}(s\gamma) \equiv f(s)/s$ is increasing. The numerator of $\frac{\partial}{\partial s}\tilde{f}(s\gamma)$ by the quotient rule is

$$s\gamma^2\tanh^2(s\gamma) - s\gamma^2\log\cosh(s\gamma)/\cosh^2(s\gamma) - \gamma\tanh(s\gamma)\log\cosh(s\gamma)$$

Dividing by γ and again substituting $z = \tanh s\gamma$ yields

$$z^2\log\left(\frac{1+z}{1-z}\right)^{1/2} - (1-z^2)\log\left(\frac{1+z}{1-z}\right)^{1/2}\log(1-z^2)^{-1/2} - z\log(1-z^2)^{-1/2}$$

Multiplying by 4 and simplifying gives

$$2z(1+z)\log(1+z) + 2z(1-z)\log(1-z) + (1-z^2)((\log(1+z))^2 - (\log(1-z))^2)$$

Expanding by power series as before and noting $\widetilde{\mathcal{H}} * \widetilde{\mathcal{H}}(n) = (-1)^n\mathcal{H} * \mathcal{H}(n)$ yields

$$2z^2(1+z)\sum_{n=0}^{\infty}(-1)^n\mathcal{H}(n)z^n - 2z^2(1-z)\sum_{n=0}^{\infty}\mathcal{H}(n)z^n$$

$$-(1-z^2)z^2\sum_{n=0}^{\infty}(1-(-1)^n)\mathcal{H} * \mathcal{H}(n)z^n$$

Low-order terms cancel, leaving

$$\sum_{n=4}^{\infty}z^n\Big[2((-1)^n - 1)\mathcal{H}(n-2) + 2(1-(-1)^n)\mathcal{H}(n-3)$$

$$-(1-(-1)^n)(\mathcal{H}*\mathcal{H}(n-2)-\mathcal{H}*\mathcal{H}(n-4))\Big]$$

For even terms, the Taylor coefficient vanishes; otherwise, the z^n coefficient is

$$-4\Big[\mathcal{H}(n-2)-\mathcal{H}(n-3)\Big]-2\Big[\mathcal{H}*\mathcal{H}(n-2)-\mathcal{H}*\mathcal{H}(n-4)\Big]$$

By the monotonicity of $\mathcal{H}$ and $\mathcal{H}*\mathcal{H}$, this is positive, thus $\check{f}$ is increasing. ■

We now show that for large κ,

$$s_{opt} \doteq \left(\frac{3b_w}{4a_w}\right)^{1/3} \kappa^{1/3}$$

Let $x = s\gamma$ and $x_{opt} = s_{opt}\gamma$. From the previous proof, $(b_w/a_w) = s_{opt}\check{f}(x_{opt})$.

We show that $x_{opt} \downarrow 0$ as $\kappa \uparrow \infty$. Note for large κ, by the definition of γ, $x \doteq 2s/\sqrt{\kappa}$. Then $x_{opt}\check{f}(x_{opt}) \doteq 2b_w/(a_w\sqrt{\kappa}) \to 0$ as $\kappa \uparrow \infty$. It was shown earlier that $\check{f}(x) \equiv x\check{f}(x)$ is continuous and strictly increasing, mapping $[0,\infty)$ to $[0,\infty)$. Thus, $\check{f}$ has a continuous inverse, and $x_{opt} = \check{f}^{-1}(x_{opt}\check{f}(x_{opt})) \to 0$ as $\kappa \to \infty$.

By expanding $\check{f}$ by power series in x and dropping low order terms, we have $\check{f}(x) \doteq x^2/3$. Substituting gives the result.

Finally, we calculate $w(s_{opt})$. From power series expansion, we obtain for sufficiently large κ

$$w(s_{opt}) = \frac{s_{opt}(a_w s_{opt} + b_w)\log\zeta^{-1}}{\log\cosh s\gamma} \doteq \frac{a_w\kappa}{2}\log\zeta^{-1}$$

We now compare this with the work estimates for other methods. For a restarted method with a small *fixed* value of s and large κ,

$$w(s) = \frac{s(a_w s + b_w)\log\zeta^{-1}}{\log\cosh s \log(1-\frac{2}{\sqrt{\kappa}+1})} \doteq \frac{s(a_w s + b_w)\log\zeta^{-1}}{\log\cosh(\frac{2s}{\sqrt{\kappa}})} \doteq \frac{a_w\kappa}{2}\log\zeta^{-1}\left(1+\frac{b_w}{a_w s}\right)$$

The conjugate residual method [2] requires approximately $\frac{\sqrt{\kappa}}{2}\log 2\zeta^{-1}$ iterations to converge to (4.1) for large κ, with fixed work per iteration. Based on this estimate, the work to converge to (4.1) by an *unrestarted* full-recurrence minimal residual algorithm such as GMRES for large κ is approximately $a_w\frac{\kappa}{4}[\log 2\zeta^{-1}]^2 + b_w\frac{\sqrt{\kappa}}{2}\log 2\zeta^{-1} \doteq a_w\frac{\kappa}{4}[\log 2\zeta^{-1}]^2$.

We conclude that a restarted method applied to the HPD case is suboptimal, in the sense that the total work is proportional to κ, even using s_{opt}, whereas the short-recurrence conjugate residual method requires work proportional to $\sqrt{\kappa}$. On the other hand, the above results show how much can be gained by using s_{opt} rather than an arbitrary small value of s. Of course, choosing s to be greater than some small fixed value may have even greater benefit for other spectral distributions, such as when A has a few isolated eigenvalues or clusters ([26], [3]).

4.2. *Case of the spectrum of A within an ellipse*

We now consider a second case in which A is normal with spectrum contained in an ellipse as described in section 3, assumed here to satisfy $c^2 > 0$, i.e. the foci are real and distinct. Inequality (3.2) yields the upper bound $\hat{\varphi}_s(\Omega) \leq \phi_s$, where ϕ_s is defined to be the right-hand side quantity of (3.2). To formulate a lower bound, note

$$\phi_n = \frac{\alpha^n}{\beta^n}\left(\frac{1+\alpha^{-2n}}{1+\beta^{-2n}}\right) \doteq \frac{\alpha^n}{\beta^n}$$

$$\alpha \equiv \frac{\sqrt{a+c}+\sqrt{a-c}}{\sqrt{a+c}-\sqrt{a-c}} > 1, \qquad \beta \equiv \frac{\sqrt{d+c}+\sqrt{d-c}}{\sqrt{d+c}-\sqrt{d-c}} > 1$$

for large n independent of a, c and d, as long as α and β are bounded above 1. Note here $a < d$ implies $\alpha < \beta$. We claim that $(\alpha/\beta)^s$ is a lower bound for $\hat{\varphi}_s(\Omega)$. Using an argument analogous to that of Theorem 3.1 part 6, we obtain $\hat{\varphi}_s(\Omega) \geq [\hat{\varphi}_n(\Omega)^{1/n}]^s$, for any n divisible by s. Then $\hat{\varphi}_s(\Omega) \geq [\lim_{n\to\infty} \hat{\varphi}_n(\Omega)^{1/n}]^s$. But, by the asymptotic optimality of the Chebyshev polynomials [21, p. 24], $\lim_{n\to\infty} \hat{\varphi}_n(\Omega)^{1/n} = \lim_{n\to\infty} \phi_n^{1/n} = (\alpha/\beta)$, giving the result.

These observations lead to the following upper and lower bounds on the work required to solve (1.1):

$$w_-(s) \equiv \frac{(a_w s + b_w)\log\zeta^{-1}}{-\log(\alpha/\beta)} \leq w(s) \equiv \frac{s(a_w s + b_w)\log\zeta^{-1}}{-\log\hat{\varphi}_s(\Omega)}$$

$$\leq w_+(s) \equiv \frac{s(a_w s + b_w)\log\zeta^{-1}}{-\log\phi_s} \tag{4.6}$$

For this case, the following assumptions are made:

- The center of the ellipse d is fixed.
- The condition of the problem is allowed to increase as follows: let $(d-a)/d = \epsilon$, i.e. $a = d(1-\epsilon)$, where $\epsilon > 0$ approaches zero. Note the condition is $\kappa = (d+a)/(d-a) = (2-\epsilon)/\epsilon \doteq 2/\epsilon$.
- The quantity $\delta \equiv c/a$ is fixed independent of ϵ, $0 < \delta < 1$. This fixes the shape of the ellipse independent of ϵ. Note that therefore $\alpha > 1$ is also fixed.

First let us examine the behavior of $w_+(s)$ for large κ and fixed s. Letting $\delta' \equiv c/d = \delta(1-\epsilon)$ yields

$$\phi_s = \frac{\cosh s \log\left(\frac{\sqrt{1+\delta}-\sqrt{1-\delta}}{\sqrt{1+\delta}+\sqrt{1-\delta}}\right)}{\cosh s \log\left(\frac{\sqrt{1+\delta'}-\sqrt{1-\delta'}}{\sqrt{1+\delta'}+\sqrt{1-\delta'}}\right)} = \frac{\cosh s \log\alpha}{\cosh s \log\beta}$$

since cosh is even. For small ϵ, power series expansion yields after some manipulation $\beta \doteq \alpha(1+\epsilon/\sqrt{1-\delta^2})$. By further power series expansion we obtain

$$\cosh(s\log\beta) \doteq \cosh(s\log\alpha) + \frac{s\epsilon}{\sqrt{1-\delta^2}}\sinh(s\log\alpha)$$

Thus

$$\phi_s \doteq 1 - \frac{s\epsilon}{\sqrt{1-\delta^2}}\tanh(s\log\alpha)$$

and

$$w_+(s) \doteq \frac{(a_w s + b_w)\sqrt{1-\delta^2}\log(\zeta^{-1})}{\epsilon\tanh(s\log\alpha)} \equiv \left(\frac{1}{\epsilon}\right)g(s)$$

Table 2. Value of $s_{opt,g}$ for large κ

$\frac{c}{a}$	0.2	0.5	0.8	0.9	0.95	0.99	0.999	0.9999
$s_{opt,g}$	0.870	1.332	2.157	2.885	3.770	6.764	15.074	32.995

Importantly, $g(s)$ has no dependence on ϵ.

Now let us consider $w_-(s)$. We have for small ϵ

$$w_-(s) = \frac{(a_w s + b_w)\log\zeta^{-1}}{\log(1+\epsilon/\sqrt{1-\delta^2})} > \frac{(a_w s + b_w)\sqrt{1-\delta^2}\log\zeta^{-1}}{2\epsilon} \equiv \left(\frac{1}{\epsilon}\right)h(s)$$

Importantly, this lower bound holds independently of s.

We now demonstrate that for any $\gamma > 0$, there exists s' such that $g(s'') + \gamma < h(s')$ for some $s'' < s'$. Otherwise, for any $s > 0$ and any integer $k > 0$, $g(s) + \gamma \geq h(sk)$. But since $\tanh(s\log\alpha) \doteq 1$ for large s, $g(s) \leq c_1 s + c_2$, for some $c_1, c_2 > 0$ and sufficiently large s. Then for some other constants $c_3, c_4 > 0$, $c_1 s + c_2 + \gamma \geq c_3 sk + c_4$, based on the definition of h. This condition can be violated for sufficiently large s and k.

Now, for such γ, s' and s'', it is clear that $\epsilon w(s) \geq \epsilon w_-(s) > h(s) \geq h(s')$ for any $s \geq s'$ and for any $\epsilon > 0$, by the definitions of w_- and h. Furthermore, there exists $\epsilon' > 0$ such that for all $\epsilon \leq \epsilon'$, $\epsilon w(s'') \leq \epsilon w_+(s'') \leq g(s'') + \gamma < h(s')$ for such $s'' < s'$, by the pointwise convergence of ϵw_+ to g as well as the result of the previous paragraph.

What we have shown is that for small ϵ (and thus large κ), the minimizer s_{opt} of w must be bounded below some s' which is independent of the condition κ. Thus, the ratio between $w(s_{opt})$ where s_{opt} depends on κ, and $w(s)$ for fixed s independent of κ, is itself bounded independent of κ.

We now consider in more detail the behavior of $w_+(s)$. As noted above, $\epsilon \cdot w_+$ converges pointwise to g as $\epsilon \downarrow 0$. Note also that $g(s) \to \infty$ as s tends to zero or ∞. We now show the second derivative g'' is positive on $(0, \infty)$, so g has a unique minimum.

The second derivative of $(1/\tanh(s))$ is $(2\cosh(s)/\sinh^3(s)) > 0$. Furthermore, the second derivative of $(s/\tanh(s))$ is $2(s\cosh s - \sinh s)/\sinh^3 s$. The Taylor expansion of $(s\cosh s - \sinh s)$ is $\sum_{n \text{ odd}}(n-1)s^n/n!$, whose terms are all nonnegative. We conclude that for any a_w, b_w, the function g has positive second derivative.

For illustrative purposes, computed values for the minimizer $s_{opt,g}$ of $g(s)$ are now given. Here the work per iteration $(a_w s + b_w)$ is specified by $a_w = 1.5, b_w = 7.5$. Note that ellipses with smaller aspect ratios give higher values of $s_{opt,g}$, in accord with the previous result that $\lim_{\kappa\to\infty} s_{opt} = \infty$ in the case of a degenerate ellipse (the HPD case).

Thus the work required to solve (1.1) by a restarted method for this case is proportional to $1/\epsilon$, or equivalently, κ. We now compare this to the estimated work for nonsymmetric Chebyshev acceleration, which utilizes a short recurrence but requires prior knowledge of the ellipse. For n large and κ fixed, $\phi_n \doteq (\alpha/\beta)^n$, yielding an estimated $n \doteq \log\zeta^{-1}/\log(\beta/\alpha)$ iterations. Further manipulation yields

$$n \doteq \frac{\sqrt{1-\delta^2}\log\zeta^{-1}}{\epsilon}$$

Thus, the restarted method is optimal, in the sense of having the same proportionality dependence on the condition of A as the short recurrence method.

4.3. *Results for other cases*

The actual performance of a restarted method as the problem size becomes large depends on the properties of the matrices A as the problem size grows. The behavior depends on such factors as the spectrum and the non-normality of the class of problems being considered.

When A is non-normal, the bound (3.1) suggests that convergence may be hindered by the non-normality. It is useful to consider how many iterations n are required to render $J(A)\varphi_n(\Omega) \leq 1$. For example, when the estimate (3.4) is used for the case when the Jordan form J of A is HPD, then the derivations of subsection 4.1 yield

$$s_{opt} \geq \frac{\sqrt{\kappa}}{2}\operatorname{arccosh}(J(A))$$

where $\kappa = \operatorname{cond}(J)$. This estimate suggests the need in some cases to perform a significant number of iterations in order to overcome the non-normality of A.

Convergence of the minimal residual method may be improved if the spectrum of A is in clusters or contains a few isolated eigenvalues. In such cases, the work function (4.2) may be expected to have one or more dropoffs as s increases, though it is difficult to predict such effects beforehand. For this reason it seems better to overestimate s than to underestimate it. This is also true for difficult problems such as non-Hermitian indefinite problems, for which $\psi_s(A)$ may be quite near 1 until s becomes fairly large.

5. An adaptive restarting strategy

The above theoretical results on s_{opt} in practice may only be applied to a few special cases. It is desirable to develop an automated procedure for determining when to restart, a strategy which can be applied to arbitrary matrices A. This is the goal of the current section.

Let $u^{(n-1)}$ denote the most recent iterate computed, and let $m < n-1$ denote the iteration of the most recent restart, so that $u^{(n-1)}$ results from $(n-1-m)$ steps of the minimal residual method applied to $u^{(m)}$. A strategy is necessary to decide whether the current iterate $u^{(n)}$ is to be computed either as step $(n-m)$ of the minimal residual method applied to $u^{(m)}$ (no restart), or as step 1 of the minimal residual method applied to $u^{(n-1)}$ (restart performed). Making these decisions during the run permits the flexibility of restart cycles of different lengths during the run.

To decide whether to restart, a measure of the efficiency of the iteration process is needed. Let us define

$$\operatorname{eff}(r^{(old)}, r^{(new)}) = \frac{-\log\left(\frac{\|r^{(new)}\|}{\|r^{(old)}\|}\right)}{\operatorname{work}(old, new)}$$

a measure of the efficiency of calculating $u^{(new)}$ from $u^{(old)}$. Here, $\operatorname{work}(old, new)$ is the CPU time used to calculate $u^{(new)}$ given $u^{(old)}$, measured for example by the formulas of Table 1. This efficiency function corresponds roughly to the reciprocal of the work function defined earlier.

The criterion for restarting is defined as follows. Let $M_s(r)$ be the result of applying s steps of the unrestarted minimal residual method (2,5) to residual r: $M_s : r^{(0)} \mapsto r^{(s)}$.

Thus, using the above notation, $r^{(n-1)} = M_{n-1-m}(r^{(m)})$. Then a choice is made to compute $r^{(n)} = M_1(M_{n-1-m}(r^{(m)}))$ (a restart) rather than $r^{(n)} = M_{n-m}(r^{(m)})$ (no restart) if

$$\mathrm{eff}[r^{(m)}, M_1(M_{n-1-m}(r^{(m)}))] > \mathrm{eff}[r^{(m)}, M_{n-m}(r^{(m)})]$$

The test effectively compares the polynomials associated with M_{n-1-m} and M_{n-m}, using a sequence of $(n-m)$ iterates in either case to do so.

To effect this comparison, it is necessary to obtain estimates of the two possibilities for $||r^{(n)}||$, before the quantity $r^{(n)}$ is actually computed. This can be done by computing certain inner products associated with the computation of $r^{(n)}$. This is described as follows for the Orthomin and GMRES variants.

For Orthomin,

$$||r^{(n)}||^2 = ||r^{(n-1)}||^2 - \frac{|(r^{(n-1)}, Ar^{(n-1)})|}{(Aq^{(n-1)}, Aq^{(n-1)})}$$

where

$$(Aq^{(n-1)}, Aq^{(n-1)}) = \begin{cases} (Ar^{(n-1)}, Ar^{(n-1)}), & \text{restart case} \\ (Ar^{(n-1)}, Ar^{(n-1)}) - \displaystyle\sum_{i=m}^{n-2} \frac{|(Aq^{(i)}, Ar^{(n-1)})|^2}{(Aq^{(i)}, Aq^{(i)})}, & \text{unrest. case} \end{cases}$$

The inner products $(Aq^{(i)}, Ar^{(n-1)})$ are used for the computation of the coefficients $\beta_{n-1,i}$ when no restart is performed; otherwise they are discarded. The result of this added work plus the computation of the added inner product $(Ar^{(n-1)}, Ar^{(n-1)})$ per step amounts to $2k$ extra inner products per each k step cycle, which is fairly negligible compared to the rest of the work of the iteration.

For GMRES, the estimate of $||r^{(n)}||$ for the unrestarted case may be determined before the basis vector $q^{(n)}$ itself is formed. For simplicity of notation, let $m = 0$. Then, as noted in section 2, the residual norm $||r^{(n)}||$ may be estimated using information from the QR factorization of H_{n+1}. The rightmost column of H_{n+1} may be formed using $h_{i+1,n} = (q^{(i)}, Aq^{(n-1)})$ as in the standard algorithm, while $h^2_{n+1,n} = (Aq^{(n-1)}, Aq^{(n-1)}) - \sum_{i=0}^{n-1} |(q^{(i)}, Aq^{(n-1)})|^2$ may be computed before $q^{(n)}$ is computed. If no restart is performed, then $h_{n+1,n}$ may be recomputed directly from $\hat{q}^{(n)}$, to obtain a safer value. If a restart is performed, the values $h_{i+1,n}$ are left unused, and the total extra work per k-step cycle is $(2k+1)$ inner products, which again is negligible, amounting to about 2 extra inner products per iteration. Furthermore, if a restart is performed, then $Ar^{(n-1)}$ must be computed from $Aq^{(n-1)}$, as demonstrated below.

For the $||r^{(n)}||$ estimate for the restarted case for GMRES, the formula

$$||r^{(n)}||^2 = ||r^{(n-1)}||^2 - \frac{|(r^{(n-1)}, Ar^{(n-1)})|^2}{(Ar^{(n-1)}, Ar^{(n-1)})}$$

is used. This requires values for $(r^{(n-1)}, Ar^{(n-1)})$ and $(Ar^{(n-1)}, Ar^{(n-1)})$. Note

$$Ar^{(n-1)} = h_{1,0}Q_2H_2 - Q_{n+1}H_{n+1}H_n y_{n-1} = Q_{n+1}\left[\begin{bmatrix} h_{1,0}H_2 \\ 0 \end{bmatrix} - H_{n+1}P_n \begin{bmatrix} \check{z}_{n-1} \\ 0 \end{bmatrix}\right]$$

using the notation from section 2. Letting $H_{n+1} = [\, h_1^{n+1} \quad \ldots \quad h_n^{n+1} \,]$ yields

$$H_{n+1} P_n = \left[\begin{bmatrix} H_n \\ 0 \end{bmatrix} \quad h_n^{n+1} \right] \begin{bmatrix} P_{n-1} & 0 \\ 0 & 1 \end{bmatrix} R_{n,n-1} = \left[\begin{bmatrix} H_n P_{n-1} \\ 0 \end{bmatrix} \quad h_n^{n+1} \right] R_{n,n-1}$$

Define

$$t_{n+1} = H_{n+1} P_n \begin{bmatrix} \check{z}_{n-1} \\ 0 \end{bmatrix}$$

$$= \left[\begin{bmatrix} H_n P_{n-1} \\ 0 \end{bmatrix} \begin{bmatrix} I_{n-2} \\ 0 \end{bmatrix} \quad \begin{bmatrix} H_n P_{n-1} \\ 0 \end{bmatrix} e_{n-1} \quad h_n^{n+1} \right] R_{n,n-1} \begin{bmatrix} \check{z}_{n-2} \\ \hat{\zeta}_{n-1} \\ 0 \end{bmatrix}$$

Since

$$R_{n,n-1} \begin{bmatrix} \check{z}_{n-2} \\ \hat{\zeta}_{n-1} \\ 0 \end{bmatrix} = \begin{bmatrix} \check{z}_{n-2} \\ c_{n-1} \hat{\zeta}_{n-1} \\ s_{n-1} \hat{\zeta}_{n-1} \end{bmatrix}$$

we have

$$t_{n+1} = \begin{bmatrix} t_n \\ 0 \end{bmatrix} + \begin{bmatrix} v_n \\ 0 \end{bmatrix} (c_{n-1} \hat{\zeta}_{n-1}) + h_n^{n+1} (s_{n-1} \hat{\zeta}_{n-1})$$

where

$$v_{n+1} = H_{n+1} P_n e_n = \left[\begin{bmatrix} H_n P_{n-1} \\ 0 \end{bmatrix} \begin{bmatrix} I_{n-2} \\ 0 \end{bmatrix} \quad \begin{bmatrix} H_n P_{n-1} \\ 0 \end{bmatrix} e_{n-1} \quad h_n^{n+1} \right] R_{n,n-1} e_n$$

$$= \left[\begin{bmatrix} H_n P_{n-1} \\ 0 \end{bmatrix} \begin{bmatrix} I_{n-2} \\ 0 \end{bmatrix} \quad \begin{bmatrix} v_n \\ 0 \end{bmatrix} \quad h_n^{n+1} \right] \begin{bmatrix} 0 \\ -s_{n-1}^* \\ c_{n-1}^* \end{bmatrix} = -s_{n-1}^* \begin{bmatrix} v_n \\ 0 \end{bmatrix} + c_{n-1}^* h_n^{n+1}$$

From t_{n+1} the desired inner products are computed: $(Ar^{(n-1)}, Ar^{(n-1)}) = \tilde{t}_{n+1}^* \tilde{t}_{n+1}$ and

$$(r^{(n-1)}, Ar^{(n-1)}) = \left[\begin{bmatrix} h_{1,0} \\ 0 \end{bmatrix} - \begin{bmatrix} P_n \begin{bmatrix} \check{z}_{n-1} \\ 0 \end{bmatrix} \\ 0 \end{bmatrix} \right]^* \tilde{t}_{n+1}, \quad \tilde{t}_{n+1} = \begin{bmatrix} h_{1,0} H_2 \\ 0 \end{bmatrix} - t_{n+1}$$

Finally,

$$Ar^{(n-1)} = Q_{n+1} \tilde{t}_{n+1} = \frac{e_{n+1}^* \tilde{t}_{n+1}}{h_{n+1,n}} Aq^{(n-1)} + \sum_{i=0}^{n-1} \left(e_{i+1}^* \tilde{t}_{n+1} - e_{n+1}^* \tilde{t}_{n+1} \frac{h_{i+1,n}}{h_{n+1,n}} \right) q^{(i)}$$

This computation requires $(k + 1)$ extra SAXPYs per k-step cycle, or about one extra SAXPY per iteration.

To determine the computational work per iteration, the system timer routine on the given computer is used to obtain w_{mv}, w_{sax} and w_{dot}, the cost for the matrix–vector product, SAXPY operation and dot product, respectively. The formulas of Table 1 are used, with the additional work per iteration of $(2s) w_{dot}/s$ for Orthomin and $(2s+1) w_{dot}/s + (s+1) w_{sax}/s$ for GMRES, s here denoting the length of the current cycle.

The modified Orthomin and GMRES algorithms are summarized as follows:

Orthomin with adaptive restarting

- 1. Initialization. $n = 1$.

- 2. Perform stopping test.
- 3. Form $A \cdot r^{(n-1)}$.
- 4. Calculate all dot products for new basis vector.
- 5. Estimate norms for next residual with/without restart.
- 6. Perform test to determine whether to restart.
- 7. Calculate new basis vector $p^{(n-1)}$ based on decision.
- 8. Form new iterate $u^{(n)}$ and residual $r^{(n)}$.
- 9. $n \leftarrow n + 1$; go to 2.

GMRES with adaptive restarting

- 1. Initialization. $n = 1$.
- 2. Perform stopping test.
- 3. Compute dot products for new basis vector $\hat{q}^{(n)}$. Estimate norm.
- 4. Update QR factorization. Estimate residual norms with/without restart.
- 5. Perform test to determine whether to restart.
- 6. If necessary, restart and obtain $u^{(n-1)}$, $r^{(n-1)}$, $Ar^{(n-1)}$.
- 7. Compute new basis vector based on decision; compute norm explicitly.
- 8. Recompute QR update from true vector norm. Recompute residual norm estimates.
- 9. Calculate other GMRES parameters. Compute coefficients of $u^{(n)}$, $r^{(n)}$ in terms of basis vectors.
- 10. $n \leftarrow n + 1$; go to 2.

6. Numerical results

In this section numerical experiments with these algorithms are given. The model problem used is based on a two-dimensional convection–diffusion problem:

$$-u_{xx}(x,y) - u_{yy}(x,y) + Du_x(x,y) = G(x,y) \text{ on } \Omega = [0,1]^2, \quad u(x,y) = 1 + xy \text{ on } \partial\Omega$$

Central five-point finite differences are used, with uniform mesh spacing h, yielding a matrix of size $N = (n_h - 1)^2$, with $n_h = 1/h$. The true solution is $u(x,y) = 1 + xy$. The constant D is used to control the amount of nonsymmetry of the problem.

The experiments are performed on the University of Texas System Cray Y-MP 8/864 vector computer, used in single-processor mode with single precision. The system timer routine SECOND is used to time the SAXPY, dot product and matrix–vector product operations. Initial iterate $u^{(0)} = 0$ and stopping test $\|r^{(n)}\|/\|r^{(0)}\| < \zeta = 10^{-5}$ are used.

The restarted GMRES algorithm with several fixed choices of s is compared to GMRES with the adaptive restarting strategy. Both the unpreconditioned case and the case of left preconditioning with the modified incomplete LU (MILU) [11] preconditioning are considered.

For the unpreconditioned cases, the adaptive strategy in all cases does nearly as well as the best choice of s among those tried for the standard method. This can offer considerable savings, since a good choice of s is usually not known beforehand. For the symmetric case, the adaptive strategy actually performed considerably better than the methods with fixed restart frequency. For the MILU-preconditioned cases, all methods performed fairly well, owing to the fact that a rather small number of iterations was required in all cases.

Table 3. Model problem, $n_h = 256$, no preconditioning. CPU seconds

Meth\Dh:	0	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{4}	2^{5}
Adaptive	11.093	11.496	6.219	5.523	4.821	5.485	5.791	5.243	6.692	9.043
$s = 5$	45.397	25.508	12.002	5.957	5.921	5.807	6.064	5.992	6.420	7.629
$s = 10$	56.737	12.242	7.793	7.445	8.061	7.947	7.974	7.552	7.015	6.551
$s = 15$	64.222	10.686	8.690	9.420	9.848	10.374	10.218	9.680	8.814	8.668
$s = 20$	56.422	10.090	10.438	11.832	12.209	12.051	12.526	11.471	11.039	10.459
$s = 25$	52.504	11.339	12.619	14.331	14.923	15.110	14.790	14.079	13.264	13.165

Table 4. Model problem, $n_h = 256$, MILU preconditioning. CPU seconds

Meth\Dh:	0	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{4}	2^{5}
Adaptive	1.338	1.278	1.439	1.838	2.484	2.362	1.723	1.042	0.595	0.370
$s = 5$	1.908	1.190	1.456	1.895	2.345	2.335	1.807	1.109	0.637	0.372
$s = 10$	1.341	1.491	1.909	2.038	2.417	2.440	1.815	1.085	0.647	0.396
$s = 15$	1.349	1.580	1.858	2.337	2.663	2.494	1.776	1.107	0.615	0.399
$s = 20$	1.500	1.408	1.854	2.231	2.582	2.288	1.748	1.055	0.636	0.399

The adaptive restarting strategy seems most well-suited to problems for which A is definite and thus convergence is assured for any s. Even for these cases, however, it is often impossible to predict beforehand whether the spectrum of A has certain properties such as clusters which might cause a sharp drop in the work function for some (possibly high) value of the restart frequency. This appears to be an inherent limitation of the Krylov space information. Furthermore, when A is non-Hermitian and indefinite, the convergence behavior of full GMRES may be fairly flat and nearly stagnant for a large period of the convergence history. For such cases it may be desirable to make the restart frequency comparatively large to overcome near-stagnation and seek a drop in the work function. Again, it is difficult for any strategy to predict beforehand the existence of such a drop in the residual based merely on the information obtained during a period of slow convergence, though such dropoff behavior commonly occurs (see e.g. [15]).

7. Conclusions

In this paper theoretical results on the work requirements for restarted GMRES have been presented for various cases. Also, an adaptive strategy has been developed and shown to be effective for certain problems. Further research is needed to determine more economical and reliable means of solving general non- Hermitian linear systems.

Acknowledgements

This work was supported in part by the National Science Foundation through Grant DCR-8518722, by the Department of Energy through Grant DE-FG05-87ER25048, and by Cray Research Inc. through Grant LTRDTD 1/18/90, with the University of Texas at Austin. This work was also supported in part by the Department of Energy through grant W-7405-ENG-36, with Los Alamos National Laboratory.

REFERENCES

1. S. F. Ashby. Polynomial preconditioning for conjugate gradient methods. Ph. D. thesis, University of Illinois at Urbana-Champaign, Department of Computer Science, Report UIUCDCS-R-87-1355, 1987.
2. S. F. Ashby, Thomas A. Manteuffel and Paul E. Saylor. A taxonomy for conjugate gradient methods. *SIAM J. Numer. Anal.*, 27, 1542–1568, 1990.
3. O. Axelsson. A restarted version of a generalized preconditioned conjugate gradient method. *Communications in Applied Numerical Methods*, 4, 521–530, 1988.
4. S. C. Eisenstat, H. C. Elman and M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, 20, 345–357, 1983.
5. V. Faber, and T. Manteuffel. Necessary and sufficient conditions for the existence of a conjugate gradient method. *SIAM J. Numer. Anal.*, 21, 352–362, 1984.
6. V. Faber, and T. Manteuffel. Orthogonal error methods. *SIAM J. Numer. Anal.*, 24 170–187, 1987.
7. B. Fischer and R. Freund. Chebyshev polynomials are not always optimal. *J. Approx. Theory*, 65, 261–272, 1991.
8. R. Freund. On polynomial preconditioning and asymptotic convergence factors for indefinite Hermitian matrices. *Lin. Alg. Appl.*, 154–156, 259–288, 1991.
9. R. W. Freund and N. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numer. Math.*, 60, 315–339, 1991.
10. A. Greenbaum. Comparison of splittings used with the conjugate gradient algorithm. *Numer. Math.*, 33, 181–194, 1979.
11. I. Gustafsson. Modified incomplete Cholesky (MIC) method. In *Preconditioning Methods: Theory and Applications*, D. J. Evans, editor, pages 265–293. Gordon Breach, New York, 1983.
12. M. H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related numerical algorithms, part I. *SIAM J. Matrix Anal. Appl.*, 13, 594–639, 1992.
13. M. R. Hestenes and E. L. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, 49, 409–436, 1952.
14. W. Joubert. Iterative methods for the solution of nonsymmetric systems of linear equations. the University of Texas at Austin, Center for Numerical Analysis, Report CNA-242, 1990.
15. W. Joubert. Lanczos methods for the solution of nonsymmetric systems of linear equations. *SIAM J. Matrix Anal. Appl.*, 13, 926–943, 1992.
16. W. Joubert. A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems. *SIAM J. Sci. Comput.*, 15, 427–439, 1994.
17. W. D. Joubert and G. F. Carey. Parallelizable restarted iterative methods for nonsymmetric linear systems. Part I: Theory, Part II: Parallel implementation. *InternationalJournal ofComputer Mathematics*, 44, 243–267, 1992.
18. W. D. Joubert and T. A. Manteuffel. Iterative methods for nonsymmetric linear systems. In *Iterative Methods for Large Linear Systems*, D. R. Kincaid and L. J. Hayes, editors, pages 149–171. Academic Press, Boston MA, 1990.
19. W. D. Joubert and D. M. Young. Necessary and sufficient conditions for the simplification of generalized conjugate gradient algorithms, *Lin. Alg. Appl.*, 88/89, 449–485, 1987.
20. C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards*, 45, 255- -282, 1950.
21. T. A. Manteuffel. An iterative method for solving nonsymmetric linear systems with dynamic estimation of parameters. Ph. D. thesis, University of Illinois at Urbana-Champaign, Dept. of Computer Science, Report UIUCDCS-R-75-758, 1975.
22. T. A. Manteuffel. The Tchebychev iteration for nonsymmetric linear systems. *Numer. Math.* 28, 307–327, 1977.
23. T. A. Manteuffel. Adaptive procedure for estimation of parameters for the nonsymmetric Tchebychev iteration. *Numer. Math.*, 31, 187–208, 1978.
24. N. Nachtigal, S. C. Reddy and L. N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Appl.*, 13, 778–795, 1992.
25. N. M. Nachtigal, L. Reichel and L. N. Trefethen. A hybrid GMRES algorithm for nonsymmetric linear systems. *SIAM J. Matrix Anal. Appl.*, 13, 796–825, 1992.

26. Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 7, 856–869, 1986.
27. A. van der Sluis and H. A. van der Vorst. The rate of convergence of conjugate gradients. *Numer. Math.*, 48, 543–560, 1986.
28. H. A. van der Vorst. Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comp.*, 13, 631–644, 1992.
29. P. K. W. Vinsome. ORTHOMIN, an iterative method for solving sparse sets of simultaneous linear equations. *Fourth Symposium of Numerical Simulation of Reservoir Performance of the Society of Petroleum Engineers of the AIME*, Los Angeles, Paper SPE 5739, 1976.