BIT
1998, Vol. 38, No. 3, pp. 527–543

0006-3835/98/3803-0527 $12.00

MINIMUM RESIDUAL METHODS FOR AUGMENTED SYSTEMS *

B. FISCHER[1], A. RAMAGE[2], D. J. SILVESTER[3], and A. J. WATHEN[4] †

[1] *Institute of Mathematics, Medical University of Lübeck, Lübeck, Germany*
email: fischer@math.mu-luebeck.de

[2] *Department of Mathematics, University of Strathclyde*
Glasgow G1 1XH, Scotland. email: A.Ramage@strath.ac.uk

[3] *Department of Mathematics, UMIST, P.O. Box 88*
Manchester M60 1QD, England. djs@lanczos.ma.umist.ac.uk

[4] *Oxford University Computing Laboratory, Oxford OX1 3QD, England*
email: wathen@comlab.ox.ac.uk

Abstract.

For large systems of linear equations, iterative methods provide attractive solution techniques. We describe the applicability and convergence of iterative methods of Krylov subspace type for an important class of symmetric and indefinite matrix problems, namely augmented (or KKT) systems. Specifically, we consider preconditioned minimum residual methods and discuss indefinite versus positive definite preconditioning. For a natural choice of starting vector we prove that when the definite and indefinite preconditioners are related in the obvious way, MINRES (which is applicable in the case of positive definite preconditioning) and full GMRES (which is applicable in the case of indefinite preconditioning) give residual vectors with identical Euclidean norm at each iteration. Moreover, we show that the convergence of both methods is related to a system of normal equations for which the LSQR algorithm can be employed. As a side result, we give a rare example of a non-trivial normal(1) matrix where the corresponding inner product is explicitly known: a conjugate gradient method therefore exists and can be employed in this case.

AMS subject classification: 65F10, 65F50.

Key words: Indefinite systems, preconditioners, minimum residual methods, conjugate gradients.

1 Introduction.

The efficient solution of sparse symmetric indefinite linear systems plays an important role in many different applications of scientific computation. In this paper we consider the iterative solution of a class of such systems, namely those of the form

*Received February 1997. Communicated by Robert S. Schreiber.

†This work was supported by British Council/German Academic Exchange Service Research Collaboration Project 465 and NATO Collaborative Research Grant CRG 960782.

$$\begin{bmatrix} H & \mathcal{B}^T \\ \mathcal{B} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}, \tag{1.1}$$

where the $n \times n$ matrix H is symmetric positive definite and the $m \times n$ matrix $\mathcal{B}$ ($n > m$) has rank $m - r$ ($r \geq 0$). Note that the zero block in the right-hand side vector ensures consistency of the system. By applying the congruence transform

$$\begin{bmatrix} H & \mathcal{B}^T \\ \mathcal{B} & 0 \end{bmatrix} = \begin{bmatrix} H & 0 \\ \mathcal{B} & I \end{bmatrix} \begin{bmatrix} H^{-1} & 0 \\ 0 & -\mathcal{B}H^{-1}\mathcal{B}^T \end{bmatrix} \begin{bmatrix} H & \mathcal{B}^T \\ 0 & I \end{bmatrix}$$

it is clear that the coefficient matrix will have n positive, $m - r$ negative and r zero eigenvalues.

When using iterative methods, preconditioning is always an important issue. Given the structure in (1.1), an obvious choice is to use block diagonal preconditioning. This will avoid coupling of the block equations and is relatively cheap to implement. One question arises immediately: should the preconditioner for an indefinite matrix itself be indefinite? Although it might be argued that a "good" preconditioner for a symmetric indefinite matrix should indeed be indefinite, in order to apply a symmetric iterative method such as MINRES [15], the preconditioner must be symmetric and positive definite. We therefore consider preconditioners of the form

$$\begin{bmatrix} \frac{1}{\eta}H & 0 \\ 0 & \pm M \end{bmatrix}, \quad \eta \in \mathbf{R}^+, \tag{1.2}$$

where M is symmetric positive definite (typically an approximation to $\mathcal{B}H^{-1}\mathcal{B}^T$), and note that the inclusion of $\pm$ offers a choice between a positive definite or indefinite matrix. Such preconditioners will be of interest in practice when H is easy to invert. Examples include linear least squares problems [2], linear KKT systems [4] and mixed finite element approximations to the Laplacian [9, p. 134].

Applying preconditioner (1.2) to (1.1) results in the preconditioned system

$$\begin{bmatrix} \eta I & B^T \\ \pm B & 0 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{f}} \\ \mathbf{0} \end{bmatrix}. \tag{1.3}$$

Here $B = Q^{-1}\mathcal{B}L^{-T}$, $\bar{\mathbf{x}} = \frac{1}{\eta}L^T\mathbf{x}$, $\bar{\mathbf{y}} = Q^T\mathbf{y}$ and $\bar{\mathbf{f}} = L^{-1}\mathbf{f}$, where L and Q are the Cholesky factors of H and M respectively. For convenience, we will use the notation

$$A_\eta^\pm = \begin{bmatrix} \eta I & B^T \\ \pm B & 0 \end{bmatrix}. \tag{1.4}$$

Note that η is often used with direct solution techniques to tune the spectral condition number [2]. With $+B$, the inertia of the coefficient matrix remains unchanged after preconditioning but for $-B$ (indefinite preconditioning) we have that

$$\begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix} \begin{bmatrix} \eta I & B^T \\ -B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \eta\, \mathbf{u}^T\mathbf{u} > 0, \quad \mathbf{u} \neq \mathbf{0},$$

that is, the preconditioned matrix is positive semi-definite (at the expense of destroying symmetry). In this paper we will analyse the effect of the choice of η and $\pm$ on the convergence of iterative methods.

Some of the most effective iterative methods available are those of Krylov subspace type, which have in-built minimisation properties whilst being cheap to implement and parameter free. Here we consider minimal residual methods, whose iterates minimise the Euclidean norm of the residual $\mathbf{r}_k$ at each step. For the symmetric indefinite matrix A_η^+, our method of choice is MINRES [8, 15], while for the nonsymmetric positive semi-definite case A_η^- we use GMRES [17]. Note that GMRES compares unfavourably with MINRES in terms of operation count and storage requirements. For the systems under consideration, the quasi-minimum residual method QMR [10] is precisely as expensive as MINRES. Here, however, we restrict our attention to the GMRES approach because it is easier to analyse and note that, in our experience, GMRES and QMR produce comparable results in terms of the number of iterations required to achieve a prescribed reduction in the norm of the residual. The QMR method cannot converge in this sense in fewer iterations than GMRES.

We note also that for problems of the form (1.3) there are connections with the LSQR method of Paige and Saunders [16]. The solution of (1.3) in the plus case is equivalent to the least squares problem $\min \|B^T\bar{\mathbf{y}} - \bar{\mathbf{f}}\|$ for which the LSQR method may be readily applied (see also [18]).

In Section 2, we describe some properties of the coefficient matrices $A_\eta^\pm$ which affect the behaviour of the iterative methods. These are used in the proofs of the main results in Section 3: for a natural choice of starting vector, MINRES applied to the symmetric indefinite matrix A_η^+ and GMRES applied to the nonsymmetric positive semi-definite matrix A_η^- are equivalent, in the sense that they generate the same iterates, so that

$$\left\|\mathbf{r}_k^{\text{MINRES}}\right\|_2 = \left\|\mathbf{r}_k^{\text{GMRES}}\right\|_2$$

at each step. This is true for any choice of η. The implications of this result are discussed in Section 4.

2 Matrix properties.

Standard convergence estimates for MINRES and GMRES depend on the eigenvalues (and eigenvectors in the nonsymmetric case) of the coefficient matrix (see for example [12, 14]). We therefore state here the eigenvalues of $A_\eta^\pm$, which are straightforward to calculate. For the + case with $\eta = 1$ and B of full rank, these are contained in [2]. We include the proof of the more general case here as it is instructive for later arguments.

LEMMA 2.1. *If σ_k are the non-zero singular values of the preconditioned matrix B, then the $n + m$ eigenvalues of A_η^+ are*

(i) *0 with multiplicity r;*

(ii) *η with multiplicity $n - m + r$;*

(iii) $\frac{1}{2}\left(\eta \pm \sqrt{4\sigma_k^2 + \eta^2}\right)$ *for* $k = 1, \ldots, m - r$.

PROOF. Any real eigenvalue λ of A_η^+ (and corresponding right eigenvector $[\mathbf{u} \;\; \mathbf{v}]^T$) must satisfy the equations

$$\eta\mathbf{u} + B^T\mathbf{v} = \lambda\mathbf{u}, \tag{2.1a}$$

$$B\mathbf{u} = \lambda\mathbf{v}. \tag{2.1b}$$

Three cases now arise:

(i) $\lambda = 0$: (2.1a) becomes $\mathbf{u} = -\frac{1}{\eta}B^T\mathbf{v}$. Now, premultiplying (2.1b) by $\mathbf{v}^T$ implies $-\frac{1}{\eta}\mathbf{v}^T BB^T\mathbf{v} = \mathbf{0}$, whence $B^T\mathbf{v} = \mathbf{0}$ and $\mathbf{u} = \mathbf{0}$. There are r non-trivial vectors which satisfy $B^T\mathbf{v} = \mathbf{0}$ so there are r zero eigenvalues with right eigenvectors of the form

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{v}_i \end{bmatrix}.$$

(ii) $\lambda = \eta$: equations (2.1) here give $\mathbf{u}^T B^T B\mathbf{u} = \eta\mathbf{u}^T B^T\mathbf{v} = 0$ hence $B\mathbf{u} = \mathbf{v} = \mathbf{0}$. This is satisfied by $n - (m - r)$ independent vectors $\mathbf{u}$, thus there are $n - m + r$ independent right eigenvectors of the form

$$\begin{bmatrix} \mathbf{u}_i \\ \mathbf{0} \end{bmatrix} \tag{2.2}$$

where $B\mathbf{u}_i = \mathbf{0}$. These can be taken to be orthonormal.

(iii) $\lambda \neq \eta$: here (2.1a) implies $\mathbf{u} = \frac{1}{\lambda - \eta}B^T\mathbf{v}$ so that, from (2.1b), we have $BB^T\mathbf{v} = \lambda(\lambda - \eta)\mathbf{v}$. Since BB^T has eigenvalues equal to the squares of the singular values of B, we have that all real roots of the quadratic equation

$$\lambda(\lambda - \eta) = \sigma_k^2 \quad \text{for } k = 1, \ldots, m - r$$

give real eigenvalues of A_η^+. These are the $2(m - r)$ eigenvalues in part (iii) of the lemma. □

LEMMA 2.2. *If the non-zero singular values, σ_k, of the preconditioned matrix B are such that*

$$0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_t \leq \frac{\eta}{2} < \sigma_{t+1} \leq \cdots \leq \sigma_{m-r}$$

then the $n + m$ eigenvalues of A_η^- are

(i) *0 with multiplicity r;*

(ii) *η with multiplicity $n - m + r$;*

(iii) $\frac{1}{2}\left(\eta \pm \sqrt{\eta^2 - 4\sigma_k^2}\right)$ *for* $k = 1, \ldots, t$;

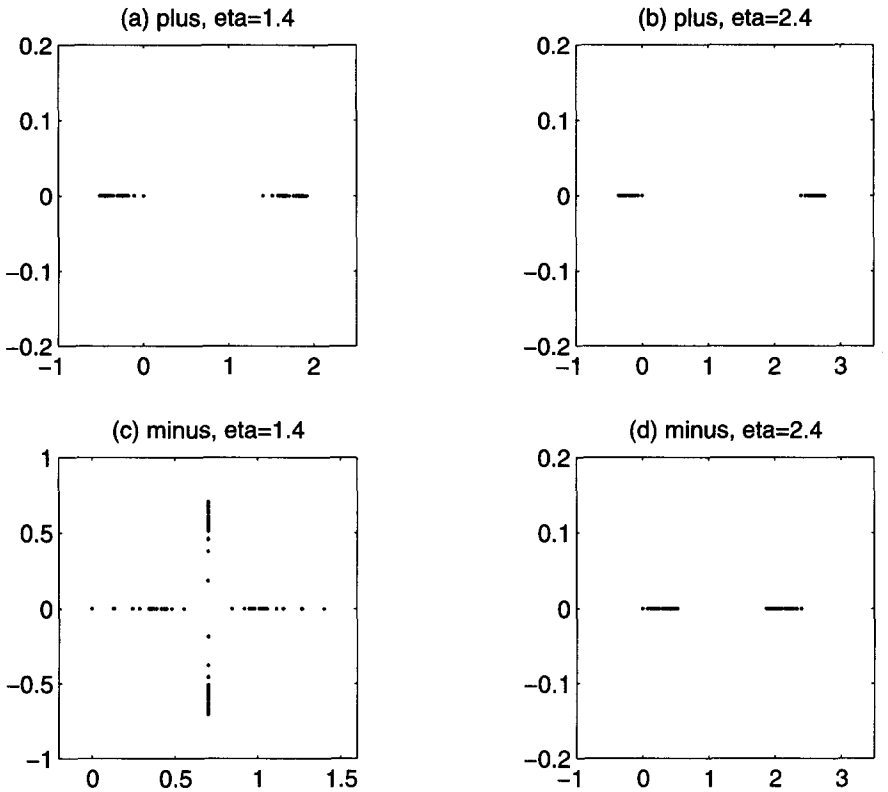

Figure 2.1: Eigenvalue distribution of $A_\eta^\pm$.

(iv) $\frac{1}{2}\left(\eta \pm i\sqrt{4\sigma_k^2 - \eta^2}\right)$ *for* $k = t+1, \ldots, m-r$.

PROOF. Similar to that of the previous lemma. □

It is clear that the parameter η plays an important role in the spectra for both the + and − cases. As an example, we consider a matrix $A_\eta^\pm$ with $n = 578$, $m = 81$, $r = 1$ and sub-block B with singular values which are roughly evenly spread between $\sigma_1 = 0.4068$ and $\sigma_{m-1} = 0.9931$. The source of this matrix will be explained in more detail later. Plots of the eigenvalues with two different values of η are given in Figure 2.1. The eigenvalues of A_η^+ always lie in two intervals symmetric about the point $\frac{\eta}{2}$ and changing η simply scales this picture. For A_η^-, the effect of changing η is more dramatic. For $\eta < 2\sigma_1$, all of the eigenvalues would lie on a vertical line-segment in the complex plane (except for the multiple eigenvalues at 0 and η). For $\eta > 2\sigma_{m-1}$ (Figure 2.1d), all eigenvalues of A_η^- are real and we again have two distinct intervals symmetric about $\frac{\eta}{2}$. For intermediate values of η, the eigenvalues will lie on a cross in the complex plane

as in Figure 2.1c. Note that although η is traditionally used to tune the spectral condition number for direct solution methods, this quantity has little bearing on iterative convergence as we will show. A further interesting point is that as all of the eigenvalues of A_η^- are real for $\eta > 2\sigma_{m-1}$, there exists an inner product with respect to which A_η^- is normal(1) [7]. That is, even though A_η^- is nonsymmetric in the usual Euclidean inner product, there exists an inner product in which A_η^- is self-adjoint and positive definite: namely, the inner product defined by

$$\left\langle \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix}, \begin{bmatrix} \mathbf{v} \\ \mathbf{q} \end{bmatrix} \right\rangle = \begin{bmatrix} \mathbf{u}^T & \mathbf{p}^T \end{bmatrix} \begin{bmatrix} I & \frac{2}{\eta}B^T \\ \frac{2}{\eta}B & I \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{q} \end{bmatrix}. \tag{2.3}$$

Hence the standard conjugate gradient method with inner product defined by (2.3) could be applied directly to A_η^-. This is a rare example of a non-trivial normal(1) matrix where the corresponding inner product is explicitly known. Implementation of the conjugate gradient method would, however, involve overheads in terms of the computation of (2.3). For $\eta < 2\sigma_{m-1}$, (2.3) does not define an inner product. Moreover, the eigenvalues of $\mathcal{A}_\eta^-$ do not lie on a single line in the complex plane and so no conjugate gradient method exists.

One common feature of the spectra of $A_\eta^\pm$ is the repeated eigenvalues at 0 and η. It is useful, therefore, to consider their possible effect on iterative convergence. The eigenvectors corresponding to the multiple eigenvalue at η are known to be of the form (2.2), so in practice it is easy to remove their influence: any initial residual for MINRES or GMRES of the form

$$\mathbf{r}_0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{w} \end{bmatrix} \tag{2.4}$$

for some m-vector $\mathbf{w}$ will have no component in the direction of this eigenspace. A natural choice of starting vector for either method is therefore

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} \frac{1}{\eta}\bar{\mathbf{f}} \\ \mathbf{0} \end{bmatrix}, \tag{2.5}$$

where $\bar{\mathbf{f}} = L^{-1}\mathbf{f}$: this will provide an initial residual of the form (2.4) exactly as required, with $\mathbf{w} = \mp\frac{1}{\eta}B\bar{\mathbf{f}}$ for $A_\eta^\pm$. Note that a vector $\hat{\mathbf{x}}_0$ of this form has no component in the nullspace of $A_\eta^\pm$.

3 Convergence of minimum residual methods.

For a general linear system with real coefficient matrix A and right-hand side vector $\mathbf{b}$, a polynomial iterative method of minimum residual type constructs a sequence of iterates $\mathbf{x}_k$ with associated residuals $\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k$. At each step, the residual satisfies the relations

$$\mathbf{r}_k = p_k(A)\mathbf{r}_0, \quad p_k(0) = 1, \quad p_k \in \Pi_k, \tag{3.1a}$$

$$\|\mathbf{r}_k\|_2 = \min\{\|p(A)\mathbf{r}_0\|_2; p \in \Pi_k, p(0) = 1\}, \tag{3.1b}$$

where Π_k is the set of polynomials of degree k and minimisation is over the Krylov subspace

$$K_{k+1}(A, \mathbf{r}_0) = \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, \ldots, A^k\mathbf{r}_0\}.$$

We define the inner product

$$\langle p, q \rangle = \mathbf{r}_0^T p(A)^T q(A)\mathbf{r}_0 \tag{3.2}$$

with associated orthonormal polynomials ψ_j, that is,

$$\langle \psi_j, \psi_k \rangle = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k \end{cases},$$

and note that

$$\|\mathbf{r}_k\|_2^2 = \langle p_k, p_k \rangle.$$

The vectors

$$\mathbf{v}_j = \psi_{j-1}(A)\mathbf{r}_0, \quad j = 1, 2, \ldots, k,$$

constitute a basis for the Krylov space

$$K_k(A, \mathbf{r}_0) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}, \quad \|\mathbf{v}_j\|_2 = 1,$$

which is orthonormal with respect to the Euclidean inner product. Here we have assumed that k does not exceed the grade of $\mathbf{r}_0$ with respect to A.

It is well-known (see for example [20]) that the solution p_k of (3.1) can be expressed in terms of scaled kernel polynomials as

$$p_k(t) = \frac{\sum_{j=0}^{k} \psi_j(0)\psi_j(t)}{\sum_{j=0}^{k} \psi_j(0)^2}. \tag{3.3}$$

We now consider the computation of the orthonormal polynomials ψ_j, each of which satisfies the equation

$$h_{j+1,j}\psi_j(t) = t\psi_{j-1}(t) - \sum_{i=1}^{j} h_{i,j}\psi_{i-1}(t), \quad h_{j+1,j} > 0, \tag{3.4}$$

where an explicit expression for the coefficients is given by

$$h_{i,j} = \langle t\psi_{j-1}, \psi_{i-1} \rangle = \mathbf{v}_i^T A\mathbf{v}_j, \quad i = 1, 2, \ldots, j+1. \tag{3.5}$$

Notice that $h_{j+1,j}$ depends on ψ_j. To develop a procedure which alternates recursively between the recurrence relation (3.4) and the formulae for the coefficients (3.5), we observe that

$$1 = \langle \psi_j, \psi_j \rangle = \frac{\langle \hat{\psi}_j, \hat{\psi}_j \rangle}{h_{j+1,j}^2} = \frac{\|\hat{\mathbf{v}}_{j+1}\|_2^2}{h_{j+1,j}^2}$$

and hence

$$h_{j+1,j} = \|\hat{\mathbf{v}}_{j+1}\|_2, \quad \mathbf{v}_{j+1} = \frac{\hat{\mathbf{v}}_{j+1}}{h_{j+1,j}}, \tag{3.6}$$

where

$$\hat{\mathbf{v}}_{j+1} = \hat{\psi}_j(A)\mathbf{r}_0 = A\psi_{j-1}(A)\mathbf{r}_0 - \sum_{i=1}^{j} h_{i,j}\psi_{i-1}(A)\mathbf{r}_0 = A\mathbf{v}_j - \sum_{i=1}^{j} h_{i,j}\mathbf{v}_i. \tag{3.7}$$

Now (3.5), together with (3.6) and (3.7), is simply Arnoldi's method [1] with starting vector

$$\mathbf{v}_1 = \psi_0(A)\mathbf{r}_0 = \frac{\mathbf{r}_0}{\|\mathbf{r}_0\|_2}. \tag{3.8}$$

Note that there is a fundamental difference here between a symmetric and a nonsymmetric matrix A. If A is symmetric, we have

$$h_{i,j} = \mathbf{r}_0^T\psi_{j-1}(A)^T A\psi_{i-1}(A)\mathbf{r}_0 = \langle \psi_{j-1}, t\psi_{i-1}\rangle = 0 \quad \text{for } i < j-1,$$

that is, the polynomials fulfill a three-term recurrence. The corresponding basis vectors $\mathbf{v}_j$, or more precisely the coefficients (3.5), may be generated by the Lanczos method [13].

Let us now return to the preconditioned matrix (1.4) and describe the notation used in the remainder of the section. For clarity, the superscript $\pm$ has in general been omitted and is reinstated only when differences between the symmetric and nonsymmetric cases are to be emphasised. We therefore define the inner product $\langle \cdot, \cdot \rangle_\eta$ in accordance with (3.2) by

$$\langle p, q\rangle_\eta = \mathbf{r}_0^T p(A_\eta)^T q(A_\eta)\mathbf{r}_0 \tag{3.9}$$

and retain the notation ψ_j for the associated orthonormal polynomials (although these will be different polynomials in the $+$ and $-$ cases). Further, we note that Arnoldi's method could be used to generate the coefficients (3.5) in both the symmetric and nonsymmetric cases, although for A_η^+ most of the Arnoldi coefficients would be zero by definition, resulting in a three-term recurrence (the recurrence automatically constructed by the Lanczos algorithm). In the light of this observation, we consider Arnoldi's method only and highlight the symmetric case simplifications where relevant.

The following useful lemma is easy to verify.

LEMMA 3.1.

$$\begin{bmatrix} \mathbf{u}^T & \mathbf{0} \end{bmatrix} \mathcal{A}_\eta \begin{bmatrix} \mathbf{v} \\ \mathbf{0} \end{bmatrix} = \eta \mathbf{u}^T\mathbf{v}, \qquad \begin{bmatrix} \mathbf{0} & \mathbf{w}^T \end{bmatrix} \mathcal{A}_\eta \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} = 0.$$

THEOREM 3.2. *If*

$$\mathbf{r}_0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{w} \end{bmatrix},$$

then the Arnoldi method applied to A_η and $\mathbf{r}_0$ generates the coefficients

$$\begin{aligned} h_{1,2j-1} = h_{3,2j-1} = \cdots = h_{2j-1,2j-1} &= 0, \\ h_{2,2j} = h_{4,2j} = \cdots = h_{2j-2,2j} &= 0, \quad j = 1, 2, \ldots, \\ h_{2j,2j} &= \eta. \end{aligned}$$

PROOF. The proof is by induction and repeated application of Lemma 3.1. We will show that (cf. (3.7))

$$\hat{\mathbf{v}}_{2m+1} = \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{w}}_{2m+1} \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{v}}_{2m+2} = \begin{bmatrix} \hat{\mathbf{w}}_{2m+2} \\ \mathbf{0} \end{bmatrix},$$

which implies

$$\mathbf{v}_{2m+1} = \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_{2m+1} \end{bmatrix}, \qquad \|\mathbf{w}_{2m+1}\|_2 = 1$$

and

$$\mathbf{v}_{2m+2} = \begin{bmatrix} \mathbf{w}_{2m+2} \\ \mathbf{0} \end{bmatrix}, \qquad \|\mathbf{w}_{2m+2}\|_2 = 1.$$

From (3.5), (3.7) and (3.8) we have the relations

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{r}_0/\|\mathbf{r}_0\|_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_1 \end{bmatrix}, \\ h_{1,1} &= \begin{bmatrix} \mathbf{0} & \mathbf{w}_1^T \end{bmatrix} A_\eta \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_1 \end{bmatrix} = 0, \\ \hat{\mathbf{v}}_2 &= A_\eta \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_1 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{w}}_2 \\ \mathbf{0} \end{bmatrix}, \\ h_{2,2} &= \begin{bmatrix} \mathbf{w}_2^T & \mathbf{0} \end{bmatrix} A_\eta \begin{bmatrix} \mathbf{w}_2 \\ \mathbf{0} \end{bmatrix} = \eta \mathbf{w}_2^T \mathbf{w}_2 = \eta, \end{aligned}$$

so the theorem holds for $j = 1$. Now assume it also holds for $j \leq m$. Then we have from (3.7)

$$\begin{aligned} \hat{\mathbf{v}}_{2m+1} &= A_\eta \mathbf{v}_{2m} - \sum_{i=1}^{2m} h_{i,2m} \mathbf{v}_i \\ &= \begin{bmatrix} \eta \mathbf{w}_{2m} \\ \pm B \mathbf{w}_{2m} \end{bmatrix} - \eta \begin{bmatrix} \mathbf{w}_{2m} \\ \mathbf{0} \end{bmatrix} - \sum_{i=1}^{m} h_{2i-1,2m} \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_{2i-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{w}}_{2m+1} \end{bmatrix}, \end{aligned}$$

and, for $k = 1, \ldots, m+1$,

$$\begin{aligned} h_{2k-1,2m+1} &= \mathbf{v}_{2k-1}^T A_\eta \mathbf{v}_{2m+1} \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{w}_{2k-1}^T \end{bmatrix} A_\eta \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_{2m+1} \end{bmatrix} = 0. \end{aligned}$$

Finally, we have

$$\begin{aligned}\hat{\mathbf{v}}_{2m+2} &= A_\eta \mathbf{v}_{2m+1} - \sum_{i=1}^{2m+1} h_{i,2m+1}\mathbf{v}_i \\ &= \begin{bmatrix} B^T\mathbf{w}_{2m+1} \\ \mathbf{0} \end{bmatrix} - \sum_{i=1}^{m} h_{2i,2m+1}\begin{bmatrix} \mathbf{w}_{2i} \\ \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{w}}_{2m+2} \\ \mathbf{0} \end{bmatrix},\end{aligned}$$

and, for $k = 1, \ldots, m+1$,

$$\begin{aligned}h_{2k,2m+2} &= \mathbf{v}_{2k}^T A_\eta \mathbf{v}_{2m+2} \\ &= \begin{bmatrix} \mathbf{w}_{2k}^T & \mathbf{0} \end{bmatrix} A_\eta \begin{bmatrix} \mathbf{w}_{2m+2} \\ \mathbf{0} \end{bmatrix} \\ &= \eta \mathbf{w}_{2k}^T \mathbf{w}_{2m+2} \\ &= \eta \delta_{k,m+1}.\end{aligned}$$

Hence, by induction, the proposition is true. □

COROLLARY 3.3. *If*

$$\mathbf{r}_0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{w} \end{bmatrix},$$

then the orthonormal polynomials with respect to $\langle \cdot, \cdot \rangle_\eta$ *satisfy*

$$\psi_{2j-1}(t) = t q_{j-1}(t(t-\eta)), \qquad \psi_{2j}(t) = s_j(t(t-\eta)), \qquad j = 1, 2, \ldots,$$

where q_{j-1} *and* s_j *are polynomials of exact degree* $j-1$ *and* j *respectively.*

PROOF. The proof is again by induction. Using (3.4) we have

$$\psi_1(t) = \frac{t\psi_0(t) - h_{1,1}\psi_0(t)}{h_{2,1}} = \frac{t\psi_0(t)}{h_{2,1}} = t q_0(t(t-\eta))$$

and

$$\begin{aligned}\psi_2(t) &= \frac{t\psi_1(t) - h_{1,2}\psi_0(t) - h_{2,2}\psi_1(t)}{h_{3,2}} \\ &= \frac{t(t-\eta)q_0(t(t-\eta)) - h_{1,2}\psi_0(t)}{h_{3,2}} \\ &= s_1(t(t-\eta)),\end{aligned}$$

hence result holds for $j = 1$. Now assume it also holds for $j \le m$. Using (3.4) again gives

$$\begin{aligned}h_{2m+1,2m}\psi_{2m}(t) &= (t-\eta)\psi_{2m-1}(t) - \sum_{i=1}^{2m-1} h_{i,2m}\psi_{i-1}(t) \\ &= (t-\eta)\psi_{2m-1}(t) - \sum_{i=1}^{m} h_{2i-1,2m}\psi_{2i-2}(t)\end{aligned}$$

and

$$\begin{aligned} h_{2m,2m-1}\psi_{2m-1}(t) &= t\psi_{2m-2}(t) - \sum_{i=1}^{2m-1} h_{i,2m-1}\psi_{i-1}(t) \\ &= t\psi_{2m-2}(t) - \sum_{i=1}^{m-1} h_{2i,2m-1}\psi_{2i-1}(t). \end{aligned}$$

Hence, by induction, the proposition is true. □

We now state an important theorem concerning the convergence behaviour of MINRES and GMRES. A result relating the iterates themselves follows.

THEOREM 3.4. *Let*

$$\mathbf{r}_0^{\pm} = \begin{bmatrix} \mathbf{0} \\ \pm\mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mp\frac{1}{\eta}B\bar{\mathbf{f}} \end{bmatrix},$$

and consider the polynomial iteration method (3.1) *with*

$$\mathbf{r}_k^{\pm} = p_k^{\pm}(A_\eta^{\pm})\mathbf{r}_0^{\pm}.$$

Then

$$\|\mathbf{r}_{2k+1}^{+}\|_2 = \|\mathbf{r}_{2k}^{+}\|_2 = \|\mathbf{r}_{2k+1}^{-}\|_2 = \|\mathbf{r}_{2k}^{-}\|_2, \qquad k = 0, 1, \ldots,$$

for any value of η.

PROOF. We apply Corollary 3.3 and (3.3) to obtain

$$p_{2k}^{\pm}(t) = p_{2k+1}^{\pm}(t) = \frac{\sum_{j=0}^{2k}\psi_j^{\pm}(0)\psi_j^{\pm}(t)}{\sum_{j=0}^{2k}\psi_j^{\pm}(0)^2} = \frac{\sum_{j=0}^{k}s_j^{\pm}(0)s_j^{\pm}(t(t-\eta))}{\sum_{j=0}^{k}s_j^{\pm}(0)^2} = \hat{p}_k^{\pm}(t(t-\eta)).$$

Using (3.1) therefore gives

$$\begin{aligned} \|\mathbf{r}_{2k+1}^{\pm}\|_2 &= \|\mathbf{r}_{2k}^{\pm}\|_2 \\ &= \left\|\hat{p}_k^{\pm}\left(A_\eta^{\pm}(A_\eta^{\pm} - \eta I)\right)\mathbf{r}_0^{\pm}\right\|_2 \\ &= \min_{p\in\Pi_k,\ p(0)=1}\left\|p\left(A_\eta^{\pm}(A_\eta^{\pm} - \eta I)\right)\mathbf{r}_0^{\pm}\right\|_2 \\ &= \min_{p\in\Pi_k,\ p(0)=1}\left\|p\left(\begin{bmatrix} \pm B^TB & 0 \\ 0 & \pm BB^T \end{bmatrix}\right)\begin{bmatrix} \mathbf{0} \\ \pm\mathbf{w} \end{bmatrix}\right\|_2 \\ &= \min_{p\in\Pi_k,\ p(0)=1}\left\|p\left(\begin{bmatrix} B^TB & 0 \\ 0 & BB^T \end{bmatrix}\right)\begin{bmatrix} \mathbf{0} \\ \mathbf{w} \end{bmatrix}\right\|_2 \\ &= \frac{1}{\eta}\min_{p\in\Pi_k,\ p(0)=1}\left\|p\left(\begin{bmatrix} B^TB & 0 \\ 0 & BB^T \end{bmatrix}\right)\begin{bmatrix} \mathbf{0} \\ B\bar{\mathbf{f}} \end{bmatrix}\right\|_2. \end{aligned}$$

Hence, as the minimisation is independent of $\pm$ and of η, we have the required result. □

COROLLARY 3.5. *Let*

$$\mathbf{r}_0^{\pm} = \begin{bmatrix} \mathbf{0} \\ \pm\mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mp\frac{1}{\eta}B\bar{\mathbf{f}} \end{bmatrix},$$

and consider the polynomial iteration method (3.1) *as before. Then*

$$\mathbf{x}_{2k+1}^{+} = \mathbf{x}_{2k}^{+} = \mathbf{x}_{2k+1}^{-} = \mathbf{x}_{2k}^{-}, \qquad k = 0, 1, \ldots,$$

for any value of η.

PROOF. In the proof of Theorem 3.4 we have seen that

$$p_{2k}^{\pm}(t) = p_{2k+1}^{\pm}(t) = \hat{p}_k^{\pm}(t(t-\eta))$$

where $\hat{p}_k^{+}, \hat{p}_k^{-}$ are polynomials of degree k which do not depend on η and are related via $\hat{p}_k^{-}(t) = \hat{p}_k^{+}(-t)$. Furthermore, the constraint $\hat{p}_k^{\pm}(0) = 1$ implies that there exist polynomials $q_{k-1}^{\pm}$ of degree $k-1$ with

$$\hat{p}_k^{\pm}(t) = 1 - tq_{k-1}^{\pm}(t).$$

We now introduce the notation

$$A_\eta^{\pm}(A_\eta^{\pm} - \eta I) = N^{\pm} = \begin{bmatrix} \pm B^T B & 0 \\ 0 & \pm BB^T \end{bmatrix}, \qquad C_\eta^{\pm} = \begin{bmatrix} 0 & B^T \\ \pm B & -\eta I \end{bmatrix},$$

so that

$$\mathbf{r}_{2k}^{\pm} = \mathbf{r}_{2k+1}^{\pm} = \hat{p}_k^{\pm}(N^{\pm})\mathbf{r}_0^{\pm}$$

and

$$N^{\pm} = A_\eta^{\pm} C_\eta^{\pm}.$$

In the plus case, we now have

$$\begin{aligned} \mathbf{r}_{2k}^{+} = \hat{p}_k^{+}(N^{+})\mathbf{r}_0^{+} &= (1 - N^{+}q_{k-1}^{+}(N^{+}))\mathbf{r}_0^{+} \\ &= \mathbf{b} - A_\eta^{+}\hat{\mathbf{x}}_0 - N^{+}q_{k-1}^{+}(N^{+})\mathbf{r}_0^{+} \\ &= \mathbf{b} - A_\eta^{+}(\hat{\mathbf{x}}_0 + C_\eta^{+}q_{k-1}^{+}(N^{+})\mathbf{r}_0^{+}) \end{aligned}$$

(where $\mathbf{b}$ is the right-hand side vector of (1.3)) so the corresponding iterate is given by

$$\begin{aligned} \mathbf{x}_{2k}^{+} &= \hat{\mathbf{x}}_0 + C_\eta^{+}q_{k-1}^{+}(N^{+})\mathbf{r}_0^{+} \\ &= \begin{bmatrix} \frac{1}{\eta}\bar{\mathbf{f}} - \frac{1}{\eta}B^T q_{k-1}(BB^T)B\bar{\mathbf{f}} \\ q_{k-1}(BB^T)B\bar{\mathbf{f}} \end{bmatrix}. \end{aligned}$$

Similarly, in the minus case,

$$\begin{aligned} \mathbf{r}_{2k}^{-} = \hat{p}_k^{-}(N^{-})\mathbf{r}_0^{-} &= \hat{p}_k^{+}(N^{+})\mathbf{r}_0^{-} \\ &= (1 - N^{+}q_{k-1}^{+}(N^{+}))\mathbf{r}_0^{-} \\ &= (1 + N^{-}q_{k-1}^{+}(N^{+}))\mathbf{r}_0^{-} \\ &= \mathbf{b} - A_\eta^{-}(\hat{\mathbf{x}}_0 - C_\eta^{-}q_{k-1}^{+}(N^{+})\mathbf{r}_0^{-}) \end{aligned}$$

so the corresponding iterate is given by

$$\begin{aligned} \mathbf{x}_{2k}^{-} &= \hat{\mathbf{x}}_0 - C_\eta^- q_{k-1}^+(N^+)\mathbf{r}_0^- \\ &= \begin{bmatrix} \frac{1}{\eta}\bar{\mathbf{f}} - \frac{1}{\eta}B^T q_{k-1}(BB^T)B\bar{\mathbf{f}} \\ q_{k-1}(BB^T)B\bar{\mathbf{f}} \end{bmatrix}. \end{aligned}$$

As Theorem 3.4 directly implies equality of the odd and even iterates in both cases, we now have

$$\mathbf{x}_{2k}^{+} = \mathbf{x}_{2k+1}^{+} = \mathbf{x}_{2k}^{-} = \mathbf{x}_{2k+1}^{-}$$

which is the required result. □

As well as demonstrating the equivalence of MINRES and GMRES in terms of residual norms, Theorem 3.4 shows that both methods make no progress on every second step. An alternative proof of this result for the case $\eta = 0$ is given in [8]. It is readily seen that the proofs presented here rely on the choice of starting vector (2.5). Only for such a starting vector is the odd/even symmetry required in the proofs exactly preserved. However, in practical computation when B is ill-conditioned and roundoff is significant, this is unlikely to be a good starting vector: see [16, Sections 2.2 and 7.4]. We show numerical results below with both the starting vector (2.5) and the numerically safer starting vector $\hat{\mathbf{x}}_0 = \mathbf{0}$: the results show very similar convergence although in the latter case there are no exact plateaux (as described in Corollary 3.5). This is to be expected. We would recommend the use of a starting vector like $\hat{\mathbf{x}}_0 = \mathbf{0}$ in practical computations especially when B is ill-conditioned; using a similar technique to the above one can show that the equivalence of the Euclidean norms of MINRES and GMRES still holds, but in this case the minimisation does depend on η and we do not have no progress on every second step.

To end this section, we discuss the results of some numerical experiments. The linear systems involved come from mixed finite element approximations to a Stokes flow problem with preconditioning of the form (1.2) [19]. We stress that this preconditioning is not a practical option for real-life Stokes applications: these problems were chosen as the matrices were readily available from a Matlab test set. Note that the matrix B always has rank $m - 1$.

With starting vector (2.5), the behaviour of MINRES with A_η^+ and GMRES with A_η^- was compared for a wide range of values of n and m, using a convergence criterion of

$$\frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_0\|_2} \leq 10^{-6}.$$

As expected, in each case the residuals satisfied

$$\|\mathbf{r}_k^{\text{MINRES}}\|_2 = \|\mathbf{r}_k^{\text{GMRES}}\|_2$$

at every step, with both methods as a result always converging in exactly the same number of iterations, independent of the choice of η. This is due to the cancellation of the $\frac{1}{\eta}$ factor in the convergence test. In addition, the methods made no progress at every odd step.

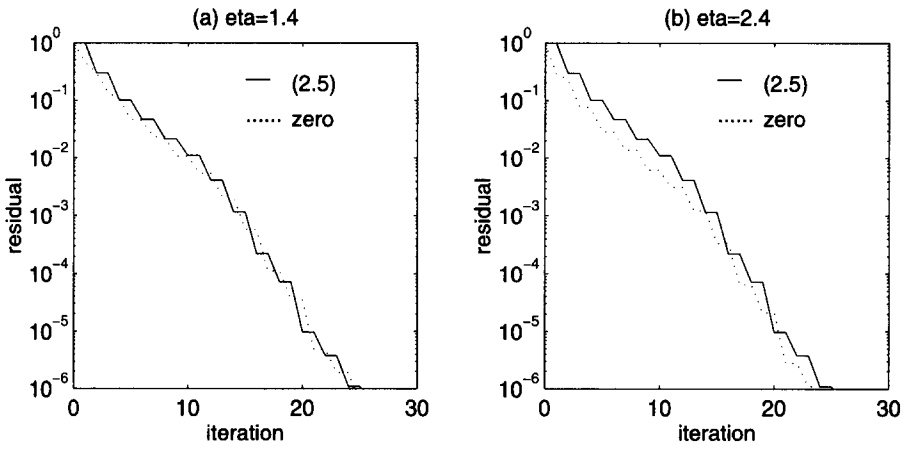

Figure 3.1: Behaviour of MINRES/GMRES for different starting vectors and two values of η (cf. Fig. 2.1).

Two plots of $\log(\|\mathbf{r}_k\|_2/\|\mathbf{r}_0\|_2)$ obtained using MINRES for the case $n = 578$ and $m = 81$ are shown in Figure 3.1 (the analogous plots for GMRES would be identical). This is precisely the matrix whose spectrum is displayed in Figure 2.1 and results are shown for the same two values of η. The solid line corresponds to a starting $\hat{\mathbf{x}}_0$ of the form (2.5): the dotted line comes from the $\hat{\mathbf{x}}_0 = 0$ case. The results in the former case are independent of η while those in the latter are not (the η dependence does not simply cancel). In addition, the exact plateaux predicted by Theorem 3.4 for starting vector (2.5) are clearly visible. We also observe here that for a random starting vector, although the exact equivalence of the residual norms of MINRES and GMRES is lost, convergence curves for the two methods obtained experimentally are almost indistinguishable.

The above results are discussed further in the next section. Here we simply note two relevant points. Firstly, Theorem 3.4 implies that for A_η^-, convergence of GMRES is independent of the associated eigenvector matrix. The standard GMRES convergence estimate [14, equation (3.8)] could therefore be very pessimistic in this case. Secondly, using MINRES with the indefinite system A_η^+ would be preferable to applying QMR to the nonsymmetric system A_η^-: MINRES minimises the residual, whereas QMR does not. Using this type of indefinite preconditioner to gain positivity of the real parts of the eigenvalues at the ex-

pense of symmetry (see for example [11]) is therefore not a sensible thing to do. Note that for the positive case, right preconditioning could be applied: this would change the above pictures as the method is minimising in a different norm. We have not explored this possibility.

4 Relation to the normal equations.

The results of the previous section have a number of important implications for the efficient iterative solution of (1.1). From the proof of Theorem 3.4, it is clear that applying MINRES to a system with coefficient matrix

$$A_\eta^+(A_\eta^+ - \eta I) = \begin{bmatrix} B^TB & 0 \\ 0 & BB^T \end{bmatrix}$$

will generate the same iterates as MINRES applied to the original matrix A_η^+. Furthermore, as we have

$$\begin{bmatrix} B^TB & 0 \\ 0 & BB^T \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ BB^T\mathbf{w} \end{bmatrix},$$

the residual vector $\mathbf{r}_k$ must retain the zero n-vector in its top part throughout. The method is therefore equivalent to applying MINRES to a system with coefficient matrix

$$BB^T = Q^{-1}\mathcal{B}H^{-1}\mathcal{B}^TQ^{-T},$$

that is, to the preconditioned Schur complement matrix. Using the proof of Corollary 3.5, it can in fact be shown that if $\mathbf{z}_j$ is the jth MINRES iterate for

$$BB^T\mathbf{z} = B\bar{\mathbf{f}}, \tag{4.1}$$

then the k^{th} MINRES (GMRES) iterate for (1.3) is given by

$$\hat{\mathbf{x}}_k = \begin{bmatrix} \frac{1}{\eta}(\bar{\mathbf{f}} - B^T\mathbf{z}_j) \\ \mathbf{z}_j \end{bmatrix},$$

where $j = \lfloor \frac{k}{2} \rfloor$. Note that in the case where B has full rank, the coefficient matrix in (4.1) is symmetric positive definite so the cheaper conjugate gradient method can be applied: the most efficient and stable implementation would probably be the LSQR algorithm [16] applied to $\min\|B^T\bar{\mathbf{y}} - \bar{\mathbf{f}}\|$.

The convergence of MINRES applied to (4.1) or LSQR applied to the equivalent problem $\min\|B^T\bar{\mathbf{y}} - \bar{\mathbf{f}}\|$ will exhibit no plateaux and the methods will therefore converge in half the number of iterations needed for MINRES with A_η^+. For the plus case where B has full rank and $\eta = 1$, a comparable result is proved in [3]: namely, that using the SYMMBK method for the matrix

$$\begin{bmatrix} I & B^T \\ B & 0 \end{bmatrix}$$

is equivalent to applying the conjugate gradient method to (4.1). For a nonsymmetric system preconditioned by its symmetric part to give a normal matrix,

Eisenstat [5] has proved the equivalence of the even iterates of the appropriate conjugate gradient method and the iterates of conjugate gradients applied to the normal equations.

We also note here that if we are prepared to solve a nonsymmetric system, then a preconditioner of form (1.2) is not necessarily the best choice. For example, applying the block upper triangular preconditioner

$$\begin{bmatrix} H & \mathcal{B}^T \\ 0 & -M \end{bmatrix}$$

on the right of system (1.1) gives a system with coefficient matrix

$$\begin{bmatrix} I & 0 \\ \mathcal{B}H^{-1} & \mathcal{B}H^{-1}\mathcal{B}^T M^{-1} \end{bmatrix}.$$

This may be solved using GMRES (see, for example, [6]).

5 Summary.

The clear conclusion of the above observations on the results of Section 3 is that the most efficient way of solving (1.1) iteratively where H is readily invertible is to solve a Schur complement system involving $\mathcal{B}H^{-1}\mathcal{B}^T$ using MINRES or conjugate gradients, or, in practical situations where roundoff is likely to be significant, to employ LSQR on the equivalent problem $\min\|H^{-1/2}\mathcal{B}^T\bar{\mathbf{y}} - H^{-1/2}\bar{\mathbf{f}}\|$. We have shown that use of indefinite preconditioning and GMRES cannot be superior. When H is not readily factorised or inverted, use of MINRES with a positive definite preconditioner is still a feasible approach.

Acknowledgement.

We are very grateful to Professors Stan Eisenstat and Michael Saunders for detailed comments on earlier drafts, which have improved and strengthened the results presented.

REFERENCES

1. W. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
2. Å. Björck, *Iterative refinement of linear least squares solutions*, BIT, 7 (1967), pp. 257–278.
3. R. Chandra, *Conjugate Gradient Methods for Partial Differential Equations*, PhD thesis, Yale University Res. Rep. 129, 1978.
4. I. S. Duff, *The solution of augmented systems*, in Numerical Analysis 1993, D.F. Griffiths and G.A. Watson, eds., Longman, London, 1994, pp. 40–55.
5. S. Eisenstat, *Some observations on the generalized conjugate gradient method*, in Numerical Methods, V. Pereyra and A. Reinoza, eds., Lecture Notes in Mathematics, Volume 1005, Springer-Verlag, Berlin, 1983, pp. 99–107.
6. H. C. Elman and D. J. Silvester, *Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations*, SIAM J. Sci. Comput., 17 (1996), pp. 33–46.

7. V. Faber and T. Manteuffel, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., 21 (1984), pp. 352–362.
8. B. Fischer, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, Wiley-Teubner Series in Advances in Numerical Mathematics, Wiley-Teubner, Chichester, 1996.
9. M. Fortin and F. Brezzi, *Mixed and Hybrid Finite Elements Methods*, Springer-Verlag, Berlin, 1991.
10. R. W. Freund and N. M. Nachtigal, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.
11. R. W. Freund and N. M. Nachtigal, *A new Krylov-subspace method for symmetric indefinite linear systems*, in Proceedings of the 14th IMACS World Congress on Computational and Applied Mathematics, W. F. Ames, ed., IMACS, 1994, pp. 1253–1256.
12. G. H. Golub and C. F. Van Loan, *Matrix Computations, 2nd ed.*, Johns Hopkins University Press, London, 1989.
13. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Stand., 45 (1950), pp. 255–282.
14. N. M. Nachtigal, S. C. Reddy and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795.
15. C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. 12 (1975), pp. 617–629.
16. C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software 8 (1982), pp. 43–71.
17. Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
18. M. A. Saunders, *Solution of sparse rectangular systems using LSQR and Craig*, BIT, 35 (1995), pp. 588–604.
19. D. J. Silvester and A. J. Wathen, *Fast iterative solution of stabilised Stokes systems part II: using general block preconditioners*, SIAM J. Numer. Anal., 31 (1994), pp. 1352–1367.
20. G. Szegö, *Orthogonal Polynomials*, AMS Colloquium Publications XXIII (revised ed.), AMS, New York, 1959.