NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2005; **12**:453–470
Published online 3 May 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.436

Superlinear convergence in minimum residual iterations

I. Kaporin*,†

Center for Supercomputer and Massively Parallel Applications, Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 119991, Russia

SUMMARY

The superlinear convergence of minimum residual-type methods for solving systems of linear equations with diagonalizable non-singular unsymmetric matrix is estimated using a special conditioning measure. For the construction of the latter, the distance from the spectrum of the matrix to the origin and the Frobenius distance between the matrix and the identity is used. Asymptotical exactness of the presented result is discussed theoretically and illustrated numerically. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: linear system of equations; unsymmetric matrix; minimum residual-type methods; superlinear convergence; ILU preconditioning; approximate inverse preconditioning

1. INTRODUCTION

Consider the solution of the linear system

$$Mx = b \tag{1}$$

where $x, b \in \mathbf{C}^n$ and M is a general non-singular unsymmetric matrix.

We shall estimate the rate of convergence of the minimum residual Krylov subspace method to solve (1). For an efficient numerical implementation of such methods one can refer to References [1, 2]. When the preconditioning is used, the (implicitly defined) matrix M approaches the identity matrix in a certain sense, and this can be used in the iteration convergence analysis.

*Correspondence to: I. E. Kaporin, Center for Supercomputer and Massively Parallel Applications, Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 119991, Russia.
†E-mail: kaporin@ccas.ru

Contract/grant sponsor: National Science Foundation (NWO); contract/grant number: 047.008.007
Contract/grant sponsor: Program for Fundamental Research of Russian Academy of Sciences (RAS); contract/grant number: 17

Received 18 April 2004
Revised 2 October 2004

Let x^k be the kth iterate in the minimum residual method and let x^0 be an initial vector. The residual vector $r^k = b - Ax^k$ is then related to the initial residual by

$$r^k = P_k(M)r^0 \tag{2}$$

for a particular $P_k \in \pi_k$, where π_k is the set of polynomials of the degree not greater than k and normalized by the condition $P_k(0) = 1$. By the definition of the method, the polynomial is optimal in the sense that for the Euclidean norm $\|v\| = (v^{\mathrm{H}}v)^{1/2}$ one has

$$\|r^k\| \leqslant \|\widetilde{P}_k(M)r^0\| \tag{3}$$

for any $\widetilde{P}_k \in \pi_k$. Considering problems of a large size n, we are primarily interested in the cases where the residual norm is sufficiently small for $1 \ll k \ll n$. Relation (3) shows that

$$\|r^0\| \geqslant \|r^1\| \geqslant \cdots \geqslant \|r^k\| \geqslant \|r^{k+1}\| \geqslant \cdots \geqslant \|r^n\| = 0 \tag{4}$$

Moreover, (3) will further be used to derive upper and lower bounds for the residual norm depending on certain spectral characteristics of A.

Currently, there is a couple of results establishing the convergence of $\|r^k\|$ to zero with linear or even superlinear rates, see References [1, 3–7] and references cited therein. Note that an analysis of Reference [6] based on the use of relationships between Krylov subspaces and certain invariant subspaces of M, allows to estimate the convergence of minimum residual-type methods with no assumptions on the diagonalizability of M.

Typically, these results refer to spectral properties of M which cannot be easily related to the use of a preconditioning strategy. For instance, when the spectrum of M does not fit a straight segment of the complex plane, there is no 'classical' Chebyshev polynomial-based estimate for the convergence rate. Here, one would use Faber polynomials for a simply connected area $\mathscr{D}$ in the complex plane containing the spectrum of M. However, such $\mathscr{D}$, first, cannot be defined uniquely and, second, even if specified, its shape can hardly be controlled at the preconditioning stage. At the same time, the asymptotic behaviour of Faber polynomials critically depends on the shape of $\mathscr{D}$ in the neighbourhood of the origin.

In contrast, in the present paper we actually discuss the conditions under which the squared Frobenius distance

$$\gamma \equiv \|I - M\|_{\mathrm{F}}^2 = (\mathrm{trace}(I - M)^{\mathrm{H}}(I - M))$$

between the preconditioned matrix and the identity can be a satisfactory measure of the convergence in the minimum residual method. Note that in the so-called sparse approximate inverse (SAI) preconditioning the quantity γ is directly minimized, see References [3, 8], while in the other cases a certain upper bound for it is actually reduced (e.g. in drop-tolerance ILU-based preconditionings, see Section 4).

A similar result on superlinear rate of convergence for the 'conjugate gradient' (CG) method can be found in References [9, 10], where the so-called K-condition number is used. For earlier presentations of related topics see Reference [3] and references cited therein. It should be stressed that spectral condition number bounds of the type

$$\kappa(M) \equiv \|M\|\|M^{-1}\| \leqslant (1 + \sqrt{\gamma})/(1 - \sqrt{\gamma}), \quad \gamma < 1$$

(cf. Theorem 8.10 in Reference [3] and related papers), cannot be a justification to the above-mentioned SAI preconditionings. Typically, even after the use of a preconditioning, γ is still

much larger than 1 because its reduction below 1 would involve excessive computational costs for large-scale problems. In other words, as far as $\|I - M\|_F > 1$, the reduction of $\|I - M\|_F$ is not correlated, in general, with an improvement of the conditioning of M. For this reason, we also consider an additional characterization of M, that is, $\delta = \min|\lambda(M)|$, determining the separation of the spectrum of M apart from the origin.

The remainder of the paper is organized as follows. In Section 2 we present a new convergence estimate and the corresponding iteration number bound for the minimum residual method. Asymptotical correctness of the estimate is studied theoretically in Section 3, where a lower bound for the residual norm is presented. In Section 4, some issues related to preconditioning are discussed, while Section 5 considers illustrative numerical examples. In conclusion, a couple of auxiliary technical results are given in the Appendix.

2. SUPERLINEAR CONVERGENCE OF THE RESIDUAL NORM

For the minimum residual method we have

$$\|r^k\| = \min_{P_k \in \pi_k} \|P_k(M)r^0\| \tag{5}$$

where $\|u\| = (u^H u)^{1/2}$. For the sake of simplicity, let M be diagonalizable,

$$M = V\Lambda V^{-1} \tag{6}$$

Here, the columns of V are the (normalized) eigenvectors $v_1, v_2, \ldots, v_n$ of M and the entries of the diagonal matrix Λ are the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of M. Using (3), one finds

$$\begin{aligned} \|r^k\| &= \|P_k(M)r^0\| \leqslant \|\widetilde{P}_k(M)r^0\| = \|\widetilde{P}_k(V\Lambda V^{-1})r^0\| = \|V\widetilde{P}_k(\Lambda)V^{-1}r^0\| \\ &\leqslant \|V\|\|\widetilde{P}_k(\Lambda)\|\|V^{-1}\|\|r^0\| = \kappa\|\widetilde{P}_k(\Lambda)\|\|r^0\| = \kappa \max_{1 \leqslant i \leqslant n} |\widetilde{P}_k(\lambda_i)|\|r^0\| \end{aligned} \tag{7}$$

which holds for any $\widetilde{P}_k \in \pi_k$. Note that hereafter, the notation

$$\kappa = \|V\|\|V^{-1}\| \tag{8}$$

is used to denote the condition number of V, where $\|\cdot\|$ is the matrix norm induced by the Euclidean scalar product.

Non-trivial estimates for $\max_{1 \leqslant i \leqslant n} |\widetilde{P}_k(\lambda_i)|$ are typically obtained via separation of the spectrum of M into cluster part and outlying part, see, for instance References [11, 12] for the case of SPD matrix M, and References [1, 4, 7] for the general case.

2.1. A new residual norm estimate

The rate of convergence of the iteration error $\|r^k\|$ can be characterized by the average convergence factor $q_k = (\|r^k\|/\|r^0\|)^{1/k}$. Next we present a simple but non-trivial superlinear convergence result which shows that $q_k = O(k^{-1/2})$.

Theorem 1
Let $k<n$ and the eigenvalues of M satisfy

$$|\lambda_i| \geqslant \delta, \quad i=1,\ldots,n \tag{9}$$

where $0<\delta\leqslant 1$. Then the following residual norm bound:

$$\frac{\|r^k\|}{\|r^0\|} \leqslant \kappa \left(\frac{\mathrm{e}}{\delta}\right)^{\gamma} \left(\frac{6.924\gamma}{k}\right)^{k/2} \tag{10}$$

where κ is determined by (6), (8),

$$\gamma = \|I-M\|_{\mathrm{F}}^2 \tag{11}$$

and $\mathrm{e}=\exp(1)\approx 2.718$, holds for the minimum residual method.

Proof
Let the eigenvalues of M be numbered as follows:

$$|1-\lambda_1| \geqslant |1-\lambda_2| \geqslant \cdots |1-\lambda_n| \geqslant 0$$

Consider the polynomial $\widetilde{P}_k$ of the form

$$\widetilde{P}_k(\lambda) = \prod_{i=1}^{k} \left(1-\frac{\lambda}{\lambda_i}\right)$$

In view of $\widetilde{P}_k(\lambda_j)=0,\ 1\leqslant j\leqslant k$, and the eigenvalue ordering used, residual norm bound (7) yields

$$\begin{aligned}
\frac{1}{\kappa}\cdot\frac{\|r^k\|}{\|r^0\|} &\leqslant \max_{1\leqslant j\leqslant n} |\widetilde{P}_k(\lambda)| = \max_{k<j\leqslant n} |\widetilde{P}_k(\lambda)| = \max_{k<j\leqslant n} \left|\prod_{i=1}^{k}\left(1-\frac{\lambda_j}{\lambda_i}\right)\right| \\
&= \max_{k<j\leqslant n} \prod_{i=1}^{m} \frac{|(1-\lambda_i)-(1-\lambda_j)|}{|\lambda_i|} \leqslant \max_{k<j\leqslant n} \prod_{i=1}^{k} \frac{|1-\lambda_i|+|1-\lambda_j|}{|\lambda_i|} \\
&= \prod_{i=1}^{k} \frac{|1-\lambda_i|+|1-\lambda_{k+1}|}{|\lambda_i|} \leqslant \left(2^k \prod_{i=1}^{k} \frac{1}{|\lambda_i|}\right) \prod_{i=1}^{k} |1-\lambda_i|
\end{aligned}$$

We now use the inequality

$$2^k \prod_{i=1}^{k} \frac{1}{|\lambda_i|} \leqslant 2^k \prod_{i=1}^{k} (\mathrm{e}^{0.2743}(\mathrm{e}/\delta)^{(1-|\lambda_i|)^2}) \leqslant \mathrm{e}^{0.96745k} \left(\frac{\mathrm{e}}{\delta}\right)^{\sum_{1\leqslant i\leqslant k}|1-\lambda_i|^2}$$

which readily follows from the assertion of Lemma A1, taken with $t=|\lambda_i|$, condition (9), and the elementary inequality

$$(1-|\lambda|)^2 \leqslant |1-\lambda|^2, \quad \lambda\in C$$

Using also the estimate

$$\prod_{i=1}^{k} |1-\lambda_i|^2 \leqslant \left(\frac{1}{k}\sum_{i=1}^{k} |1-\lambda_i|^2\right)^k$$

which holds by the inequality between the arithmetic and geometric mean values, one has

$$\frac{1}{\kappa}\cdot\frac{\|r^k\|}{\|r^0\|}\leqslant\left(\frac{\mathrm{e}}{\delta}\right)^{\sum_{1\leqslant i\leqslant k}|1-\lambda_i|^2}\left(\frac{\mathrm{e}^{1.9349}}{k}\sum_{i=1}^{k}|1-\lambda_i|^2\right)^{k/2}$$

It only remains to use the Weil inequality [13]

$$\sum_{i=1}^{n}|\lambda_i(C)|^2\leqslant\sum_{i=1}^{n}\sigma_i^2(C)=\mathrm{trace}(C^{\mathrm{H}}C)=\|C\|_{\mathrm{F}}^2$$

with $C=I-M$. By $\lambda_i(C)=1-\lambda_i(M)$, one has

$$\sum_{i=1}^{k}|1-\lambda_i|^2\leqslant\sum_{i=1}^{n}|1-\lambda_i|^2\leqslant\|I-M\|_{\mathrm{F}}^2=\gamma$$

which completes the proof. □

The above estimate can be considered as an adaptation and an essential refinement of the techniques presented in Reference [14]. In our case, the latter corresponds to the use of the trivial estimate $2^k\prod_{i=1}^{k}|\lambda_i|^{-1}\leqslant(2/\delta)^k$ instead of the result of Lemma A1 and yields a rather rough upper bound

$$\frac{\|r^k\|}{\|r^0\|}\leqslant\kappa\left(\frac{(2/\delta)^2\gamma}{k}\right)^{k/2}$$

Note that in practice δ is typically a small parameter, the numerical value of which can hardly be estimated *a priori*. However, the estimate in Theorem 1 (cf. also Theorem 2) demonstrates a rather slight dependence on δ. This estimate depends most essentially on the Frobenius distance to the identity $\sqrt{\gamma}=\|I-M\|_{\mathrm{F}}$ which represents an 'integral' characterization of the preconditioned matrix M. Hence, one can consider γ as a target function to be minimized when applying a preconditioning (provided that the preconditioned eigenvalues stay well bounded apart from zero). A further discussion of the relation of estimate (10) to certain preconditionings can be found in Section 4.

Remark 1

The constant 6.924 in (10) can be taken somewhat smaller (namely, not larger than 5.260) which can be shown using

$$\widetilde{P}_k(\lambda)=(1-\lambda)^{k-m}\prod_{i=1}^{m}\left(1-\frac{\lambda}{\lambda_i}\right)$$

with an appropriate $1\leqslant m<k$. Such an improvement requires rather cumbersome analysis, and therefore is not discussed here. Moreover, in the next section it is actually shown that a particular value of this constant does not affect the main term $(2+2/\mathrm{e})\gamma\log(\mathrm{e}/\delta)/\log\log(\mathrm{e}/\delta)$ in the corresponding iteration number estimate (14) of Theorem 2.

Remark 2

In view of (4), the above estimate is non-trivial only if $c(\delta)\gamma<k<n$. (The value of $c(\delta)=O(\log(\mathrm{e}/\delta)/\log\log(\mathrm{e}/\delta))$ grows with $1/\delta$, cf. the next section.) Hence, the superlinear behaviour can be observed only if $\gamma<n/c(\delta)$ and if the iteration number k is sufficiently large.

Remark 3
Actually, it is implied that the matrix M is already scaled in such a way that the squared Frobenius distance $\|I-M\|_{\mathrm{F}}^2$ attains its minimum. Hence, in general, one considers the quantity

$$\gamma_* = \min_{\sigma \in C^1} \|I - \sigma M\|_{\mathrm{F}}^2 = n - |\mathrm{trace}(M)|^2 / \|M\|_{\mathrm{F}}^2$$

where the optimum scaling factor is $\sigma_* = \mathrm{trace}(\overline{M})/\|M\|_{\mathrm{F}}^2$. This actually gives us an estimate depending on homogeneous functions in M, which seems to be a natural property for any iteration error bound. In practice, the exact value of such scaling constant σ can hardly be known, but it is typically close to 1 for sufficiently strong preconditionings (such as ones considered in Section 4).

2.2. The corresponding iteration number bound

To estimate the iteration number k from the residual norm bound (10) we will use the following proposition.

Lemma
Let $t > 0$ and

$$s = \frac{1 + (1 + \mathrm{e}^{-1})t}{\log(\mathrm{e} + t)} \tag{12}$$

where $\mathrm{e} = \exp(1)$; then the inequality

$$s \log s \geqslant t \tag{13}$$

holds.

Proof
Using the inequality $-\log z \geqslant -\mathrm{e}^{-1} z$ with appropriate $z > 0$, one has

$$\begin{aligned}
s \log s &= \frac{1 + (1 + \mathrm{e}^{-1})t}{\log(\mathrm{e} + t)} \log\left(\frac{1 + (1 + \mathrm{e}^{-1})t}{\log(\mathrm{e} + t)}\right) \\
&\equiv \frac{1 + (1 + \mathrm{e}^{-1})t}{\log(\mathrm{e} + t)} \left(\log(\mathrm{e} + t) - \log\left(\frac{\mathrm{e} + t}{1 + (1 + \mathrm{e}^{-1})t} \log(\mathrm{e} + t)\right)\right) \\
&\geqslant \frac{1 + (1 + \mathrm{e}^{-1})t}{\log(\mathrm{e} + t)} \left(\log(\mathrm{e} + t) - \mathrm{e}^{-1} \frac{\mathrm{e} + t}{1 + (1 + \mathrm{e}^{-1})t} \log(\mathrm{e} + t)\right) \\
&= 1 + (1 + \mathrm{e}^{-1})t - \mathrm{e}^{-1}(\mathrm{e} + t) = t
\end{aligned}$$

□

Note that $t > 0$ and (12) readily yield $s > 1$, and by the monotone increase of the function $s \log s$ for such s, the equality in condition (12) can be replaced by the '$\geqslant$' sign with the same assertion (13) holding.

Remark 4
The above lemma gives a fairly close approximation to the solution of the equation $s \log s = t$. Indeed, it can be shown that (12) implies $0 < (s \log s - t)/t < 0.064$ for all $t > 0$.

Now we can prove the following minimum residual iteration number bound.

Theorem 2
The iteration number k sufficient for the ε times reduction of the residual norm in the minimum residual method satisfies

$$k \leqslant \left\lceil \frac{6.924\gamma + (2 + 2\mathrm{e}^{-1})(\gamma \log(\mathrm{e}/\delta) + \log(\kappa/\varepsilon))}{\log(\mathrm{e} + (2/6.924)(\log(\mathrm{e}/\delta) + (1/\gamma)\log(\kappa/\varepsilon)))} \right\rceil \tag{14}$$

with γ, δ, and e defined as in Theorem 1.

Proof
Denoting

$$s = \frac{k}{6.924\gamma}, \quad t = \frac{2}{6.924} \log \frac{\mathrm{e}}{\delta} + \frac{2}{6.924\gamma} \log \frac{\kappa}{\varepsilon}$$

one can see that the required inequality $\|r^k\|/\|r^0\| \leqslant \varepsilon$ follows from (10) if condition (13) holds. By the above lemma, a sufficient condition for the latter is (12), which takes the form

$$k \leqslant \frac{6.924\gamma + (2 + 2\mathrm{e}^{-1})\gamma \log(\mathrm{e}/\delta) + (2 + 2\mathrm{e}^{-1}) \log(\kappa/\varepsilon)}{\log(\mathrm{e} + (2/6.924) \log(\mathrm{e}/\delta) + (2/6.924\gamma) \log(\kappa/\varepsilon))}$$

The use of the closest integer from above is valid, since the function $s \log s$ increases for $s \geqslant 1/\mathrm{e}$ and $s \geqslant 1$ by (12). □

3. ON OPTIMALITY OF THE ITERATION NUMBER BOUND

In this section, we will show that if the above iteration number estimate is applicable, it cannot be essentially improved.

Theorem 3
Let δ and $\tilde{\gamma}$ satisfy

$$0 < \delta < 0.0409 \tag{15}$$

$$1 < \tilde{\gamma} < \frac{1}{\Phi(\delta)} \left(\max \left(\frac{1}{n}, \log \frac{1}{1 - \delta} \right) \right)^{-1} \tag{16}$$

where

$$\Phi(\delta) = \frac{\log((1 - \delta)/\rho) + \log((1 - \rho)/\delta)}{(1 + \delta^2) \log((1 - \delta)/\rho) + \rho^2 \log((1 - \rho)/\delta)} = O\left(\frac{\log(1/\delta)}{\log\log(1/\delta)} \right) \tag{17}$$

and $\rho = \rho(\delta)$ is defined by

$$\frac{1}{\rho} = \sqrt{1 + 2 \log \frac{1}{\delta}} \tag{18}$$

Then there exist a diagonalizable matrix M with eigenvalues λ_j such that

$$\min_{1\leqslant j\leqslant n} |\lambda_j| = \delta \tag{19}$$

satisfying

$$\tilde{\gamma} - 0.4059 \leqslant \gamma = \|I - M\|_{\mathrm{F}}^2 \leqslant \tilde{\gamma} + 1 \tag{20}$$

and an n-vector r^0 such that

$$\|r^l\| = \min_{P_l(0)=1} \|P_l(M)r^0\| \geqslant \frac{1}{2\mathrm{e}} \|r^0\|, \quad l = 1, 2, \ldots, k-1 \tag{21}$$

where $r^l = P_l(M)r^0$ is generated by the minimum residual method, $\mathrm{e} = \exp(1)$, and

$$k = \lfloor \tilde{\gamma}\Phi(\delta) \rfloor \geqslant \left\lfloor \frac{(\gamma - 1)(1 + 2\log(1/\delta))}{1 + \log(1 + 2\log(1/\delta))} \right\rfloor \tag{22}$$

Proof
It is sufficient to consider the case of diagonal matrix $M = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, so that $V = I$ and $\kappa = 1$. Consider the following construction:

$$\lambda_j = \begin{cases} \delta \exp\left(\dfrac{2\pi \mathrm{i}}{m} j\right), & 1 \leqslant j \leqslant m \\ 1 - \rho \exp\left(\dfrac{2\pi \mathrm{i}}{k-m}(j-m)\right), & m < j \leqslant k \\ 1, & k < j \leqslant n \end{cases} \tag{23}$$

where $\mathrm{i} = \sqrt{-1}$, the quantities ρ and k are defined by (18) and (22), respectively, and

$$\begin{aligned} m &= \left\lceil k \frac{\log((1-\delta)/\rho)}{\log((1-\delta)/\rho) + \log((1-\rho)/\delta)} \right\rceil \\ &= k \frac{\log((1-\delta)/\rho)}{\log((1-\delta)/\rho) + \log((1-\rho)/\delta)} + \theta, \quad 0 \leqslant \theta < 1 \end{aligned} \tag{24}$$

Note also that (22) is equivalent to

$$k = \tilde{\gamma}\Phi(\delta) - \vartheta, \quad 0 \leqslant \vartheta < 1 \tag{25}$$

The condition $|\lambda_j| \geqslant \delta$ is clearly satisfied by $|\lambda_j| \geqslant \min(\delta, 1-\rho)$ whenever

$$\rho < 1 - \delta \tag{26}$$

holds. The latter is a consequence of (18) and (15).

Furthermore, by (23) one has

$$\|I - M\|_{\mathrm{F}}^2 = \sum_{j=1}^{n} |1 - \lambda_j|^2 = m(1 + \delta^2) + (k - m)\rho^2 \tag{27}$$

Substituting into the latter equality the expressions for m and k given in (24) and (25), respectively, yields

$$\|I-M\|_{\mathrm{F}}^2=\frac{k}{\Phi(\delta)}+(1+\delta^2-\rho^2)\theta=\tilde{\gamma}-\frac{1}{\Phi(\delta)}\vartheta+(1+\delta^2-\rho^2)\theta$$

Using $\rho\geqslant\delta$ (which holds by (18) and by the inequality $\log(\delta^{-2})\leqslant\delta^{-2}-1$), one gets

$$\|I-M\|_{\mathrm{F}}^2\leqslant\tilde{\gamma}+1$$

On the other hand, using the result of Lemma A4 and (15) one has

$$\|I-M\|_{\mathrm{F}}^2\geqslant\tilde{\gamma}-1/\Phi(0.0409)\geqslant\tilde{\gamma}-1/\tilde{\Phi}(0.0409)\geqslant\tilde{\gamma}-0.4059$$

and therefore (20) is proved.

It only remains to deduce the lower bound on $\|P_{k-1}(M)r^0\|$ based on the result of Lemma A2. Using the first k eigenvalues of M (which are pairwise different by (26)), one has by (23) and (A4)

$$Q_k(z)=(-1)^{k-m}(z^m-\delta^m)((1-z)^{k-m}-\rho^{k-m})$$

$$zQ_k'(z)=(-1)^{k-m}mz^m((1-z)^{k-m}-\rho^{k-m})+(z^m-\delta^m)(k-m)z(1-z)^{k-m-1}$$

Therefore,

$$|Q_k(0)|=\delta^m(1-\rho^{k-m})$$

$$|\lambda_jQ_k'(\lambda_j)|=m\delta^m\left|\left(1-\delta\exp\left(\frac{2\pi\mathrm{i}j}{m}\right)\right)^{k-m}-\rho^{k-m}\right|\geqslant m\delta^m((1-\delta)^{k-m}-\rho^{k-m})$$

for $j=1,\ldots,m$ and

$$|\lambda_jQ_k'(\lambda_j)|=(k-m)\rho^{k-m-1}\left|\left(1-\rho\exp\left(\frac{2\pi\mathrm{i}(j-m)}{k-m}\right)\right)^{m}-\delta^m\right|\left|1-\rho\exp\left(\frac{2\pi\mathrm{i}(j-m)}{k-m}\right)\right|$$

$$\geqslant(k-m)(1-\rho)\rho^{k-m-1}((1-\rho)^m-\delta^m)$$

for $j=m+1,\ldots,k$. By Lemma A2, this yields

$$\frac{\|r^0\|}{\|r^{k-1}\|}\leqslant\sum_{j=1}^{k}\left|\frac{Q_k(0)}{\lambda_jQ_k'(\lambda_j)}\right|\leqslant\frac{1-\rho^{k-m}}{(1-\delta)^{k-m}-\rho^{k-m}}+\frac{\delta^m(1-\rho^{k-m})}{(1-\rho)\rho^{k-m-1}((1-\rho)^m-\delta^m)}$$

In order to simplify the latter estimate, let us strengthen condition (26) by replacing it with

$$1-\rho^{k-m}\leqslant\frac{(1-\delta)^{k-m}-\rho^{k-m}}{(1-\delta)^k}\tag{28}$$

which is equivalent to

$$\rho\leqslant(1-\delta)\left(\frac{1-(1-\delta)^m}{1-(1-\delta)^k}\right)^{1/(k-m)}$$

Comparing the latter inequality with the result of Lemma A3 taken with $t = 1 - \delta$, one can see that a sufficient condition for (28) to hold is simply

$$\rho \leqslant (1 - \delta)/2 \tag{29}$$

which, in its turn, follows from (15) and (18). Indeed, combining (18) with (29) gives the condition

$$\frac{4}{(1 - \delta)^2} \leqslant 1 + 2 \log \frac{1}{\delta}$$

an analysis of which shows its validity for all $0 < \delta \leqslant 0.1226$.

Hence, condition (26) can further be replaced by (29). Using now (28) as the upper bound for $1 - \rho^{k-m}$, one can obtain

$$\begin{aligned} \frac{\|r^0\|}{\|r^{k-1}\|} &\leqslant \frac{1}{(1 - \delta)^k} \left(1 + \frac{\delta^m((1 - \delta)^{k-m} - \rho^{k-m})}{(\rho^{-1} - 1)\rho^{k-m}((1 - \rho)^m - \delta^m)}\right) \\ &= \frac{1}{(1 - \delta)^k} \left(1 + \left(\frac{\rho}{1 - \rho}\right) \frac{\delta^m(1 - \delta)^{k-m} - \delta^m\rho^{k-m}}{\rho^{k-m}(1 - \rho)^m - \delta^m\rho^{k-m}}\right) \end{aligned} \tag{30}$$

We now observe that definition (24) of m yields

$$m \geqslant k \frac{\log(1 - \delta)/\rho}{\log[(1 - \delta)/\rho] + \log(1 - \rho)/\delta}$$

which is equivalent to

$$\delta^m(1 - \delta)^{k-m} \leqslant \rho^{k-m}(1 - \rho)^m \tag{31}$$

Using the latter with (30) readily gives

$$\|r^{k-1}\|/\|r^0\| \geqslant (1 - \rho)(1 - \delta)^k \tag{32}$$

Since $1 - \rho > 1/2$ by (29) and $k \leqslant \tilde{\gamma}\Phi \leqslant 1/(-\log(1 - \delta))$ by the equality in (22) and the right inequality in (16), inequality (32) immediately yields (21). Finally, the inequality in (22) follows from the result of Lemma A4 and the right inequality in (20). □

Remark 5

Using (18), (17), and (22), one can see that the minimum residual iteration number needed to satisfy the convergence criterion $\|r^k\|/\|r^0\| \leqslant \varepsilon$ with $\varepsilon \leqslant 1/(2e)$ is $k = O(\gamma \log(\delta^{-1})/\log\log(\delta^{-1}))$. Comparing this with the result of Theorem 2 shows the asymptotic optimality of the latter with respect to the matrix-dependent parameters $1/\delta \gg 1$ and $\gamma = \|I - M\|_F^2 \gg 1$. Moreover, the ratio of the upper to the lower bound tends to $(2 + 2\exp(-1))/2 < 1.368$ asymptotically.

Remark 6

The above 'optimality' of the result of Theorem 2 implies only its asymptotic 'unimprovability', quite similar to the standard Chebyshev polynomial-based estimate for the CG iterations via the spectral condition number. One should keep in mind that a relatively small γ is

only sufficient and *not necessary* for a good convergence in GMRES method, just similar to the condition number not necessarily being small for the CG method to be efficient.

4. RELATION TO PRECONDITIONING ISSUES

The conditioning measures and the convergence estimate presented above are not aimed to an accurate prediction of the iteration number. Rather, these give a qualitative characterization of the effect which can be expected when applying certain classes of preconditionings in the solution of unsymmetric systems.

4.1. Sparse approximate inverse preconditionings

We first note that the iteration number estimate presented above seems to give the lacking theoretical background for the numerous Frobenius norm minimization techniques introduced during the last two decades, see Reference [8] and references cited therein. These SAI preconditionings are of the form $M = GA$, where G is a sparse matrix chosen to minimize the quadratic functional $\|I - M\|_{\mathrm{F}}^2$. The SAI preconditionings were intended to achieve high parallel performance in linear iterative solvers. In some simple cases, e.g. when A is diagonally dominant, it is even possible to prove nice spectral bounds for M, see Reference [3] and references cited therein. In such cases, one can guarantee that the spectrum of the preconditioned matrix is separated from the origin. Unfortunately, in general there may be no lower bound on δ in (9). For instance, in Reference [15] for the related factorized SAI preconditioning (referred there as Incomplete Inverse Cholesky) applied to general 'symmetric positive definite' matrices, a family of matrices A was found for which the condition number of M is considerably larger than that of A.

4.2. Incomplete LU preconditionings

Let us consider the well-known incomplete LU preconditioning. Being carefully implemented, the ILU methods can be quite robust and efficient [16, 17]. It appears that the above GMRES convergence estimate can be related to such preconditionings. If

$$A = LU + E$$

is an ILU decomposition of A and

$$M = AU^{-1}L^{-1}$$

is the right preconditioned matrix, then the size of the error matrix E is well controlled if the ILU decomposition by values is used. One can show that such ILU preconditionings with small drop-tolerance parameter guarantee a good separation of the spectrum from the origin by

$$|\lambda(M)| \geqslant \delta = 1/(1 + \|A^{-1}\|\|E\|) \tag{33}$$

Moreover, the Frobenius distance to the identity $\|I - M\|_{\mathrm{F}}$ is bounded above by

$$\|I - M\|_{\mathrm{F}} = \sqrt{\gamma} \leqslant \|U^{-1}L^{-1}\|\|E\|_{\mathrm{F}}$$

Meanwhile, one gets another evidence of the importance of the ILU stability, i.e. an appropriate boundedness of the quantity $\|U^{-1}L^{-1}\|$. An ILU factorization strategy which directly takes into the account the bounds on $\|U^{-1}\|$ and $\|L^{-1}\|$ was presented in Reference [17].

At the same time, separation of the spectrum apart from the origin, appears to be of less importance due to (a) slight dependence on δ in the iteration number bound, and (b) the existence of guaranteed lower bound on δ according to (33).

5. NUMERICAL EXAMPLE

In this section we are considering the parametrized set of test problems which was actually constructed above in Theorem 3.

We will consider a 3×3 block-diagonal matrix M with diagonal blocks of the size m, $k-m$, and $n-k$ constructed as follows:

$$M_{11} = \delta S_m, \quad M_{22} = I_{k-m} - \rho S_{k-m}, \quad M_{33} = I_{n-k}$$

where S_n is the $n \times n$ cyclic shift permutation matrix, i.e.

$$S_n = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

The eigenvalue spectrum of M is exactly (23). As the independent parameters, we choose n, $\tilde{\gamma}$, and δ, while the remaining quantities ρ, m, and k, were obtained using (18), (24), and (27), respectively. We also calculated the actual values γ and γ_*.

In Table I we present some matrix statistics and iteration results for a set of such problems of the order $n = 2000$ defined by various $\delta = \min|\lambda(M)|$ and $\tilde{\gamma}$:

$\gamma = \|I - M\|_{\mathrm{F}}^2$ (the matrix is almost optimally normalized in the sense of Remark 3);

$\gamma_* = \min_{\sigma \in R^1} \|I - \sigma M\|_{\mathrm{F}}^2$;

$\rho = \min_{\lambda \neq 1} |1 - \lambda(M)|$ is the radius of the eigenvalue cluster around $\lambda = 1$;

m is the number of small eigenvalues such that $|\lambda(M)| = \delta$;

k is the total number of eigenvalues satisfying $\lambda(M) \neq 1$, where M is the coefficient matrix (on the other hand, k is the iteration number lower bound);

k_0 is the observed iteration number for the standard GMRES(∞) method [1], needed for the 10^{-6} reduction of the Euclidean norm of the residual;

$\hat{k}$ is the upper bound for k_0 given by Theorem 2.

Table I. Near 'worst-case' test problems ($n = 2000$).

δ	$\tilde{\gamma}$	γ	γ_*	ρ	m	k	k_0	$\hat{k}$	$\hat{k}/k$
10^{-2}	8.00	8.06	8.06	0.313	6	27	28	138	5.11
10^{-2}	16.00	16.11	16.11	0.313	12	54	55	261	4.83
10^{-2}	32.00	32.33	32.29	0.313	24	109	110	507	4.65
10^{-5}	8.00	8.79	8.79	0.204	7	50	51	209	4.18
10^{-5}	16.00	16.62	16.61	0.204	13	100	101	384	3.84
10^{-5}	32.00	32.28	32.26	0.204	25	200	201	733	3.67
10^{-8}	8.00	8.64	8.64	0.163	7	69	73	257	3.72
10^{-8}	16.00	16.33	16.32	0.163	13	139	143	475	3.42
10^{-8}	32.00	32.69	32.66	0.163	26	279	284	938	3.36
10^{-11}	8.00	8.57	8.57	0.139	7	88	178	301	3.42
10^{-11}	16.00	16.16	16.15	0.139	13	176	347	557	3.16
10^{-11}	32.00	32.33	32.31	0.139	26	353	708	1105	3.13

The right-hand side for the linear systems was defined as $b = [b_1, \ldots, b_n]^{\mathrm{T}}$ with

$$b_j = \begin{cases} 1, & j = 1 \\ \sqrt{m/(k-m)}, & j = m+1 \\ \sqrt{m}/(n-k), & j = k+1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

the initial guess was $x^0 = 0$ and the GMRES stopping criterion was $\|r^k\| \leqslant 10^{-6}\|r^0\|$. Note that the above choice of $b = r^0$ is different from that implied by Lemma A2, since the observed iteration number k_0 was not very sensitive to deviations from the exact 'worst-case' initial residual determined in Lemma A2.

In Table I one can observe a satisfactory consistency between the actual GMRES iteration number and its upper and (especially) lower bounds.

A certain loss of GMRES orthogonality was observed for the problems with $\delta \leqslant 10^{-8}$, which explains the increase of the gap between k_0 and k observed in these cases. (However, with the stopping criterion $\|r^k\| \leqslant \|r^0\|/(2e)$, one can observe $k_0 = k + 1$ even with $\delta = 10^{-11}$.)

The last column of Table I shows the decrease of ratio $\hat{k}/k$ with the increase of δ^{-1} and γ. However, the theoretical 37% gap between the upper and lower bounds (see Remark 5) is attained only for δ far smaller than that usable in numerical tests. Considerably better upper bounds for realistic scale of condition numbers would result from the improvement of the constant 6.924 in (10) as noted in Remark 1.

6. CONCLUSION

In the present paper, an upper bound for GMRES residual norm error is obtained, which:

(a) shows the superlinear convergence of the method;

(b) mainly depends on the Frobenius distance between the identity and the (preconditioned) matrix, only slightly depends on the relative closeness of the matrix spectrum to the origin, and provides an asymptotically correct iteration number bound;
(c) may be useful for the construction and comparative studies of various preconditionings, in particular of the ILU type.

APPENDIX A

Here we present the proofs of auxiliary results used in the paper.

Lemma A1
For any $c>1$ and $0<\delta<1$ one has

$$\max_{t\geqslant\delta}\left(\log\frac{1}{t}-(1-t)^2\log\frac{c}{\delta}\right)\leqslant\max_{0<t<1}\psi(t) \tag{A1}$$

where

$$\psi(t)=-(1-t)^2\log c-t(2-t)\log t$$

In particular, if $c=\mathrm{e}\equiv\exp(1)$, then the estimate

$$\log\frac{1}{t}\leqslant 0.2743+(1-t)^2\log\frac{\mathrm{e}}{\delta} \tag{A2}$$

holds for any $t\geqslant\delta$.

Proof
First we note that the right-hand side of (A1) is always positive and therefore it suffices to consider only the case of $0<t<1$. Using $t\geqslant\delta$, one has

$$\log\frac{1}{t}-(1-t)^2\log\frac{c}{\delta}\equiv(1-t)^2\log\frac{\delta}{t}+\psi(t)\leqslant\psi(t)$$

which readily yields the general estimate by

$$\max_{t\geqslant\delta}\left(\log\frac{1}{t}-(1-t)^2\log\frac{c}{\delta}\right)=\max_{\delta\leqslant t<1}\left(\log\frac{1}{t}-(1-t)^2\log\frac{c}{\delta}\right)\leqslant\max_{\delta\leqslant t<1}\psi(t)\leqslant\max_{0<t<1}\psi(t)$$

Setting $c=\mathrm{e}$ in (A1), one can prove (A2) using the upper bound

$$\max_{0<t<1}\psi(t)=\max_{0<t<1}(-(1-t)^2-t(2-t)\log t)=\psi(0.5468\ldots)\leqslant 0.2743$$

The latter inequality is found numerically using the root $t=0.5468\ldots$ of the equation $f(t)=0$ (obtained from $\psi'(t)=0$), where $f(t)=2\log(c/t)-1-(1-t)^{-1}$ with $c=\mathrm{e}$. Note that $f'(t)=-2t^{-1}-(1-t)^{-2}x$ yields $|f'(t)|>3$ for $0<t<1$ and therefore Newton iterations $t:=t-f(t)/f'(t)$ quickly converge starting, e.g. with $t=1/2$. □

Lemma A2
For any diagonalizable matrix

$$M = V\Lambda V^{-1}$$

with k pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ and any polynomial of the $(k-1)$th degree $P_{k-1}(z)$ normalized by $P_{k-1}(0) = 1$ there exists n-vector r^0 such that

$$\frac{\|P_{k-1}(M)r^0\|}{\|r^0\|} \geqslant \frac{1}{\kappa}\left(\sum_{j=1}^{k}\left|\frac{Q_k(0)}{\lambda_j Q_k'(\lambda_j)}\right|\right)^{-1} \tag{A3}$$

where

$$Q_k(z) = \prod_{j=1}^{k}(z-\lambda_j) \tag{A4}$$

and $\kappa = \|V\|\|V^{-1}\|$.

Proof
Let us denote

$$r^k = P_{k-1}(M)r^0$$

and

$$s^0 = V^{-1}r^0, \quad s^k = V^{-1}r^k$$

First we note that $\|s^k\| \leqslant \|V^{-1}\|\|r^k\|$ and $\|r^0\| = \|Vs^0\| \leqslant \|V\|\|s^0\|$ yield

$$\frac{\|r^k\|}{\|r^0\|} \geqslant \frac{\|V^{-1}\|^{-1}\|s^k\|}{\|V\|\|s^0\|} = \frac{1}{\kappa}\frac{\|s^k\|}{\|s^0\|} \tag{A5}$$

Further, by $P_{k-1}(M) = VP_{k-1}(\Lambda)V^{-1}$, one has

$$s^k = P_{k-1}(\Lambda)s^0$$

and therefore,

$$\|s^k\|^2 \equiv \sum_{j=1}^{n}|s_j^k|^2 = \sum_{j=1}^{n}|s_j^0|^2|P_{k-1}(\lambda_j)|^2$$

Comparing the latter with (A5) yields

$$\frac{\|r^k\|^2}{\|r^0\|^2} \geqslant \frac{1}{\kappa^2}\frac{\sum_{j=1}^{n}|s_j^0|^2|P_{k-1}(\lambda_j)|^2}{\sum_{j=1}^{n}|s_j^0|^2} \tag{A6}$$

On the other hand, any $(k-1)$th degree polynomial P_{k-1} can be represented using interpolation with k pairwise distinct nodes $\lambda_1, \ldots, \lambda_k$:

$$P_{k-1}(z) \equiv \sum_{j=1}^{k} w_j(z)P_{k-1}(\lambda_j)$$

where

$$w_j(z) = \prod_{l \neq j} \frac{z - \lambda_l}{\lambda_j - \lambda_l}$$

Using the condition $P_{k-1}(0) = 1$, one has

$$1 = \left| \sum_{j=1}^{k} w_j(0) P_{k-1}(\lambda_j) \right|^2 \leqslant \left(\sum_{j=1}^{k} |w_j(0)||P_{k-1}(\lambda_j)| \right)^2$$

$$\leqslant \left(\sum_{j=1}^{k} |w_j(0)||P_{k-1}(\lambda_j)|^2 \right) \left(\sum_{j=1}^{k} |w_j(0)| \right) \tag{A7}$$

Therefore, choosing $r^0 = Vs^0$ with

$$|s_j^0| = \begin{cases} \sqrt{|w_j(0)|}, & 1 \leqslant j \leqslant k \\ 0, & k < j \leqslant n \end{cases}$$

one can see that (A7) takes the form

$$\sum_{j=1}^{n} |s_j^0|^2 |P_{k-1}(\lambda_j)|^2 \geqslant \frac{1}{\sum_{j=1}^{n} |s_j^0|^2}$$

and thus (A6) yields

$$\frac{\|r^k\|^2}{\|r^0\|^2} \geqslant \frac{1}{\kappa^2} \frac{\sum_{j=1}^{n} |s_j^0|^2 |P_{k-1}(\lambda_j)|^2}{\sum_{j=1}^{n} |s_j^0|^2} \geqslant \frac{1}{\kappa^2} \left(\sum_{j=1}^{n} |s_j^0|^2 \right)^{-2} = \frac{1}{\kappa^2} \left(\sum_{j=1}^{k} |w_j(0)| \right)^{-2}$$

It only remains to note that by (A4) one has

$$w_j(z) \equiv \frac{Q_k(z)}{(z - \lambda_j) Q_k'(\lambda_j)}$$

and (A3) readily follows. □

Remark A1
The above condition 'any polynomial of the $(k-1)$th degree $P_{k-1}(z)$' can obviously be replaced by 'any polynomial $P(z)$ of the degree smaller than k'.

Lemma A3
For any $1 \leqslant m < k$ and $0 < t < 1$ one has

$$\left(\frac{1 - t^m}{1 - t^k} \right)^{1/(k-m)} \geqslant \frac{1}{2}$$

Proof
We first prove that

$$\frac{1 - t^m}{1 - t^k} \geqslant \frac{m}{k}$$

Indeed, by $0<t<1$ the above is equivalent to the inequality $kt^m - mt^k \leqslant k - m$, the validity of which can readily be checked for $0 \leqslant t \leqslant (k/m)^{1/(k-m)} > 1$. It only remains to show that

$$\left(\frac{m}{k}\right)^{1/(k-m)} \geqslant \frac{1}{2}$$

The latter follows from

$$\left(\frac{k}{m}\right)^{1/(k-m)} = \exp\left(\frac{\log k - \log m}{k-m}\right)$$

$$= \exp\left(\int_0^1 \frac{\mathrm{d}s}{ks + m(1-s)}\right) \leqslant \exp\left(\int_0^1 \frac{\mathrm{d}s}{1+s}\right) = 2 \qquad \square$$

Lemma A4
For any $0<t<0.0409$ it holds

$$\Phi(t) \geqslant \tilde{\Phi}(t)$$

where

$$\Phi(t) = \frac{\log((1-t)/s) + \log((1-s)/t)}{(1+t^2)\log((1-t)/s) + s^2 \log((1-s)/t)}, \quad s = \left(1 + 2\log\frac{1}{t}\right)^{-1/2}$$

and

$$\tilde{\Phi}(t) = \frac{1 + 2\log(1/t)}{1 + \log(1 + 2\log(1/t))}$$

Moreover, the function $\tilde{\Phi}(t)$ decreases on $0<t<\exp(1/2)$.

Proof
The monotonicity of $\tilde{\Phi}(t)$ follows immediately from $\tilde{\Phi}'(t) = -2t^{-1}\log(s^{-2})(1+\log(s^{-2})^{-2}$. In order to prove the inequality, let us note that

$$\tilde{\Phi}(t) \equiv \frac{\frac{1}{2} + \log(1/t)}{\frac{1}{2} + \log(1/s)}$$

and hence it suffices to show that

$$\log\frac{1-t}{s} + \log\frac{1-s}{t} \geqslant \frac{1}{2} + \log\frac{1}{t}$$

and

$$(1+t^2)\log\frac{1-t}{s} + s^2\log\frac{1-s}{t} \leqslant \frac{1}{2} + \log\frac{1}{s}$$

The former inequality is equivalent to $f(t) = 1 - 2\log t - (1+\sqrt{\mathrm{e}}/(1-t))^2 \geqslant 0$, and, by $f'(t)<0$ clearly holds for any $0<t\leqslant 0.0409<t_*$, where $f(t_*)=0$.

The latter inequality is a consequence of the estimate

$$(1+t^2)\log\frac{1-t}{s}+s^2\log\frac{1-s}{t}-\frac{1}{2}-\log\frac{1}{s}\leqslant t^2\log\frac{1}{s}+s^2\log\frac{1}{t}-\frac{1}{2}$$

$$=\frac{t^2}{2+4\log(1/t)}\left(\left(1+\log\frac{1}{t^2}\right)\log\left(1+\log\frac{1}{t^2}\right)-\frac{1}{t^2}\right)\leqslant-\frac{t^2}{6+12\log(1/t)}$$

where the last inequality holds by $(1+x)\log(1+x)-\exp(x)\leqslant x(1+x)-1-x-x^2/2-x^3/6=-1+x^2/2-x^3/6\leqslant-1/3$, where $x=\log(t^{-2})>0$. □

ACKNOWLEDGEMENTS

The author gratefully acknowledges a useful discussion of the draft version of Section 2 with Prof. Owe Axelsson, as well as a careful and instructive review of an anonymous referee which helped to properly focus the content of the paper and to essentially improve its presentation.

REFERENCES

1. Saad Y, Schultz MH. GMRES: a generalized minimal residual algorithm for solving nonsymmetric systems of linear equations. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**:856–869.
2. Walker HF. Implementation of the GMRES method using householder transformations. *SIAM Journal on Scientific and Statistical Computing* 1988; **9**:152–163.
3. Axelsson O. *Iterative Solution Methods*. Cambridge University Press: New York, 1994.
4. Campbell SL, Ipsen IC, Kelley CT, Meyer CD. GMRES and the minimal polynomial. *BIT* 1996; **36**:664–675.
5. Moret I. A note on the superlinear convergence of GMRES. *SIAM Journal on Numerical Analysis* 1997; **34**:513–516.
6. Simoncini V, Szyld DB. On the occurrence of superlinear convergence of exact and inexact Krylov subspace methods. Department of Mathematics, *Temple University Report 03-3-13*, Philadelphia, Pennsylvania, March, 2003; 25. (*SIAM Review* 2005; **47**, to appear.)
7. van der Vorst HA, Vuik C. The superlinear convergence behaviour of GMRES. *Journal of Computational and Applied Mathematics* 1993; **48**:327–341.
8. Grote MJ, Huckle T. Parallel preconditioning with sparse approximate inverses. *SIAM Journal on Scientific Computing* 1997; **18**:838–853.
9. Kaporin IE. Two-level explicit preconditionings for the conjugate gradient method. *Differential Equations* 1992; **28**:280–289.
10. Kaporin IE. New convergence results and preconditioning strategies for the conjugate gradient method. *Numerical Linear Algebra with Applications* 1994; **1**:179–210.
11. Axelsson O, Lindskog G. On the rate of convergence of the preconditioned conjugate gradient method. *Numerische Mathematik* 1986; **48**:499–523.
12. Jennings A. Influence of the eigenvalue spectrum on the convergence rate of the conjugate gradient method. *Journal of the Institute of Mathematics and its Applications* 1977; **20**:61–72.
13. Marcus M, Minc H. *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Inc.: Boston, 1964.
14. Winther R. Some superlinear convergence results for the conjugate gradient method. *SIAM Journal on Numerical Analysis* 1980; **17**:14–17.
15. Kaporin IE. Spectral bound estimates for two-sided explicit preconditionings (Russian). *Vestnik Moskovskogo Universiteta, Ser.15* (*Computational Mathematics and Cybernetics*) 1993; **2**:28–42.
16. de Almeida VF, Chapman AM, Derby JJ. On equilibration and sparse factorization of matrices arising in finite element solutions of partial differential equations. *Numerical Methods in Partial Differential Equations* 2000; **16**:11–29.
17. Bollhöfer M. A robust ILU with pivoting based on monitoring the growth of the inverse factors. *Linear Algebra and Applications* 2001; **338**(1–3):201–218.

