A DIFFERENT APPROACH TO BOUNDING THE MINIMAL RESIDUAL NORM IN KRYLOV METHODS

ILSE C. F. IPSEN*

Abstract. In the context of Krylov methods for solving systems of linear equations, expressions and bounds are derived for the norm of the minimal residual, like the one produced by GMRES or MINRES.

It is shown that the minimal residual norm is large as long as the Krylov basis is well-conditioned. In the context of non-normal matrices, examples are given where the minimal residual norm is a function of the departure of the matrix from normality, and where the decrease of the residual norm depends on how large the departure from normality is compared to the eigenvalues. With regard to normal matrices, the Krylov matrix is unitarily equivalent to a row-scaled Vandermonde matrix and the minimal residual norm in iteration i is proportional to a product of i relative eigenvalue separations. Arguments are given for why normal matrices with complex eigenvalues can produce larger residual norms than Hermitian matrices, and why indefinite matrices can produce larger residual norms than definite matrices.

Key words. linear system, Krylov methods, GMRES, MINRES, Vandermonde matrix, eigenvalues, departure from normality

AMS subject classification. 15A03, 15A06, 15A09, 15A12, 15A18, 15A60, 65F10, 65F15, 65F20, 65F35.

1. Introduction.

We consider the problem of solving a system of linear equations $Ax = b$ by Krylov methods which in iteration i approximate the solution x by a vector x_i from the Krylov space

$$\operatorname{span}\{b, Ab, \ldots, A^{i-1}b\}.$$

We are interested in the residual $b - Ax_i$ with minimal two-norm, and how fast the minimal residual norm decreases during the course of the iterations. Such a residual is produced (in exact arithmetic) by GMRES [46, 53] and, if the matrix is Hermitian, by MINRES [44].

Usually the convergence of Krylov methods is analysed by deriving upper bounds on the norm of the residual or of the error in terms of a polynomial in the matrix, e.g. [1, §13], [26, §3], [38, §§2,3], [45, §6.11]. In the particular case of GMRES, upper bounds on the residual norm in terms of polynomials are given in [4], [41, §3], [46, §3.4], and the tightness of these bounds is examined in [27, 29, 49]. Convergence analyses based on Ritz values are given in [5, 43, 52]. The case of nearly singular matrices is analysed in [3], and comparisons with other methods are made in [2, 33].

In spite of all this work Greenbaum [26, page 58] states that '[...] it remains an open problem to describe the convergence of GMRES in terms of simple characteristic properties of the coefficient matrix', and [28, §4] 'it remains an open problem to determine the most appropriate set of system parameters for describing the behavior of GMRES'. Our motivation is to make a contribution towards the solution of these problems. Our approach differs from existing approaches because we ignore the way GMRES is implemented (e.g. via Arnoldi's method) and concentrate on the fact that the two-norm of the residual is minimized at every iteration.

* Center for Research In Scientific Computation, Department of Mathematics, North Carolina State University, P. O. Box 8205, Raleigh, NC 27695-8205 (ipsen@math.ncsu.edu, http://www4.ncsu.edu/~ipsen/info.html). This research was supported in part by NSF grant CCR-9400921.

Our expressions and lower bounds for the minimal residual norm give information about the convergence (in exact arithmetic) of a large number of Krylov methods. Although residuals of minimal norm are, in general, only produced by GMRES and MINRES, lower bounds on the minimal residual provide negative information on the convergence for Krylov methods such as QMR [19, 18], TFQMR [17], CGS [47], BiCG [39, 16], Bi-CGSTAB [51] for non-Hermitian matrices[1], all Krylov methods for Hermitian matrices, and polynomially preconditioned methods (excluded are methods based on normal equations and methods with restarting). If a lower bound on the minimal residual norm in iteration i is large, none of these Krylov methods can converge in i iterations.

1.1. Overview. In §2 the minimal residual norm in iteration i is expressed in terms of the pseudo-inverse of the Krylov matrix in iteration $i+1$. As a consequence the minimal residual norm is large if the next larger Krylov basis is well-conditioned. Therefore a Krylov method cannot converge as long as the Krylov basis is well-conditioned. However the converse is not true. An example illustrates that an ill-conditioned Krylov basis does not necessarily produce a residual of small norm.

In §3 we give an example where the minimal residual norm can be expressed in terms of the departure of the matrix from normality, and the decrease of the residual norm depends on the how large the departure from normality is compared to the eigenvalues. Therefore some relation between departure from normality and eigenvalues may be an appropriate way to describe the convergence of GMRES.

In §4 the current minimal residual norm is related to the previous one via the sine of the angle between the old Krylov space and the new Krylov vector. This results in a lower bound for the convergence rate of GMRES in terms of an angle between Krylov vectors.

In §5 a model problem for normal matrices[2] with a particular right-hand side is analysed. This choice of right-hand side allows us to focus on the effect of the eigenvalue distribution on the residual norm. In this case the Krylov matrix is a rectangular Vandermonde matrix. The minimal residual norm in iteration i is proportional to a product of i relative eigenvalue separations. We also give some arguments for why complex eigenvalues can produce larger residual norms than real eigenvalues, and why indefinite matrices can produce larger residual norms than definite matrices.

In §6 we extend the analysis from §5 to normal matrices with arbitrary right-hand sides. Now the Krylov matrix is unitarily equivalent to a row-scaled Vandermonde matrix, and the minimal residual norm in iteration i is proportional to two factors: a product of i relative eigenvalue separations and the orthogonal projection of b onto an eigenspace.

The Appendix in §7 contains statements and proofs of auxiliary results.

1.2. Notation. The norm $\|\cdot\|$ is the Euclidean two-norm, or spectral norm. The identity matrix of order n is $I = (\, e_1 \quad \dots \quad e_n \,)$ with columns e_i. The conjugate transpose of a matrix A is A^*; and the Moore-Penrose inverse of a full column rank matrix A is $A^\dagger \equiv (A^*A)^{-1}A^*$. The column vector e is the vector of all ones.

Following [14, §1], [26, p 2], [45, §6.1], we define a Krylov method in this context as a method that approximates in iteration i the solution to the linear system $Ax = b$

[1] The methods above that use the conjugate transpose of A do so only for computing coefficients in linear combinations, the approximate solution still lies in the Krylov space generated by A.

[2] Normal matrices are unitarily similar to a diagonal matrix but, unlike Hermitian matrices, their eigenvalues can be complex. Normal matrices include Hermitian, real symmetric, skew-Hermitian, real skew-symmetric, unitary, real orthogonal and diagonal matrices.

by a vector from the *Krylov space*

$$\mathcal{K}_i \equiv \text{span}\{b, Ab, \ldots, A^{i-1}b\}, \qquad i \geq 1.$$

This includes all the methods mentioned in the previous section. The *Krylov matrix*

$$K_i \equiv \begin{pmatrix} b & Ab & \ldots & A^{i-1}b \end{pmatrix}, \qquad i \geq 1$$

contains the *Krylov basis*.

If an initial guess $x_0 \neq 0$ is available, it is incorporated into the right-hand side and one considers instead the solution of the linear system

$$Ay = c, \qquad \text{where } c \equiv b - Ax_0, \qquad y \equiv x_0 + x.$$

The Krylov space for this linear system is

$$\text{span}\{c, Ac, \ldots, A^{i-1}c\}, \qquad i \geq 1.$$

Therefore, we assume without loss of generality that any initial guess has already been incorporated into the right-hand side b.

2. Nothing Happens as Long as the Krylov Basis is Well-Conditioned. We show that the minimal residual norm is related to the conditioning of the Krylov basis in the next larger space, and that the residual norm cannot be small as long as the Krylov basis is well conditioned.

The goal is to find a Krylov vector $z \in \mathcal{K}_i$ that makes the residual norm $\|b - Az\|$ small. But $z \in \mathcal{K}_i$ means that $z = K_i y$ for some y, hence

$$\|b - Az\| = \|b - AK_i y\|.$$

However since $K_{i+1} = \begin{pmatrix} b & AK_i \end{pmatrix}$, making the residual norm small means approximating the first column of K_{i+1} by the remaining columns. If the residual norm can be made small then the columns of K_{i+1} must be almost linearly dependent, which means $\|K_{i+1}^\dagger\|$ is large. The result below quantifies this observation.

THEOREM 2.1. *If K_{i+1} has full column rank then*

$$\min_{z \in \mathcal{K}_i} \|b - Az\| = \frac{1}{\|e_1^* K_{i+1}^\dagger\|}.$$

Proof. The minimal residual norm for a vector from $\mathcal{K}_i$ is

$$\min_{z \in \mathcal{K}_i} \|b - Az\| = \min_y \|b - AK_i y\|.$$

Apply Lemma 7.1 in the Appendix with $B_1 \equiv AK_i$ and $B = K_{i+1}$. □

Therefore if the columns of the next Krylov matrix, K_{i+1}, are very linearly independent then the residual norm in the current iteration must be large. The iterative methods MINRES (for Hermitian matrices) [44] and GMRES (for non-Hermitian matrices) [46, 53] produce such a residual of minimal norm. The following example represents an extreme case where the residual norms stagnate until the last iteration.

EXAMPLE 1. *Let $Ax = b$ be a linear system with a circulant matrix, i.e. [2, Example 3.1], [41, Example C],*

$$A = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 1 & & & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In iteration i the relevant Krylov matrix is

$$K_{i+1} = \begin{pmatrix} e_1 & e_n & \dots & e_{n-i+1} \end{pmatrix}.$$

Since K_{i+1} has orthonormal columns, $\|e_1^* K_{i+1}^\dagger\| = 1$ for all $i \leq n-1$. Therefore the residual norms remain maximal until the last iteration

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} = 1, \qquad 1 \leq i \leq n-1.$$

The same thing happens in the next example.

EXAMPLE 2. *Consider the linear system $Ax = b$ where*

$$A = \begin{pmatrix} 1 & & & \\ & \omega & & \\ & & \ddots & \\ & & & \omega^{n-1} \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

and $\omega \equiv e^{2\pi\sqrt{-1}/n}$ is the nth root of unity. The matrix A is unitary with eigenvalues evenly distributed around the unit circle.

In the last iteration the Krylov matrix K_n has elements $\omega^{(i-1)(j-1)}$, $1 \leq i, j \leq n$. Hence $K_n = \sqrt{n} F_n$, where F_n is the Fourier matrix. Since F_n is unitary, $K_n^{-1} = n^{-1} K_n^*$ and

$$\|e_1^* K_n^{-1}\| = \|e_1^* K_n^*\|/n = 1/\sqrt{n}.$$

Again, the residual norms remain maximal until the last iteration

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} = 1, \qquad 1 \leq i \leq n-1.$$

Theorem 2.1 implies a lower bound for the relative residual norm.

COROLLARY 2.2. *If K_{i+1} has full column rank then*

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} = \frac{1}{\|K_{i+1} e_1\| \, \|e_1^* K_{i+1}^\dagger\|} \geq \frac{1}{\|K_{i+1}\| \, \|K_{i+1}^\dagger\|}.$$

This means, any vector in the current Krylov space has a large relative residual norm if the Krylov basis in the next iteration is well-conditioned. Therefore a Krylov method cannot converge as long as the Krylov basis is well-conditioned.

EXAMPLE 3. *For the linear system in Example 1 Corollary 2.2 is tight for all i because K_{i+1} has orthonormal columns.*

For the linear system in Example 2 Corollary 2.2 is tight for $i = n-1$ because K_n is a Vandermonde matrix whose nodes are roots of unity, hence its two-norm condition number is one [21, Example 6.4].

Unfortunately the converse of Corollary 2.2 does not hold: An ill-conditioned Krylov basis does not necessarily produce a residual of small norm, as the example below illustrates. In particular this implies that the lower bound in Corollary 2.2 is not always tight.

EXAMPLE 4. *Let A be a square matrix whose field of values contains 0, and choose b such that $b^* A b = 0$.*

The relevant Krylov matrix in the first iteration, $K_2 = (\, b \quad Ab\,)$, has orthonormal columns and

$$K_2^* K_2 = \begin{pmatrix} \|b\| & \\ & \|Ab\| \end{pmatrix}^2 .$$

The condition number of K_2 is

$$\|K_2\|\,\|K_2^\dagger\| = \max\{\frac{\|Ab\|}{\|b\|}, \frac{\|b\|}{\|Ab\|}\}.$$

Since $\|e_1^* K_2^\dagger\| = 1/\|b\|$, the residual norm in the first iteration is maximal,

$$\min_{z \in \mathcal{K}_1} \frac{\|b - Az\|}{\|b\|} = 1.$$

Although the residual norm is maximal, the condition number of K_2 can be arbitrarily large. For instance, let A be a diagonal matrix of even order n with diagonal elements

$$\lambda_j = jh, \qquad \lambda_{n-j+1} = -jh, \qquad 1 \leq j \leq n/2,$$

where $h > 1$, and let b be the vector of all ones. Then $b^* Ab = 0$, and

$$\|Ab\|^2 = 2h^2 \sum_{j=1}^{n/2} j^2 = \frac{h^2}{12}\, n(n+1)(n+2).$$

Thus

$$\|K_2\|\,\|K_2^\dagger\| = \frac{\|Ab\|}{\|b\|} = h\sqrt{\frac{(n+1)(n+2)}{12}} = O(hn) \to \infty \quad \text{as} \quad h \to \infty.$$

Thus, as the spacing between eigenvalues increases, the Krylov matrix becomes more ill conditioned, yet the minimal residual norm in the first iteration remains maximal.

3. The Residual Norm Can Depend on the Departure From Normality. We present examples where the residual norm can decrease as a function of the departure of the matrix from normality – even if there is only a single eigenvalue. Thus the convergence of a Krylov method can be governed by the departure from normality.

Although it is known that the convergence of GMRES for a highly non-normal matrix is not determined by eigenvalues alone [28, §4], [41, §3], [42, §2], the examples below present a quantitative dependence on the departure from normality: The decrease in the minimal residual norm depends on how large the departure from normality is compared to the magnitude of the eigenvalue.

Every square matrix A has a Schur decomposition $A = Q(\Lambda + N)Q^*$, where Q is unitary, Λ is diagonal with the eigenvalues of A on its diagonal, and N is upper triangular with zero diagonal. If $N = 0$ then A is *normal*. Henrici calls $\|N\|$ the *two-norm departure of A from normality* [31, §1.2] (or simply: *non-normality*).

Let

$$A = \begin{pmatrix} 1 & \eta & & \\ & 1 & \ddots & \\ & & \ddots & \eta \\ & & & 1 \end{pmatrix}.$$

The two-norm departure of A from normality is $|\eta|$. If $\eta = 0$ then $A = I$ is normal. When $\eta \neq 0$ then A is diagonally similar to a Jordan block, $A = XJX^{-1}$, where

$$J = \begin{pmatrix} 1 & 1 & & \\ & 1 & \ddots & \\ & & \ddots & 1 \\ & & & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & & & \\ & \eta & & \\ & & \ddots & \\ & & & \eta^{n-1} \end{pmatrix},$$

and n is the order of A. Hence the single eigenvalue 1 is maximally defective for $\eta \neq 0$.

Consider the linear system

$$(3.1) \qquad Ax = b \qquad \text{where } A = \begin{pmatrix} 1 & \eta & & \\ & 1 & \ddots & \\ & & \ddots & \eta \\ & & & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

When $\eta = 0$ then b is an eigenvector of A, while for $\eta \neq 0$ b is a principal vector of maximal grade.

In iteration $i \geq 1$, the Krylov matrix is

$$K_{i+1} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \cdot^{\cdot^{\cdot}} & \eta^i \\ 0 & 0 & 0 & \cdot^{\cdot^{\cdot}} & \alpha_{i,i-1}\eta^{i-1} \\ 0 & 0 & \eta^2 & \cdot^{\cdot^{\cdot}} & \vdots \\ 0 & \eta & \alpha_{21}\eta & \dots & \alpha_{i1}\eta \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix},$$

where $\alpha_{ij} = \binom{i}{j}$ (see Lemma 7.2). The Krylov basis is well conditioned when $|\eta| \geq 1$.

Below the minimal residual norm is expressed as a function of the departure from normality.

THEOREM 3.1. *Let $Ax = b$ be as in (3.1). Then*

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} = \sqrt{\frac{|\eta|^{2i}}{1 + |\eta|^2 + \cdots + |\eta|^{2i}}}.$$

Proof. Lemma 7.2 in the Appendix states

$$\|e_1^* K_{i+1}^\dagger\| = \sqrt{\frac{|\eta|^{2i}}{1 + |\eta| + \cdots + |\eta|^{2i}}}.$$

The proof follows from Theorem 2.1. □

Therefore, although the matrix A looks perfect when it comes to the eigenvalue distribution, it is the departure from normality that governs the convergence. In particular, the residual norm decreases very slowly when the departure from normality is large.

COROLLARY 3.2. *Let $Ax = b$ be as in (3.1). If $|\eta| \geq 1$ then*

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \geq \frac{1}{\sqrt{i+1}},$$

and if $|\eta| \ll 1$ then

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \approx |\eta|^i.$$

This means the minimal residual norm in iteration i is large when the non-normality of the matrix is large. In this case a Krylov method can be expected to use the maximal number of iterations in order to attain reasonable accuracy. When the non-normality is small, the minimal residual norm decreases as a power of the non-normality.

We extend the above statements to eigenvalues other than one. Consider the linear system

$$(3.2) \quad Ax = b \qquad \text{where } A = \begin{pmatrix} \lambda & \eta & & \\ & \lambda & \ddots & \\ & & \ddots & \eta \\ & & & \lambda \end{pmatrix}, \quad \lambda \neq 0, \qquad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

When $\lambda = 0$ then no solution to $Ax = b$ lies in a Krylov space $\mathcal{K}_i$ for any $i \geq 1$ [36, Theorem 2]. So the interesting case is $\lambda \neq 0$.

THEOREM 3.3. *Let $Ax = b$ be as in (3.2). Then*

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} = \sqrt{\frac{|\tau|^{2i}}{1 + |\tau|^2 + \cdots + |\tau|^{2i}}}, \qquad \tau \equiv \frac{\eta}{\lambda}.$$

Proof. Factor out λ and write $A = \lambda \hat{A}$, where $\hat{A}$ is as the matrix in (3.1),

$$\hat{A} = \begin{pmatrix} 1 & \tau & & \\ & 1 & \ddots & \\ & & \ddots & \tau \\ & & & 1 \end{pmatrix}.$$

This implies for the Krylov matrix

$$K_{i+1} = (\, b \quad Ab \quad \ldots \quad A^i b \,) = (\, b \quad \lambda \hat{A} b \quad \ldots \quad \lambda^i \hat{A}^i b \,) = \hat{K}_{i+1} D,$$

where

$$\hat{K}_{i+1} \equiv (\, b \quad \hat{A} b \quad \ldots \quad \hat{A}^i b \,), \qquad D \equiv \begin{pmatrix} 1 & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda^i \end{pmatrix}.$$

For the residual norm we need

$$K_{i+1}^{\dagger} = (K_{i+1}^* K_{i+1})^{-1} K_{i+1}^* = D^{-1} \hat{K}_{i+1}^{\dagger}.$$

Since the first diagonal element of D is equal to one,

$$e_1^* K_{i+1}^\dagger = e_1^* \hat{K}_{i+1}^\dagger.$$

This implies for the minimal residual of $Ax = b$,

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} = \frac{1}{\|e_1^* \hat{K}_{i+1}^\dagger\|} = \min_{z \in \mathcal{K}_i} \frac{\|b - \hat{A}z\|}{\|b\|}.$$

Since the linear system $\hat{A}x = b$ is of the form (3.1), application of Theorem 3.1 gives the desired result. □

Therefore the minimal residual norm is a function of non-normality divided by eigenvalue. It makes sense to distinguish between two cases. The matrix A in (3.2) is *highly non-normal in the context of Krylov methods* if the non-normality is at least as large in magnitude as the eigenvalue, i.e. $|\eta| \geq |\lambda|$. The matrix is *weakly non-normal* if the non-normality is much smaller in magnitude than the eigenvalue, $|\eta| \ll |\lambda|$.

COROLLARY 3.4. *Let $Ax = b$ be as in (3.2). If $|\eta| \geq |\lambda|$ then*

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \geq \frac{1}{\sqrt{i+1}}$$

and if $|\eta| \ll |\lambda|$ then

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \approx \left|\frac{\eta}{\lambda}\right|^i.$$

This means the residual norm is large when the matrix is highly non-normal, and it decreases faster as the non-normality of the matrix becomes weaker.

4. Relation Between Successive Residuals. We show that two successive minimal norm residuals are related by the sine of the angle between the current Krylov space and the new Krylov vector. This leads to a lower bound on the convergence rate of GMRES and MINRES.

Let

$$\rho_i \equiv \min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|}$$

be the relative residual in iteration i.

THEOREM 4.1. *If K_{i+1} has full column rank then*

$$\sin\theta_i \, \rho_{i-1} \leq \rho_i \leq \rho_{i-1},$$

where $0 < \theta_i \leq \pi/2$ is the angle between $\mathcal{K}_i$ and $A^i b$, and

$$\sin\theta_i = \|(I - K_i K_i^\dagger) A^i b\| / \|A^i b\|.$$

Proof. Applying Lemmas 7.3 and 7.4 in the Appendix to $A \equiv K_{i+1}$, $B \equiv K_i$, $c \equiv A^i b$ and $\theta \equiv \theta_i$ gives

$$\|e_1^* K_{i+1}^\dagger\| \leq \|e_1^* K_i^\dagger\| \, \frac{1}{\sin\theta_i},$$

The desired inequalities follow from Theorem 2.1. $\square$

Thus the current residual norm is not much smaller than the previous residual norm if the most recent Krylov vector is almost orthogonal to the old Krylov space $\mathcal{K}_i$. Therefore the convergence rate of GMRES can be bounded below by the sine of an angle between Krylov vectors.

EXAMPLE 5. *The linear system from Example 1,*

$$A = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 1 & & & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

presents an extreme case, where each new Krylov vector is orthogonal to the previous Krylov space, until the very last iteration. Since

$$K_i = \begin{pmatrix} e_1 & e_n & \dots & e_{n-i+2} \end{pmatrix}, \qquad A^i b = e_{n-i+1},$$

we have $K_{i+1}^\dagger A^i b = 0$. *Hence*

$$(I - K_i K_i^\dagger) A^i b = A^i b = e_{n-i+1}$$

and $\sin\theta_i = 1$ *for all* $i \leq n-1$.

5. A Model Problem for Normal Matrices. Before deriving general residual bounds for normal matrices we consider a model problem with a particular right-hand side to focus on the effect of the eigenvalue distribution on the residual. For this model problem we derive lower and upper bounds on the minimal residual norm in terms of relative eigenvalue separations. We also give some arguments for why complex eigenvalues can produce larger residual norms than real eigenvalues, and why indefinite matrices can produce larger residual norms than definite matrices.

In the model problem matrix the linear system is

$$Ax = b \quad \text{where} \quad A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{5.1}$$

The diagonal matrix represents any normal matrix since the two-norm is invariant under multiplication by unitary matrices. The Krylov matrix

$$K_i = \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{i-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{i-1} \end{pmatrix}$$

is a $n \times i$ Vandermonde matrix. Corollary 2.2 implies that the relative residual norm for the model problem is bounded below by the condition number of the rectangular Vandermonde K_{i+1}.

EXAMPLE 6. *Let* $Ax = b$ *be as in (5.1) where* A *has equally spaced positive eigenvalues*

$$\lambda_j = j/n, \qquad 1 \leq j \leq n.$$

Fix the number of iterations i. As the matrix order n grows the condition number of the Vandermonde matrix approaches the square root of the condition number of a Hilbert matrix [15, Proposition 4.2]

$$\lim_{n\to\infty} \|K_{i+1}\|\,\|K_{i+1}^{\dagger}\| = \sqrt{\|H_{i+1}\|\,\|H_{i+1}^{-1}\|},$$

where H_j is the jth segment of the Hilbert matrix with elements $\frac{1}{l+k-1}$, $1 \le k,l \le j$. Corollary 2.2 implies that the minimal residual norm in the model problem has the asymptotic lower bound

$$\lim_{n\to\infty} \min_{z\in\mathcal{K}_i} \frac{\|b-Az\|}{\|b\|} \ge \frac{1}{\sqrt{\|H_{i+1}\|\,\|H_{i+1}^{-1}\|}} \approx e^{-1.9\frac{i+1}{2}}$$

since the condition number of a Hilbert matrix grows exponentially [32, §26.1]. For instance [32, Table 26.1]

$$\|H_{11}\|\,\|H_{11}^{-1}\| \approx 5\cdot 10^{-14}.$$

This means when the matrix in the model problem is large it takes at least 10 iterations to get the relative residual norm down to single precision accuracy of 10^{-7}.

However we have no information about the accuracy of this bound, it could be very loose. The matrix of order 199 in [43] required many more iterations.

Because the results for rectangular Vandermonde matrices in [15] are essentially asymptotic, we could not see how to fully exploit them here. At this point it seems easiest to just convert the rectangular Vandermonde matrix to a square Vandermonde matrix.

LEMMA 5.1. *Let A and b be as in (5.1). If A has $i+1$ distinct eigenvalues then K_{i+1} has a submatrix V_{i+1} of order $i+1$ such*

$$\frac{1}{\kappa_{i+1}} \le \min_{z\in\mathcal{K}_i} \frac{\|b-Az\|}{\|b\|} \le \sqrt{(i+1)(n-i)}\,\frac{1}{\kappa_{i+1}},$$

where

$$\kappa_{i+1} \equiv \|V_{i+1}e_1\|\,\|e_1^* V_{i+1}^{-1}\|,$$

and $|\det(V_{i+1})|$ is maximal among all submatrices of order $i+1$ of K_{i+1}.

Proof. According to Lemma 7.6 in the Appendix, K_{i+1} has a submatrix V_{i+1} of order $i+1$ such that $|\det(V_{i+1})|$ is maximal and

$$\frac{1}{\sqrt{(i+1)(n-i)}}\,\|e_1^*V_{i+1}^{-1}\| \le \|e_1^*K_{i+1}^{\dagger}\| \le \|e_1^*V_{i+1}^{-1}\|.$$

Applying this bound to Theorem 2.1 gives

$$\frac{1}{\|e_1^*V_{i+1}^{-1}\|} \le \min_{z\in\mathcal{K}_i} \|b-Az\| \le \sqrt{(i+1)(n-i)}\,\frac{1}{\|e_1^*V_{i+1}^{-1}\|}.$$

□

Therefore the minimal residual norm in iteration i can be bounded in terms of a (square) Vandermonde matrix of order $i+1$ whose nodes are eigenvalues of A.

The fact that the determinant of this Vandermonde matrix is maximal suggests that GMRES and MINRES process outlying, far-apart eigenvalues first.

Since

$$\kappa_{i+1} \leq \|V_{i+1}\| \, \|V_{i+1}^{-1}\|$$

one can apply existing results for condition numbers of square Vandermonde matrices to obtain bounds for the minimal residual norm. According to [8, §1] 'Vandermonde matrices have a reputation of being ill conditioned. This reputation is well-deserved for Vandermonde matrices whose nodes are real'. In particular, the condition number of a Vandermonde matrix grows exponentially when the nodes are positive [23, §2], or are located symmetrically around the origin on the real line [23, §3], or are either all less than one in magnitude or larger than one [50, §4]. Thus there is hope for fast convergence of GMRES and MINRES (in exact arithmetic) when the matrix is positive-definite or even indefinite. However, as Example 2 illustrates, allowing complex nodes makes it easier to construct well-conditioned Vandermonde matrices [8, §1], which in turn produce large residual norms for many iterations. In Example 2 all residual norms are large because the eigenvalues of A are nth roots of unity. This can also happen when the eigenvalues lie on the circle but are not necessarily roots of unity.

EXAMPLE 7. *Let A and b be as in (5.1), and suppose that the eigenvalues of A are constructed from a Van der Corput sequence as follows. Let*

$$k = \sum_{j=0}^{\infty} k_j 2^j, \qquad k_j \in \{0, 1\}$$

be the binary representation of a non-negative number k. The sequence $\{\alpha_k\}_{k=0}^{\infty}$, where

$$\alpha_k \equiv \sum_{j=0}^{\infty} k_j 2^{-j-1}, \qquad k_j \in \{0, 1\},$$

is called 'Van der Corput sequence' [8, §1]. Let the eigenvalues of A be $\lambda_k = e^{2\pi\sqrt{-1}\alpha_k}$, $1 \leq k \leq n$. Then [8, Corollary 3]

$$\|V_n\| \, \|V_n^{-1}\| < \sqrt{2n}.$$

Therefore

$$\min_{z \in \mathcal{K}_{n-1}} \frac{\|b - Az\|}{\|b\|} \geq \frac{1}{\sqrt{2n}},$$

which means that even after $n-1$ iterations the minimal residual norm is still not small.

However, a Vandermonde matrix can be ill conditioned while the first row of its inverse has small norm, as Example 4 illustrates. According to Corollary 2.2, the condition number is only a lower bound for the residual norm, and it may not be tight. Therefore the existing results on condition numbers of Vandermonde matrices [20, 21, 22, 23, 50] appear to be of limited use in this context.

Below we use the fact that the elements in the first row of the inverse of a Vandermonde matrix have simple expressions to bound the norm of the first row of the inverse.

THEOREM 5.2. *Let A and b be as in (5.1). If A has $i+1$ distinct eigenvalues then*

$$\frac{1}{\sqrt{n(i+1)}}\,\gamma_{i+1} \leq \min_{z\in\mathcal{K}_i} \frac{\|b-Az\|}{\|b\|} \leq \sqrt{\frac{(i+1)(n-i)}{n}}\,\gamma_{i+1},$$

where

$$\gamma_{i+1} \equiv \min_{1\leq j\leq i+1} \prod_{l=1,l\neq j}^{i+1} \frac{|\lambda_l-\lambda_j|}{|\lambda_l|},$$

and $\lambda_1,\ldots,\lambda_{i+1}$ are $i+1$ distinct eigenvalues of A that maximise

$$\prod_{j=1}^{i+1}\prod_{l=j+1}^{i+1} |\lambda_l-\lambda_j|.$$

Proof. The norm of $e_1^*V_{i+1}^{-1}$ in Lemma 5.1 can be bounded by an element of largest magnitude,

$$\max_{1\leq j\leq i+1} |(V_{i+1}^{-1})_{1j}| \leq \|e_1^*V_{i+1}^{-1}\| \leq \sqrt{i+1}\max_{1\leq j\leq i+1} |(V_{i+1}^{-1})_{1j}|.$$

Define

$$\gamma_{i+1} \equiv \min_{1\leq j\leq i+1} \frac{1}{|(V_{i+1}^{-1})_{1j}|},$$

and express the residual bounds in terms of γ_{i+1},

$$\frac{1}{\sqrt{i+1}}\,\gamma_{i+1} \leq \min_{z\in\mathcal{K}_i} \|b-Az\| \leq \sqrt{(i+1)(n-i)}\,\gamma_{i+1}.$$

The expression for the elements in the first row of the inverse of a Vandermonde matrix in Lemma 7.7 implies

$$|(V_{i+1}^{-1})_{1j}| = \prod_{l=1,l\neq j}^{i+1} \frac{|\lambda_l|}{|\lambda_l-\lambda_j|},$$

which gives the desired expression for γ_{i+1}. □

Since both, lower and upper bounds contain the factor γ_{i+1}, the minimal residual norm in iteration i is a multiple of γ_{i+1}. This means the minimal residual norm is proportional to a product of pair-wise relative distances between $i+1$ eigenvalues, and these $i+1$ eigenvalues maximise a product of pair-wise absolute distances. In particular in early iterations the residual is proportional to a product of relative eigenvalue distances where the eigenvalues are far apart in the absolute sense. In this sense Theorem 5.2 corroborates the convergence model for GMRES in [4].

5.1. Iteration $i=1$. We interpret the bounds in the first iteration for different types of matrices.

The minimal residual norm in the first iteration is bounded by

$$\frac{1}{\sqrt{2n}}\,\gamma_2 \leq \min_{z\in\mathcal{K}_1} \frac{\|b-Az\|}{\|b\|} \leq \sqrt{\frac{2(n-1)}{n}}\,\gamma_2,$$

where

$$\gamma_2 = \frac{|\lambda_2 - \lambda_1|}{\max\{|\lambda_1|, |\lambda_2|\}},$$

and λ_1 and λ_2 are eigenvalues at a maximal absolute distance. That is $|\lambda_1 - \lambda_2| = \mathrm{spread}(A)$, where the spread of A is defined as [35, p 264]

$$\mathrm{spread}(A) \equiv \max_{1 \le j < l \le n} |\lambda_j - \lambda_l|.$$

Thus

$$\frac{1}{\sqrt{2n}} \frac{\mathrm{spread}(A)}{\rho(A)} \le \min_{z \in \mathcal{K}_1} \frac{\|b - Az\|}{\|b\|},$$

where $\rho(A) \equiv \max_{1 \le j \le n} |\lambda_j|$ is the spectral radius of the normal matrix A. The residual norm in the first iteration is large if the the relative size of the eigenvalue distribution is large.

We consider three special cases.

1. Let A be Hermitian positive-definite. Then

$$\frac{1}{\sqrt{2n}} \left(1 - \frac{1}{\kappa(A)}\right) \le \min_{z \in \mathcal{K}_1} \frac{\|b - Az\|}{\|b\|},$$

where $\kappa(A) \equiv \rho(A)/\min_{1 \le j \le n} |\lambda_j|$ is the condition number of A. The residual norm in the first iteration is large if the Hermitian positive-definite matrix is ill-conditioned.

2. Let A be singular Hermitian positive semi-definite. Then $\mathrm{spread}(A) = \rho(A)$ and the residual norm in the first iteration is not small,

$$\frac{1}{\sqrt{2n}} \le \min_{z \in \mathcal{K}_1} \frac{\|b - Az\|}{\|b\|}.$$

3. Let A be Hermitian indefinite with

$$\lambda_1 = -\,\mathrm{spread}(A)/2, \qquad \lambda_2 = \mathrm{spread}(A)/2.$$

Then the residual norm can be larger than for a positive semi-definite matrix,

$$\sqrt{\frac{2}{n}} \le \min_{z \in \mathcal{K}_1} \frac{\|b - Az\|}{\|b\|}.$$

The following examples illustrate how the bounds in Theorem 5.2 can be interpreted for different eigenvalue distributions.

5.2. One cluster and One Outlier. Suppose the normal matrix A in (5.1) has one cluster of eigenvalues centered at a point c in the complex plane with radius $\epsilon > 0$, and a single outlier $c + \delta$. The number $|\delta|$ is the absolute distance between cluster and outlier. We make two assumptions: First, the absolute separation between cluster and outlier is much larger than the absolute cluster radius,

$$|\delta| \gg \epsilon.$$

Second, the relative cluster radius is small

$$\frac{\epsilon}{|c|} < 1.$$

In iteration $i = 1$ two eigenvalues with maximal absolute distance are $\lambda_1 = c + \delta$ and $\lambda_2 \approx c$. Then

$$\gamma_2 \approx \min\left\{\left|\frac{\delta}{c}\right|, \left|\frac{\delta}{c+\delta}\right|\right\},$$

and the minimal residual norm in the first iteration is proportional to the relative separation between cluster and outlier.

In iteration $i = 2$ three eigenvalues that maximise the product of pair-wise absolute distances are $\lambda_1 = c + \delta$, and $\lambda_2 \approx \lambda_3 \approx c$. Then

$$\gamma_3 \approx \min\left\{\left|\frac{\delta}{c}\right|^2, \left|\frac{\delta}{c+\delta}\right| \frac{\epsilon}{|c|}\right\}.$$

The first term in the minimum represents the relative separation of the outlier from the two eigenvalues in the cluster. The second term in the minimum consists of two factors, the relative separation of cluster and outlier, and the relative separation between the two eigenvalues in the cluster.

In iteration i,

$$\gamma_{i+1} \approx \min\left\{\left|\frac{\delta}{c}\right|^i, \left|\frac{\delta}{c+\delta}\right| \left(\frac{\epsilon}{|c|}\right)^{i-1}\right\}.$$

If the outlier is farther away from zero than the cluster, i.e. $|c + \delta| \geq |c|$, then

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \approx \left|\frac{\delta}{c+\delta}\right| \left(\frac{\epsilon}{|c|}\right)^{i-1},$$

and the minimal residual norm decreases as a power of the relative cluster radius, with the relative separation between cluster and outlier being only a multiplicative factor. This agrees with the bounds [4, Corollary 4.2] and [46, Theorem 5].

To compare definite versus indefinite matrices, consider two special cases. When cluster and outlier lie in the same quadrant of the complex plane, say $c = \frac{1}{2}\delta$, so $c + \delta = \frac{3}{2}\delta$, then the minimal residual norm in iteration i is proportional to

$$\frac{1}{3}\, 2^i \left(\frac{\epsilon}{|\delta|}\right)^{i-1}.$$

When cluster and outlier lie in different quadrants, say $c = -\frac{1}{2}\delta$, so $c + \delta = \frac{1}{2}\delta$, then the minimal residual norm in iteration i is proportional to

$$2^i \left(\frac{\epsilon}{|\delta|}\right)^{i-1}.$$

In this case there appears to be little difference between definite and indefinite matrices. Although the indefinite matrix has a larger relative separation between cluster and outlier, this separation has little influence on the residual because it occurs only as a multiplicative factor.

5.3. Two Clusters. Suppose the normal matrix A in (5.1) has one cluster of eigenvalues centered at c and a second cluster centered at $c+\delta$. The two clusters have the same number of eigenvalues and the same absolute cluster radius $\epsilon > 0$. The number $|\delta|$ is the absolute cluster separation. We assume again that the absolute cluster separation is much larger than the absolute cluster radius, $|\delta| \gg \epsilon$, and that one of the clusters has a small relative cluster radius, $\epsilon/|c| < 1$.

The minimal residual norm in iteration i is proportional to

$$\min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \approx \min\left\{ \left|\frac{\delta}{c}\right|, \left|\frac{\delta}{c+\delta}\right| \right\} \left|\frac{\delta}{c}\right|^{\frac{i-1}{2}} \left(\frac{\epsilon}{|c+\delta|}\right)^{\frac{i-1}{2}}.$$

The last factor represents a power of the relative cluster radius, and the preceding factors represent a power of the relative cluster separation. In contrast to the previous example of one cluster and one outlier the relative cluster separation now has more influence on the residual norm. This is because the relative separations in γ_{i+1} come from eigenvalues whose absolute distances are maximal, which means γ_{i+1} is made up of $(i-1)/2$ eigenvalues from the cluster around c and $(i-1)/2$ eigenvalues from the cluster around $c+\delta$. Again, this agrees with the more qualitative bound [4, Proposition 5.1].

To compare definite versus indefinite matrices, consider again two special cases. When cluster and outlier lie in the same quadrant of the complex plane, say $c = \frac{1}{2}\delta$, so $c + \delta = \frac{3}{2}\delta$, then the minimal residual norm in iteration i is proportional to

$$\left(\frac{1}{3}\right)^{\frac{i+1}{2}} 2^i \left(\frac{\epsilon}{|\delta|}\right)^{\frac{i-1}{2}}.$$

When cluster and outlier lie in different quadrants, say $c = -\frac{1}{2}\delta$, so $c+\delta = \frac{1}{2}\delta$, then the minimal residual norm in iteration i is proportional to

$$2^i \left(\frac{\epsilon}{|\delta|}\right)^{\frac{i-1}{2}}.$$

The residual norm for the definite matrix is smaller by a factor of $(\frac{1}{3})^{\frac{i+1}{2}}$ than the residual norm for the indefinite matrix. That's because the relative cluster separation for the definite matrix is

$$\left|\frac{\delta}{c+\delta}\right| = \frac{2}{3},$$

which is smaller by a factor of $1/3$ than the relative cluster separation of the indefinite matrix

$$\left|\frac{\delta}{c+\delta}\right| = 2.$$

In general, we conjecture that indefinite matrices can produce larger residual norms than definite matrices because eigenvalues with opposite signs lead to larger relative cluster separations.

6. Normal Matrices. We extend the analysis from the previous section to arbitrary right-hand sides. The minimal residual norm in iteration i is now proportional to two factors: a product of i relative eigenvalue separations and the orthogonal projection of b onto an eigenspace.

A normal matrix A has an eigenvalue decomposition $A = Q\Lambda Q^*$, where Q is unitary,

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

and λ_i are the eigenvalues of A. The elements of the transformed right-hand side are

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \equiv Q^* b.$$

The Krylov matrix is unitarily equivalent to a row-scaled Vandermonde matrix

$$K_i = Q \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{i-1} \\ \vdots & & & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{i-1} \end{pmatrix}.$$

Conversely, in [50, §3] a row-scaled Vandermonde matrix is treated as a Krylov matrix.

Below are the bounds for the minimal residual norm for a normal matrix with arbitrary right-hand side. The bounds are weighted by the orthogonal projection of b onto some eigenspace, as is the product of relative eigenvalue distances. Denote by d the number of distinct eigenvalues of A minus the eigenvalues of A whose eigenspace is orthogonal to b, and by β_j the norm of the orthogonal projection of b onto the eigenspace associated with λ_j.

THEOREM 6.1. *If A is normal and $1 \leq i \leq d-1$ then*

$$\frac{1}{\sqrt{i+1}}\, \gamma_{i+1} \leq \min_{z \in \mathcal{K}_i} \frac{\|b - Az\|}{\|b\|} \leq \sqrt{(i+1)(d-i)}\, \gamma_{i+1},$$

where

$$\gamma_{i+1} \equiv \max_{1 \leq j \leq i+1} \left\{ \frac{\beta_j}{\|b\|} \prod_{l=1, l \neq j}^{i+1} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|} \right\},$$

and $\lambda_1, \dots, \lambda_{i+1}$ are $i+1$ distinct eigenvalues of A that maximise

$$\prod_{j=1}^{i+1} \beta_j \prod_{l=j+1}^{i+1} |\lambda_l - \lambda_j|.$$

Proof. As above, write $K_i = QDV_i$, where D is the diagonal matrix whose elements are b_j and V_i is a $n \times i$ Vandermonde matrix. Since the diagonal matrices D and Λ commute the residual norm can be written as

$$\|b - AK_i y\| = \|De - D\Lambda V_i y\|,$$

where e is the vector of all ones.

First remove rows with multiple eigenvalues in V_i to create a Vandermonde matrix with distinct nodes. Suppose $\lambda_j = \lambda_k$ so that rows j and k of V_i are identical. Eliminate one of these rows in $D\,(\Lambda V_i)$, say row k, by a plane rotation P_{jk},

$$P_{jk}\begin{pmatrix} b_j \\ b_k \end{pmatrix}\begin{pmatrix} \lambda_j & \lambda_j^2 & \dots & \lambda_j^i \end{pmatrix} = \begin{pmatrix} \gamma \\ 0 \end{pmatrix}\begin{pmatrix} \lambda_j & \lambda_j^2 & \dots & \lambda_j^i \end{pmatrix},$$

where $\gamma = \sqrt{|b_j|^2 + |b_k|^2}$ is the norm of the orthogonal projection of b onto the space spanned by columns j and k of Q. Aside from introducing a zero row, the plane rotation P_{jk} preserves the Vandermonde structure. Now look at what happens to the right-hand side De. Since P_{jk} is solely determined by the components b_j and b_k in D, it automatically eliminates the corresponding row in De. At last, permute zero rows to the bottom of the matrices.

The whole transformation can be expressed as a unitary matrix U. The down-sized quantities are

$$\begin{pmatrix} \hat{D}e \\ 0 \end{pmatrix} = U\,(De), \qquad \begin{pmatrix} \hat{D}\hat{\Lambda}\hat{V}_i \\ 0 \end{pmatrix} = U\,(D\Lambda V_i).$$

Hence

$$\|b - AK_i y\| = \|U(De - D\Lambda V_i y)\| = \|\hat{D}\,(e - \hat{\Lambda}\hat{V}_i y)\|.$$

We still have to count how many non-zero rows are left. Since the plane rotations are designed to eliminate multiple eigenvalues, the number of non-zero rows cannot exceed the number of distinct eigenvalues of A. It is less than that if all components of Q^*b associated with a particular eigenvalue λ_j are zero. This can only occur if b is orthogonal to all columns of Q associated with λ_j, which in turn means that b is orthogonal to the whole eigenspace of λ_j.

Hence

$$\hat{D} = \begin{pmatrix} \beta_1 & & \\ & \ddots & \\ & & \beta_d \end{pmatrix}$$

is a non-singular diagonal matrix where β_j is the (non-zero) norm of the orthogonal projection of b onto the eigenspace associated with λ_j. Because the Frobenius norm is invariant under multiplication by a unitary matrix, $\|\hat{D}\|_F = \|D\|_F = \|b\|$, so we have not lost any information about the right-hand side b. Hence

$$\hat{\Lambda} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix}$$

contains distinct eigenvalues of A whose eigenspace is not orthogonal to b, and $\hat{V}_i$ is the corresponding $d \times i$ Vandermonde matrix.

At this point we have

$$\min_{z \in \mathcal{K}_i} \|b - Az\| = \min_y \|\hat{D}e - \hat{\Lambda}\,(\hat{D}\hat{V}_i)\,y\|.$$

Applying Lemma 7.1 in the Appendix with

$$b \equiv De, \qquad B_1 \equiv \hat{\Lambda}\hat{D}\hat{V}_i = \hat{D}\hat{\Lambda}\hat{V}_i, \qquad B \equiv \hat{D}\hat{V}_{i+1} = D\,(\,e \quad \hat{\Lambda}\hat{V}_i\,)\,,$$

gives

$$\min_{z\in\mathcal{K}_i} \|b - Az\| = 1/\|e_1^*(\hat{D}\hat{V}_{i+1})^\dagger\|.$$

Now extract the best non-singular part from the $d \times (i+1)$ matrix $\hat{D}\hat{V}_{i+1}$. According to Lemma 7.6 in the Appendix, there exists a permutation matrix P so that

$$P(\hat{D}\hat{V}_{i+1}) = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix},$$

where $|\det(B_1)|$ is maximal and

$$\frac{1}{\sqrt{(i+1)(d-i)}}\|e_1^* B_1^{-1}\| \le \|e_1^*(\hat{D}\hat{V}_{i+1})^\dagger\| \le \|e_1^* B_1^{-1}\|.$$

Since B_1 is a submatrix of the row-scaled Vandermonde matrix $\hat{D}\hat{V}_{i+1}$, we can write $B_1 = \Delta W_{i+1}$, where

$$W_{i+1} = \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^i \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{i+1} & \dots & \lambda_{i+1}^i \end{pmatrix}, \qquad \Delta = \begin{pmatrix} \beta_1 & & \\ & \ddots & \\ & & \beta_{i+1} \end{pmatrix},$$

and the distinct eigenvalues $\lambda_1, \dots, \lambda_{i+1}$ maximise

$$|\det(B_1)| = |\det(\Delta W_{i+1})| = |\det(\Delta)|\,|\det(W_{i+1})| = \prod_{j=1}^{i+1} \beta_j \prod_{l=j+1}^{i+1} |\lambda_l - \lambda_j|.$$

At last replace W_{i+1}^{-1} in $\|e_1^* B_1^{-1}\|$ by eigenvalues of A. To this end let $w^* \equiv e_1^* W_{i+1}^{-1}\Delta^{-1}$ and bound $\|w^*\|$ in terms of an element of largest magnitude,

$$\max_{1\le j\le i+1} |w_j| \le \|w^*\| \le \sqrt{i+1} \max_{1\le j\le i+1} |w_j|.$$

Lemma 7.7 in the Appendix implies

$$|w_j| = |(W_{i+1}^{-1})_{1j}\,(\Delta^{-1})_{jj}| = \frac{1}{|\beta_j|} \prod_{l=1, l\neq j}^{i+1} \frac{|\lambda_l|}{|\lambda_l - \lambda_j|}.$$

$\square$

When b is the vector of all ones then Theorem 6.1 reduces to Theorem 5.2. In contrast to Theorem 5.2, though, $\|b\|$ has now been incorporated into γ_{i+1}. As in the model problem the minimal residual norm is small if the pair-wise relative differences among all eigenvalues are small. The factor γ_{i+1} does not change if A is multiplied by a scalar. This suggests that scalar multiplication does not affect the convergence of GMRES or MINRES (in exact arithmetic).

After d iterations, GMRES has found the exact solution and $z \in \mathcal{K}_d$ solves $Ax = b$. This fact is well-known [46, Proposition 2], [36, §10] because d is the degree of the minimal polynomial of b with respect to A.

7. Appendix. We express the problem of approximating one column of a matrix by the remaining columns as a least squares problem, and show that the residual of the least squares solution is related to a row of the pseudo-inverse.

LEMMA 7.1 (§5 IN [7], §8 IN [6], §§3,4 IN [48]). *Let*

$$B = \begin{pmatrix} b & B_1 \end{pmatrix}$$

be a matrix with leading column b. *If* B *has full column rank then the residual* r *of the solution to the least squares problem* $\min_y \|b - B_1 y\|$ *satisfies*

$$\|r\| = 1/\|e_1^* B^\dagger\|.$$

Proof. The proof is similar to the one in [6, §8].

Approximating b by columns of B_1 can be expressed as the least squares problem $\min_y \|b - B_1 y\|$. Let y_0 be the solution, and $r \equiv b - B_1 y_0$ the minimal residual.

The Moore-Penrose inverse of the full column rank matrix B is $B^\dagger = (B^*B)^{-1}B^*$. Applying the expression for the inverse in [9, §1] and [13, §2.4] to

$$B^*B = \begin{pmatrix} b^*b & b^*B_1 \\ B_1^*b & B_1^*B_1 \end{pmatrix}$$

gives

$$e_1^*(B^*B)^{-1} = (r^*b)^{-1} \begin{pmatrix} 1 & -y_0^* \end{pmatrix}.$$

From $r^*B_1 = 0$ follows

$$r^*b = r^*(b - B_1 y_0) = r^*r = \|r\|^2,$$

hence

$$e_1^*(B^*B)^{-1} = \frac{1}{\|r\|^2} \begin{pmatrix} 1 & -y_0^* \end{pmatrix}.$$

At last multiply by B^* on the right,

$$e_1^* B^\dagger = \frac{1}{\|r\|^2}(b^* - y_0^* B_1^*) = \frac{-r^*}{\|r\|^2}.$$

Therefore $\|e_1^* B^\dagger\| = 1/\|r\|$. □

Next we determine the norm of the first row of the pseudo-inverse of the Krylov matrix for a linear system whose matrix has a maximally defective eigenvalue.

LEMMA 7.2. *Let*

$$Ax = b \qquad \text{where} \quad A = \begin{pmatrix} 1 & \eta & & \\ & 1 & \ddots & \\ & & \ddots & \eta \\ & & & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

and $\eta \neq 0$. *Then*

$$\|e_1^* K_{i+1}^\dagger\| = \sqrt{\frac{|\eta|^{2i}}{1 + |\eta| + \ldots + |\eta|^{2i}}}.$$

Proof. Let A be of order n. An induction shows that

$$b=\begin{pmatrix}0_{n-1}\\ 1\end{pmatrix}, \qquad Ab=\begin{pmatrix}0_{n-2}\\ \eta\\ 1\end{pmatrix}, \qquad A^ib=\begin{pmatrix}0_{n-(i+1)}\\ \eta^i\\ \alpha_{i,i-1}\eta^{i-1}\\ \vdots\\ \alpha_{i1}\eta\\ 1\end{pmatrix}, \quad i\geq 2,$$

where

$$\begin{aligned}\alpha_{21}&=2\\ \alpha_{i1}&=1+\alpha_{i-1,1}, \qquad 2\leq i\\ \alpha_{ij}&=\alpha_{i-1,j-1}+\alpha_{i-1,j}, \qquad 2\leq i, \quad 2\leq j\leq i-2\\ \alpha_{i,i-1}&=\alpha_{i-1,i-2}+1, \qquad 2\leq i.\end{aligned}$$

Therefore[3] $\alpha_{ij}=\binom{i}{j}$. Hence

$$K_{i+1}=(\,b \quad Ab \quad \ldots \quad A^ib\,)=\begin{pmatrix}0\\ Z\end{pmatrix},$$

where

$$Z=\begin{pmatrix}0 & 0 & 0 & \iddots & \eta^i\\ 0 & 0 & 0 & \iddots & \alpha_{i,i-1}\eta^{i-1}\\ 0 & 0 & \eta^2 & \iddots & \vdots\\ 0 & \eta & \alpha_{21}\eta & \ldots & \alpha_{i1}\eta\\ 1 & 1 & 1 & \ldots & 1\end{pmatrix}$$

is of order $i+1$. Since $\eta\neq 0$, Z is non-singular and $K_{i+1}^\dagger=(\,0 \quad Z^{-1}\,)$. Hence

$$\|e_1^*K_{i+1}^\dagger\|=\|e_1^*Z^{-1}\|.$$

One can decompose $Z=DPR$, where

$$D=\begin{pmatrix}\eta^i & & & \\ & \ddots & & \\ & & \eta & \\ & & & 1\end{pmatrix}, \qquad P=\begin{pmatrix} & & 1\\ & \iddots & \\ 1 & & \end{pmatrix},$$

and

$$R=\begin{pmatrix}1 & 1 & 1 & \ldots & 1\\ & 1 & \alpha_{21} & \ldots & \alpha_{i1}\\ & & \ddots & \ddots & \vdots\\ & & & \ddots & \alpha_{i,i-1}\\ & & & & 1\end{pmatrix}.$$

[3] We thank David Rosnick for this observation.

Hence $e_1^* Z^{-1} = e_1^* R^{-1} P^T D^{-1}$. Another induction using the above recurrences for α_{ij} shows that

$$e_1^* R^{-1} = (\, 1 \quad -1 \quad \dots \quad (-1)^{i+1}\,)\,.$$

This implies

$$\begin{aligned} e_1^* Z^{-1} &= (\,(-1)^{i+1} \quad \dots \quad -1 \quad 1\,) \begin{pmatrix} \eta^{-i} & & & \\ & \ddots & & \\ & & \eta^{-1} & \\ & & & 1 \end{pmatrix} \\ &= (\,(-1)^{i+1}\eta^{-i} \quad \dots \quad -\eta^{-1} \quad 1\,)\,. \end{aligned}$$

At last

$$\|e_1^* K_{i+1}^\dagger\|^2 = \|e_1^* Z^{-1}\|^2 = \sum_{l=0}^{i} |\eta|^{-2i} = |\eta|^{-2i} \,\left(1 + |\eta|^2 + \dots + |\eta|^{2i}\right)\,.$$

□

Next we relate the first row of a pseudo-inverse to the first row of a submatrix. Although such a relation follows from [12, Theorem 2.8.2.5], the expression below is more convenient for our purposes.

Lemma 7.3. *If $A = (\, B \quad c\,)$ has full column rank then*

$$e_1^* A^\dagger = e_1^* B^\dagger\,(I - P),$$

where

$$P \equiv (1 - c^\dagger B B^\dagger c)^{-1}\, c c^\dagger (I - B B^\dagger)$$

is an oblique projector.

Proof. If A has full column rank then

$$e_1^* A^\dagger = e_1^* (A^* A)^{-1} A^*.$$

The partitioning of A gives

$$A^* A = \begin{pmatrix} B^* B & B^* c \\ c^* B & c^* c \end{pmatrix}.$$

Since $A^* A$ is Hermitian, the first row of the inverse is [9, (4)]

$$e_1^* (A^* A)^{-1} = e_1^* S^{-1}\, (\, I \quad -(c^* c)^{-1} B^* c\,)\,,$$

where S is the Schur complement

$$S \equiv B^* B - B^* c\,(c^* c)^{-1}\, c^* B = B^* B - B^* c c^\dagger B.$$

Hence the first row of the pseudo-inverse can be expressed as

$$\begin{aligned} e_1 A^\dagger &= e_1^* S^{-1}\, (\, I \quad -(c^* c)^{-1} B^* c\,) \begin{pmatrix} B^* \\ c^* \end{pmatrix} \\ &= e_1^* S^{-1} (B^* - B^* c (c c^*)^{-1} c^*) = e_1^* S^{-1} B^* (I - c c^\dagger). \end{aligned}$$

Apply the Sherman-Morrison Woodbury formula [24, (2.1.4)] to S^{-1},

$$\begin{aligned} S^{-1} &= (B^*B)^{-1} + (B^*B)^{-1}B^*c\,(1 - c^\dagger B(B^*B)^{-1}B^*c)^{-1}\,c^\dagger B(B^*B)^{-1} \\ &= (B^*B)^{-1} + \alpha^{-1}\,B^\dagger cc^\dagger B(B^*B)^{-1}, \end{aligned}$$

where

$$\alpha \equiv 1 - c^\dagger B(B^*B)^{-1}B^*c = 1 - c^\dagger BB^\dagger c.$$

Since A has full column rank, its columns are linearly independent and c does not lie in range(B). Hence $\alpha \neq 0$.

Substitute the expression for S^{-1} into the expression for $e_1^*A^\dagger$,

$$\begin{aligned} e_1^*A^\dagger &= e_1^*\left[(B^*B)^{-1} + \alpha^{-1}\,B^\dagger cc^\dagger B(B^*B)^{-1}\right]B^*(I - cc^\dagger) \\ &= e_1^*B^\dagger\,(I + \alpha^{-1}cc^\dagger BB^\dagger)(I - cc^\dagger). \end{aligned}$$

Multiply out the last two factors

$$\begin{aligned} &(I + \alpha^{-1}cc^\dagger BB^\dagger)(I - cc^\dagger) = I + \alpha^{-1}cc^\dagger BB^\dagger - cc^\dagger - \alpha^{-1}c\,c^\dagger BB^\dagger c\,c^\dagger = \\ &I + \alpha^{-1}cc^\dagger BB^\dagger - cc^\dagger - \alpha^{-1}(1-\alpha)cc^\dagger = I + \alpha^{-1}cc^\dagger BB^\dagger - \alpha^{-1}cc^\dagger = \\ &I - \alpha^{-1}cc^\dagger(I - BB^\dagger) = I - P. \end{aligned}$$

Since $P^2 = P$, P is a projector. □

As a consequence one can relate the first rows of the pseudo-inverse of a matrix to that of a submatrix. The angle $0 < \theta \leq \pi/2$ below is the largest principal angle between two subspaces [24, §12.4.3].

LEMMA 7.4. *If* $A = (B \quad c)$ *has full column rank then*

$$\|e_1^*A^\dagger\| \leq \|e_1^*B^\dagger\|\,\frac{1}{\sin\theta},$$

where $0 < \theta \leq \pi/2$ *is the angle between* range(B) *and* c*, and*

$$\sin\theta = \|(I - BB^\dagger)c\|/\|c\|.$$

Proof. The previous lemma implies $e_1^*A^\dagger = e_1^*B^\dagger(I - P)$, where

$$P = \alpha^{-1}\,cc^\dagger(I - BB^\dagger), \qquad \alpha = 1 - c^\dagger BB^\dagger c.$$

Thus

$$\|e_1^*A^\dagger\| \leq \|e_1^*B^\dagger\|\,\|I - P\|.$$

Since A has full column rank, $c \neq 0$ and $B \neq I$, and P has rank one. Hence [37, Corollary 5.3]

$$\|I - P\| = \|P\| = \|\alpha^{-1}c\|\,\|(I - BB^\dagger)c^\dagger\| = |\alpha|^{-1}\|(I - BB^\dagger)c\|/\|c\|.$$

But $I - BB^\dagger$ is the orthogonal projector onto the orthogonal complement of range(B). Hence [10, §6], [11, p 10]

$$\|(I - BB^\dagger)c\|/\|c\| = \sin\theta,$$

and

$$\|I - P\| = |\alpha|^{-1} \sin\theta.$$

At last

$$\alpha = 1 - \frac{c^* BB^\dagger c}{c^* c} = \frac{c^*(I - BB^\dagger)c}{c^* c}$$

implies

$$|\alpha| = \left(\frac{\|(I - BB^\dagger)c\|}{\|c\|}\right)^2 = (\sin\theta)^2.$$

Therefore $\|I - P\| = 1/\sin\theta$. □

Below we select the most linearly independent rows from a matrix of full column rank.

LEMMA 7.5. *Let B be a $(k+m) \times k$ matrix of full column rank. If P is a permutation matrix such that*

$$PB = \begin{array}{c} \\ k \\ m \end{array} \begin{array}{c} k \\ \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \end{array}$$

where $|\det(B_1)|$ is maximal[4] *then $\|B_2 B_1^{-1}\| \leq \sqrt{km}$.*

Proof. This proof was inspired by the existence proof of a strong rank revealing QR decomposition [30, Lemma 3.1].

Choose a permutation matrix P so that $|\det(B_1)| \geq |\det(\tilde{B}_1)|$, where

$$\tilde{P}B = \begin{pmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{pmatrix},$$

and $\tilde{P}$ is any permutation. That is, P permutes the rows of B such that the k leading rows have a determinant of maximal magnitude among all sets of k rows of B.

Since B has full row rank, B_1 is non-singular. Therefore one can apply Cramer's rule [34, §0.8.3] to the linear system with matrix B_1^* and right-hand side $B_2^* e_j$. Let $\tilde{B}_1$ be the matrix obtained by replacing row l of B_1 with row j of B_2. Then

$$|(B_2 B_1^{-1})_{jl}| = \frac{|\det(\tilde{B}_1)|}{|\det(B_1)|} \leq 1$$

since $|\det(B_1)|$ is maximal. This implies for the norm of the entire matrix [24, (2.3.8)] $\|B_2 B_1^{-1}\| \leq \sqrt{km}$. □

We use the previous result to bound the norm of the first row of a pseudo-inverse of a full column rank matrix.

LEMMA 7.6. *Let B be a $n \times k$ matrix of full column rank. If P is a permutation matrix such that*

$$PB = \begin{array}{c} \\ k \\ n-k \end{array} \begin{array}{c} k \\ \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \end{array}$$

[4] This means, P permutes the rows of B such that, among all sets of k rows of B, the k leading rows have a determinant of maximal magnitude.

where $|\det(B_1)|$ *is maximal then*

$$\frac{1}{\sqrt{k(n-k+1)}}\,\|e_1^* B_1^{-1}\| \leq \|e_1^* B^\dagger\| \leq \|e_1^* B_1^{-1}\|.$$

Proof. Writing

$$B = P^* \begin{pmatrix} I \\ Z \end{pmatrix} B_1, \qquad Z \equiv B_2 B_1^{-1}$$

gives

$$B^\dagger = (B^*B)^{-1}B^* = B_1^{-1}\,(I+Z^*Z)^{-1}\,(\,I \quad Z^*\,)\,P = B_1^{-1} \begin{pmatrix} I \\ Z \end{pmatrix}^\dagger P.$$

Thus

$$\|e_1^* B^\dagger\| = \|e_1^* B_1^{-1} \begin{pmatrix} I \\ Z \end{pmatrix}^\dagger \|.$$

Lower and upper bounds for $\|e_1^* B^\dagger\|$ in terms of $\|e_1^* B_1^{-1}\|$ are

$$\|e_1^* B_1^{-1}\|/\| \begin{pmatrix} I \\ Z \end{pmatrix} \| \leq \|e_1^* B^\dagger\| \leq \|e_1^* B_1^{-1}\|\,\| \begin{pmatrix} I \\ Z \end{pmatrix}^\dagger \|.$$

Since the singular values of $\begin{pmatrix} I \\ Z \end{pmatrix}$ are [25, §2]

$$\sigma_i \begin{pmatrix} I \\ Z \end{pmatrix} = \sqrt{1+\sigma_i^2(Z)}$$

we get for the upper bound

$$\| \begin{pmatrix} I \\ Z \end{pmatrix}^\dagger \| = \frac{1}{\sigma_{min} \begin{pmatrix} I \\ Z \end{pmatrix}} = \frac{1}{\sqrt{1+\sigma_{min}(Z)^2}} \leq 1,$$

where $\sigma_{min}(\cdot)$ denotes the smallest singular value of a matrix, and for the lower bound

$$\| \begin{pmatrix} I \\ Z \end{pmatrix} \| = \sqrt{1+\|Z\|^2}.$$

Consequently the important norm is bracketed by

$$\frac{\|e_1^* B_1^{-1}\|}{\sqrt{1+\|Z\|^2}} \leq \|e_1^* B^\dagger\| \leq \|e_1^* B_1^{-1}\|.$$

According to Lemma 7.5, choosing a permutation matrix P so that $|\det(B_1)|$ is maximal leads to $\|Z\| \leq \sqrt{(n-k)k}$. Since $k \geq 1$, $1+\|Z\|^2 \leq k(n-k+1)$. □

Now we give expressions for the elements in the first row of the inverse of a Vandermonde matrix.

LEMMA 7.7. *Let*

$$V = \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ \vdots & & & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{pmatrix}$$

be a Vandermonde matrix of order n *with distinct nodes* λ_j, $1 \leq j \leq n$. *Then*

$$|(V^{-1})_{1j}| = \prod_{k=1, k\neq j}^{n} \frac{|\lambda_k|}{|\lambda_k - \lambda_j|}.$$

Proof. Denote the first row of the inverse by $w^* \equiv e_1^* V^{-1}$. Instead of using the known expressions for the elements of an inverse Vandermonde matrix [20, §3], [40, §4], we apply Cramer's rule [34, §0.8.3] to $V^* w = e_1$ because the resulting expressions are simpler. Let $V^{(j)}$ be the matrix obtained by replacing row j of V by e_1^*, the first row of the identity matrix. This corresponds to replacing node λ_j by 0. Cramer's rule implies for the elements of w,

$$|w_j| = \frac{|\det(V^{(j)})|}{|\det(V)|}, \qquad 1 \leq j \leq n.$$

The determinant of the Vandermonde V equals [34, §0.9.11]

$$|\det(V)| = \prod_{j=1}^{n} \prod_{k=j+1}^{n} |\lambda_k - \lambda_j|.$$

Consider the case $j = 1$. To eliminate common factors from numerator and denominator in the expression for w_1, we extract the factors containing λ_1 from $\det(V)$,

$$|\det(V)| = \prod_{k=2}^{n} |\lambda_k - \lambda_1| \prod_{j=2}^{n} \prod_{k=j+1}^{n} |\lambda_k - \lambda_j|.$$

Also, since $V^{(1)}$ contains 0 instead of λ_1,

$$|\det(V^{(1)})| = \prod_{k=2}^{n} |\lambda_k| \prod_{j=2}^{n} \prod_{k=j+1}^{n} |\lambda_k - \lambda_j|.$$

Hence

$$|w_1| = \frac{|\det(V^{(1)})|}{|\det(V)|} = \prod_{k=2}^{n} \frac{|\lambda_k|}{|\lambda_k - \lambda_1|}.$$

Permuting the rows of W and arguing in a similar way yields expressions for the remaining elements of w. □

REFERENCES

[1] O. AXELSSON, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.
[2] P. BROWN, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Stat. Comput., 12 (1991), pp. 58–78.

[3] P. Brown and H. Walker, *GMRES on (nearly) singular systems*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 37–51.
[4] S. Campbell, I. Ipsen, C. Kelley, and C. Meyer, *GMRES and the minimal polynomial*, BIT, 36 (1996), pp. 664–75.
[5] Z. Cao, *A note on the convergence behaviour of GMRES*, Appl. Numer. Math., 25 (1997), pp. 13–20.
[6] S. Chandrasekaran and I. Ipsen, *Perturbation theory for the solution of systems of linear equations*, Research Report 866, Department of Computer Science, Yale University, 1991.
[7] ———, *On the sensitivity of solution components in linear systems of equations*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 93–112.
[8] A. Córdova, W. Gautschi, and S. Ruscheweyh, *Vandermonde matrices on the circle: Spectral properties and conditioning*, Numer. Math., 57 (1990), pp. 577–91.
[9] R. Cottle, *Manifestations of the Schur complement*, Linear Algebra Appl., 8 (1974), pp. 189–211.
[10] C. Davis and W. Kahan, *Some new bounds on perturbation of subspaces*, Bull. Amer. Math. Soc., 75 (1969), pp. 863–8.
[11] C. Davis and W. Kahan, *The rotation of eigenvectors by a perturbation, III*, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.
[12] P. Davis, *Circulant Matrices*, John Wiley & Sons, New York, 1979.
[13] J. Delosme and I. Ipsen, *From Bareiss' algorithm to the stable computation of partial correlations*, Journal of Computational and Applied Mathematics, 27 (1989), pp. 53–91. Also in: *Parallel Algorithms for Numerical Linear Algebra (Advances in Parallel Computing, 1)*, H. van der Vorst and P. Van Dooren, eds., North Holland, 1990.
[14] V. Faber, W. Joubert, E. Knill, and T. Manteuffel, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl, 17 (1996), pp. 707–29.
[15] D. Fasino and G. Inglese, *On the spectral condition of rectangular Vandermonde matrices*, Calcolo, 29 (1993), pp. 291–300.
[16] R. Fletcher, *Conjugate Gradient Method for Indefinite Systems*, Springer Verlag, Berlin, 1976, pp. 73–89.
[17] R. Freund, *A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems*, SIAM J. Sci. Comput., 14 (1993), pp. 470–82.
[18] R. Freund, G. Golub, and N. Nachtigal, *Iterative solution of linear systems*, in Acta Numerica 1992, Cambridge University Press, 1992, pp. 57–100.
[19] R. Freund and N. Nachtigal, *QMR: A quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–39.
[20] W. Gautschi, *On inverses of Vandermonde and confluent Vandermonde matrices*, Numer. Math., 4 (1962), pp. 117–23.
[21] ———, *Norm estimates for inverses of Vandermonde matrices*, Numer. Math., 23 (1975), pp. 337–47.
[22] ———, *Optimally conditioned Vandermonde matrices*, Numer. Math., 24 (1975), pp. 1–12.
[23] W. Gautschi and G. Inglese, *Lower bounds for the condition number of Vandermonde matrices*, Numer. Math., 52 (1988), pp. 241–50.
[24] G. Golub and C. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1996.
[25] W. Govaerts and J. Pryce, *A singular value inequality for block matrices*, Linear Algebra Appl., 125 (1989), pp. 141–8.
[26] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
[27] A. Greenbaum and L. Gurvits, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–58.
[28] A. Greenbaum, V. Pták, and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–9.
[29] A. Greenbaum and L. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–68.
[30] M. Gu and S. Eisenstat, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–69.
[31] P. Henrici, *Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices*, Numer. Math., 4 (1962), pp. 24–40.
[32] N. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.
[33] M. Hochbruck and C. Lubich, *Error analysis of Krylov methods in a nutshell*, SIAM J. Sci. Comput., 19 (1998), pp. 695–701.
[34] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
[35] ———, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[36] I. Ipsen and C. Meyer, *The idea behind Krylov methods*, Amer. Math. Monthly. To appear.
[37] ———, *The angle between complementary subspaces*, Amer. Math. Monthly, 102 (1995), pp. 904–11.
[38] C. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.
[39] C. Lanczos, *Solution of systems of linear equations by minimzed iteration*, J. Res. Nl. Bur. Stand., 49 (1952), pp. 33–53.
[40] N. Macon and A. Spitzbart, *Inverses of Vandermonde matrices*, Amer. Math. Monthly, 65 (1958), pp. 95–100.
[41] N. Nachtigal, S. Reddy, and L. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–95.
[42] N. Nachtigal, L. Reichel, and L. Trefethen, *A hybrid GMRES algorithm for nonsymmetric linear systems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 796–825.
[43] C. Paige, B. Parlett, and H. Van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra, 2 (1995), pp. 115–33.
[44] C. Paige and M. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–29.
[45] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, 1996.
[46] Y. Saad and M. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM Sci. Stat. Comput., 7 (1986), pp. 856–69.
[47] P. Sonneveld, *CGS, a fast Lanczos-type solver for nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 10 (1989), pp. 36–52.
[48] G. Stewart, *Collinearity and least squares regression*, Statist. Sci., 2 (1987), pp. 68–100.
[49] K. Toh, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–6.
[50] E. Tyrtyshnikov, *How bad are Hankel matrices?*, Numer. Math., 67 (1994), pp. 261–9.
[51] H. Van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for nonsymmetric linear systems*, SIAM J. Sci. Comput., 13 (1992), pp. 631–44.
[52] H. Van der Vorst and C. Vuik, *The superlinear convergence behaviour of GMRES*, J. Comput. Appl. Math., 48 (1993), pp. 327–41.
[53] H. Walker and L. Zhou, *A simpler GMRES*, Numer. Linear Algebra Appl., 1 (1994), pp. 571–81.