MATRICES THAT GENERATE THE SAME KRYLOV RESIDUAL SPACES

ANNE GREENBAUM* AND ZDENEK STRAKOS†

Abstract. Given an n by n nonsingular matrix A and an n-vector v, we consider the spaces of the form $AK_k(A,v)$, $k = 1,...,n$, where $K_k(A,v)$ is the k^{th} Krylov space, equal to $span\{v, Av, ..., A^{k-1}v\}$. We characterize the set of matrices B that, with the given vector v, generate the same spaces; i.e., those matrices B for which $BK_k(B,v) = AK_k(A,v)$, for all $k = 1,...,n$. It is shown that any such sequence of spaces can be generated by a unitary matrix. If zero is outside the field of values of A, then there is a Hermitian positive definite matrix that generates the same spaces, and, moreover, if A is close to Hermitian then there is a nearby Hermitian matrix that generates the same spaces. It is also shown that any such sequence of spaces can be generated by a matrix having any desired eigenvalues.

Implications about the convergence rate of the GMRES method are discussed. A new proof is given that if zero is outside the field of values of A, then convergence of the GMRES algorithm is strictly monotonic. It is shown that if A is close to Hermitian, then the eigenvalues of A essentially determine the behavior of the GMRES iteration but that, in general, eigenvalue information alone is never sufficient to ensure rapid convergence of the GMRES algorithm.

1. Introduction.

Most iterative methods for solving a linear system

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

start with an initial guess x^0 for the solution and, at each step k, generate an approximate solution x^k from the linear variety

$$x^0 + span\{r^0, Ar^0, ..., A^{k-1}r^0\}, \tag{1.2}$$

where $r^0 \equiv b - Ax^0$ is the initial residual. Several of these methods – e.g., (full) GMRES[1] [16], ORTHOMIN [21], and ORTHODIR [22] – choose the approximation x^k to minimize the Euclidean norm of the residual $r^k \equiv b - Ax^k$. A question of considerable interest is how accurate an approximation x^k can be obtained from this linear variety.

---

* Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012. This work was supported in part by DOE contract DEFG0288ER25053. Part of this work was performed while visiting IBM T.J. Watson Research Center and part was performed while visiting the Institute for Mathematics and its Applications, University of Minnesota.

† Institute of Computer and Information Sciences, Czech Academy of Sciences, Pod vod vezi 2, 18207 Praha 8, Czechoslovakia. This work was performed while visiting the Institute for Mathematics and its Applications, University of Minnesota.

[1] Throughout this paper we will refer to the GMRES algorithm, but since we always assume exact arithmetic, our statements will apply equally well to any mathematically equivalent formulation.

The residual vector r^k is the vector of smallest 2-norm in the linear variety

$$r^0 + span\{Ar^0, ..., A^k r^0\}. \tag{1.3}$$

It follows that r^k can be written in the form

$$r^k = P_k(A)r^0, \tag{1.4}$$

where P_k is the k^{th} degree polynomial with value one at the origin that minimizes $\|p_k(A)r^0\|$ over the set $\mathcal{P}_k$ of all k^{th} degree polynomials p_k with $p_k(0) = 1$.

A simple bound on the size of the residual can be derived as follows. If we assume that the matrix A has a complete set of eigenvectors Z, so that A can be written in the form $A = Z\Lambda Z^{-1}$, where $\Lambda = diag(\lambda_1, ..., \lambda_n)$ is the diagonal matrix of eigenvalues, then r^k can be written in the form

$$r^k = ZP_k(\Lambda)Z^{-1}r^0. \tag{1.5}$$

For any other k^{th} degree polynomial $p_k \epsilon \mathcal{P}_k$, we have

$$\|r^k\| \leq \|Zp_k(\Lambda)Z^{-1}r^0\| \leq \kappa(Z) \max_{i=1,...,n} |p_k(\lambda_i)| \, \|r^0\|, \tag{1.6}$$

where $\kappa(Z) = \|Z\| \, \|Z^{-1}\|$ is the condition number of Z and $\|\cdot\|$ here, and elsewhere, denotes the 2-norm for vectors and the corresponding spectral norm for matrices. The bound in (1.6) is minimized by taking p_k to be the k^{th} degree minimax polynomial on the discrete set of eigenvalues $\{\lambda_1, ..., \lambda_n\}$, and (1.6) can be written in the form

$$\frac{\|r^k\|}{\|r^0\|} \leq \kappa(Z) \min_{p_k \epsilon \mathcal{P}_k} \max_{i=1,...,n} |p_k(\lambda_i)|. \tag{1.7}$$

The columns of the matrix Z can be normalized arbitrarily in order to make $\kappa(Z)$ as small as possible. Throughout this paper, however, we will assume that the columns of Z have norm 1. This scaling usually gives a near optimal value for $\kappa(Z)$.

If the matrix A is *normal*, then $\kappa(Z)$ is one, and it was recently shown in several different ways [8,9,11] that this bound is *sharp*; i.e., that for each k, there is an initial vector r^0 (depending on k) for which equality holds in (1.7). In many cases of interest, however, the matrix A is not normal and the factor $\kappa(Z)$ in (1.7) may be quite large. (See, for instance, [15] for

some interesting physical examples.) In such cases, the bound (1.7) may be a large overestimate of the actual residual.

To deal with such problems, Trefethen has introduced bounds based on the *ϵ-pseudo-eigenvalues* of the matrix [18]. There are two equivalent definitions of the ϵ-pseudo-spectrum of a matrix A:

Definition 1. The ϵ-pseudo-spectrum of A, denoted $\Lambda_\epsilon(A)$, is the set of points $z \in C$ such that $\|(zI - A)^{-1}\| \geq \epsilon^{-1}$.

Definition 2. The ϵ-pseudo-spectrum of A is the set of points $z \in C$ which are eigenvalues of some matrix $A + E$ with $\|E\| \leq \epsilon$.

For any analytic function f, it is known [2, p. 560] that $f(A)$ can be written in the form

$$f(A) = \frac{1}{2\pi i} \int_\Gamma (zI - A)^{-1} f(z) dz, \tag{1.8}$$

where Γ is any simple closed curve or union of simple closed curves containing the spectrum of A. Taking norms on both sides and replacing the norm of the integral by the length $\mathcal{L}(\Gamma)$ of the curve times the maximum norm of the integrand factors gives

$$\|f(A)\| \leq \frac{1}{2\pi} \mathcal{L}(\Gamma) \max_{z \in \Gamma} \|(zI - A)^{-1}\| \max_{z \in \Gamma} |f(z)|. \tag{1.9}$$

Note that the inequality (1.9) will be close to an equality if and only if the integrand in (1.8) is nearly constant in norm over the curve Γ and cancellation does not cause the norm of the integral to be much smaller than the length of the curve times the norm of the integrand.

A reasonable set of curves to consider, then, are the curves Γ_ϵ on which the resolvent $(zI - A)^{-1}$ has constant norm $1/\epsilon$; i.e., the boundaries of the ϵ-pseudo-spectra of A. In this case, the bound (1.9) becomes

$$\|f(A)\| \leq \frac{\mathcal{L}(\Gamma_\epsilon)}{2\pi\epsilon} \max_{z \in \Gamma_\epsilon} |f(z)|. \tag{1.10}$$

Taking the function f to be the polynomial $p_k \in \mathcal{P}_k$ for which this bound is minimal gives the following bound on the residual in (1.4):

$$\frac{\|r^k\|}{\|r^0\|} \leq \frac{\mathcal{L}(\Gamma_\epsilon)}{2\pi\epsilon} \min_{p_k \in \mathcal{P}_k} \max_{z \in \Gamma_\epsilon} |p_k(z)|. \tag{1.11}$$

While, for properly chosen values of ϵ, the bound (1.11) sometimes gives a much better estimate than (1.7) of the actual size of the residual, there are other cases in which even the bound (1.11) is a large overestimate, for any value of ϵ. Consider, for example, a matrix A of the form $Z\Lambda Z^{-1}$, where

$$Z = \begin{pmatrix} 1 & \sqrt{1-\delta} & 0 & \dots & 0 \\ 0 & \sqrt{\delta} & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \quad \delta << 1, \quad \Lambda = \begin{pmatrix} 20 & & & & \\ & 10 & & & \\ & & 5 & & \\ & & & . & \\ & & & & 1 \end{pmatrix}. \tag{1.12}$$

A has two large, well-separated eigenvalues, 20 and 10, corresponding to an ill-conditioned block in Z. The remaining eigenvalues are uniformly distributed in the interval $[1, 5]$. The condition number of Z is approximately $2/\sqrt{\delta}$.

Fig. 1a shows the ϵ-pseudo-spectra of such a matrix A, of order $n = 19$ and with $\delta = 10^{-8}$, for $\epsilon = 10^{-5}$ and 10^{-4}. (Note that the scales on the horizontal and vertical axes in the figure are different. The pseudo-spectral regions are approximately circular.) Note that the factor $\frac{\mathcal{L}(\Gamma_\epsilon)}{2\pi\epsilon}$ in (1.11) is on the order of $2 \cdot 10^4$ for these values of ϵ, and a low degree polynomial p_k could not be small enough on these curves to make the bound (1.11) reasonable; i.e., less than 1 for $k << 19$. For larger values of ϵ ($\epsilon >\sim 10^{-2}$), the ϵ-pseudo-spectrum contains the origin, and then, because of the maximum modulus theorem, the bound (1.11) cannot be useful since $p_k(0) = 1$.

Yet for this problem the GMRES iteration will behave almost as it would if applied to the diagonal matrix Λ. It will annihilate the top 2 by 2 block quickly, just as it would annihilate the two large eigenvalues 20 and 10 of Λ quickly, and then converge as if it were working with a diagonal matrix whose eigenvalues lie between 1 and 5. Fig. 1b shows the convergence of the GMRES method applied to a linear system with coefficient matrix A (solid line) and to one with coefficient matrix Λ (dashed line), with a random initial residual. The + marks in Fig. 1b show a *sharp* upper bound on the residual at each step of the GMRES algorithm applied to A. (This is also an upper bound on the residual at each step of the GMRES algorithm applied to Λ, and it is very similar to the sharp upper bound for this problem.) This was computed using an optimization code (*fminu* in *MATLAB* [13]) to find, for each k, an initial residual r^0 for which $\|r^k\|/\|r^0\|$ was as large as possible. These bounds were checked in the following way. Taking the initial vector r^0, with $\|r^0\| = 1$, returned by the optimization routine, we computed the GMRES polynomial $P_k(A)$ in (1.4). If $\|P_k(A)\|$ is equal to $\|r^k\|$, then this must be a sharp upper bound, since it is attainable and since, for any other initial residual $\hat{r}^0$, we have $\|\hat{r}^k\| \leq \|P_k(A)\| \, \|\hat{r}^0\|$.

Although it often required carefully chosen initial guesses to enable the optimization code to find the worst initial residual, the bounds plotted in Fig. 1b were all proven correct in this way. The bound (1.11) based on the ϵ-pseudo-spectra and the bound (1.7) based on the condition number of the eigenvector matrix are both *much* larger than the sharp bound pictured in Fig. 1b.

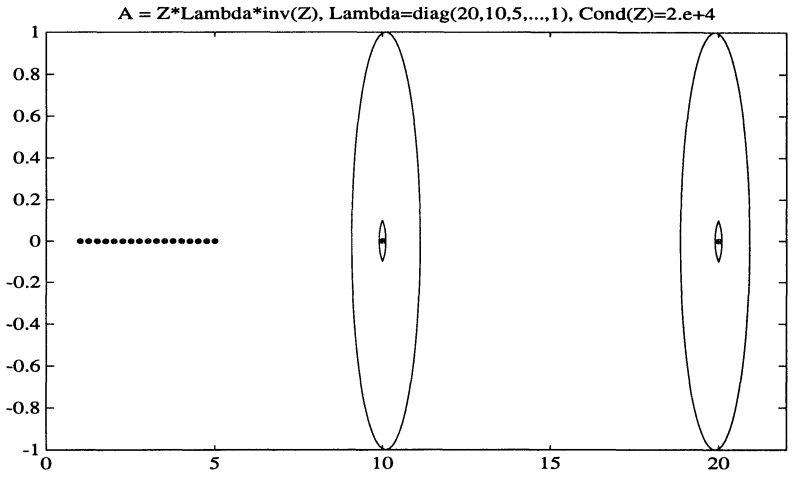

Fig. 1a. 1.e-5 and 1.e-4 Pseudo-spectra

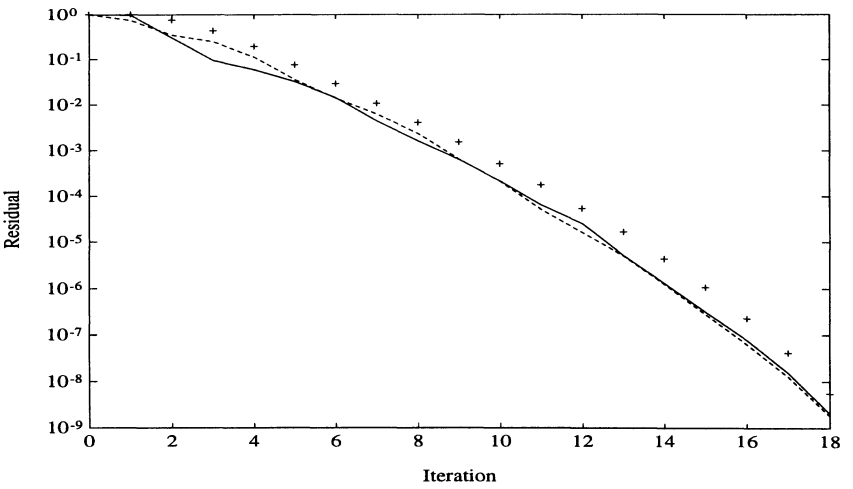

Fig. 1b. GMRES(A) (solid), GMRES(Lambda)(dashed), Sharp Bound(+)

It should also be noted that if A is reduced to Schur form, then the Frobenius norm of the strict upper triangle of the Schur form is also quite large – in this case about 10^5. Thus, the matrix A is *far* from normal in all of the usual senses, and yet the GMRES iteration converges very well.

The problem of deriving a sharp bound on the convergence rate of the GMRES method, in terms of eigenvalues or pseudo-eigenvalues or other simple properties of the matrix, appears to be a difficult one. In this paper, we take a different approach to this problem. While it does not yet achieve the ultimate goal of giving a sharp or nearly sharp bound for all cases, it does yield certain insights into the behavior of the GMRES method when applied to linear systems with non-normal coefficient matrices. We relate the behavior of the method applied to a non-normal matrix to its behavior when applied to certain normal matrices. Since the convergence rate of the method applied to normal matrices is determined by the eigenvalues of the matrix, if the eigenvalues of the normal matrix could be related to some properties of the original matrix, then (1.7) would give a bound on the convergence rate, in terms of these properties.

Unfortunately, it is only in special cases that we are able to relate the eigenvalues of the normal matrix to meaningful properties of the original matrix. It is shown that any behavior that can be seen with the GMRES method can be seen with the method applied to a unitary matrix. While certain properties of the original matrix – e.g., positive definiteness – appear to guarantee large gaps in the spectrum of this unitary matrix, we have not been able to prove this. It is shown that if zero is outside the field of values of the original matrix A, then the GMRES method behaves just as it would for a certain Hermitian positive definite matrix. If A is close to Hermitian, then the GMRES method behaves just as it would for a Hermitian matrix whose eigenvalues are close to those of A. In this case, then, the eigenvalues of A essentially determine the behavior of the GMRES algorithm. In general, however, eigenvalue information alone cannot be sufficient to ensure fast convergence of the GMRES algorithm. Any behavior that can be seen with the method can be seen with the method applied to a matrix having *any* nonzero eigenvalues.

Throughout this paper, matrices and vectors will be assumed to be complex, unless otherwise stated. A superscript * will denote the Hermitian transpose. The space spanned by a set of vectors $\{v^1, ..., v^k\}$ will be denoted $[v^1, ..., v^k]$, and the set of vectors of the form v plus a linear combination of $\{v^1, ..., v^k\}$ will be denoted $v + [v^1, ..., v^k]$. The symbol $\|\cdot\|$ will always denote the 2-norm of a vector or matrix and $\|\cdot\|_F$ will denote the Frobenius norm of a matrix. The usual Euclidean inner product will be denoted $< \cdot, \cdot >$.

In section 2 we characterize those matrices $B \equiv B(r^0)$ for which the GMRES method, with initial residual r^0, generates the same residual vectors at each step as the GMRES method applied to a linear system with coefficient matrix A and initial residual r^0. Letting $K_k(A, r^0)$ denote the Krylov space $[r^0, Ar^0, ..., A^{k-1}r^0]$, we know that the residual r^k has the smallest 2-norm among all vectors of the form

$$r^0 + AK_k(A, r^0). \tag{1.13}$$

If the spaces $BK_k(B, r^0)$ are the same as the spaces $AK_k(A, r^0)$, for all $k = 1, ..., n$, then the GMRES method applied to B, with initial residual r^0, will generate the same residual vectors at each step as the method applied to A, with initial residual r^0. For lack of a better name, we refer to these spaces as *Krylov residual spaces*, which should not be confused with the Krylov spaces $K_k(A, r^0)$ and $K_k(B, r^0)$. The Krylov spaces $K_k(A, r^0)$ and $K_k(A+\alpha I, r^0)$, where α is any scalar, are the same for all k, but the GMRES method behaves very differently when applied to A and to $A + \alpha I$! Note also that it is not required that individual vectors, $Ar^0, ..., A^k r^0$ and $Br^0, ..., B^k r^0$ be equal, only that the spaces they span be the same; i.e., Br^0 must be a scalar multiple of Ar^0, $B^2 r^0$ must be a linear combination of $A^2 r^0$ and Ar^0, etc. Note further that if r^k is written in the form (1.4), $r^k = P_k(A)r^0 = \hat{P}_k(B)r^0$, then the polynomials P_k and $\hat{P}_k$ will be different. Thus, this approach cannot be used to analyze eigenvalue approximations generated by the algorithm, since these will be very different for A and for B. In section 3, we consider normal matrices B for which the Krylov residual spaces, with a given initial vector r^0, are the same as those of a given matrix A. In section 4, we show that any sequence of Krylov residual spaces can be generated by a matrix having any nonzero eigenvalues. Section 5 gives conclusions and discusses remaining open questions and possible applications of this technique of analysis.

2. Matrices that generate the same Krylov residual spaces. Given an n by n nonsingular matrix A and an n-vector v, we wish to characterize the matrices $B = B(v)$ for which the spaces

$$(2.1) \quad AK_k(A, v) \equiv [Av, ..., A^k v] \quad \text{and} \quad BK_k(B, v) \equiv [Bv, ..., B^k v]$$

are the same, for all $k = 1, ..., n$. Throughout this paper, we will assume that the vectors $Av, ..., A^n v$ are linearly independent. If they are not, however, or if one wishes to consider conditions in which only, say, the first m of these spaces are the same, then the characterizations given here can be modified accordingly. The class of matrices B for which the first m such spaces are the same will, of course, include the class discussed in this paper.

Let $w^1, ..., w^k$ be an orthonormal basis for $[Av, ..., A^k v]$, and let W be the matrix with orthonormal columns $w^1, ..., w^n$. Then it is well-known that the unitary matrix W satisfies

$$(2.2) \qquad AW = WH,$$

where H is an unreduced upper Hessenberg matrix. The following theorem characterizes all matrices $B = B(v)$ that generate the same Krylov residual spaces (2.1) with the vector v.

THEOREM 2.1. *Using the above notation, let B be of the form*

$$B = WRHW^*, \tag{2.3}$$

where R is any nonsingular upper triangular matrix. Then

$$BK_k(B,v) = AK_k(A,v), \quad k = 1,...,n, \tag{2.4}$$

and, conversely, any matrix B that satisfies (2.4) is of the form (2.3).

Proof. Suppose the order one Krylov residual spaces are the same; i.e.,

$$Bv = cAv \tag{2.5}$$

for some nonzero scalar c. The higher order spaces $BK_k(B,v)$ and $AK_k(A,v)$, $k > 1$, will be the same if B satisfies

$$BW = W\tilde{H} \tag{2.6}$$

for some unreduced upper Hessenberg matrix $\tilde{H}$. To see that this is so, assume that $BK_{k-1}(B,v) = AK_{k-1}(A,v)$. The $(k-1)^{st}$ column of equation (2.6) can be written as

$$B\, w^{k-1} = \sum_{j=1}^{k} w^j\, \tilde{H}_{j,k-1}.$$

It follows that Bw^{k-1}, and hence $B^k v$, is a linear combination of $w^1,...,w^k$, and since the coefficient of w^k is nonzero, $B^k v$ is linearly independent of $[w^1,...,w^{k-1}] = [Bv,...,B^{k-1}v]$. Thus the order k spaces are the same and the proof is by induction. Conversely, it is clear that (2.5) and (2.6) are necessary conditions in order for (2.4) to be satisfied.

Taking B to be of the form $B = W\tilde{H}W^*$, then, the condition (2.5) becomes, upon multiplying each side by W^*,

$$\tilde{H}W^*v = cW^*Av. \tag{2.7}$$

But W^*Av is just a constant times e_1, the first unit vector,

$$W^*Av = \hat{c}e_1,$$

and W^*v satisfies

$$W^*v = W^*A^{-1}Av = W^*(WH^{-1}W^*)Av = \hat{c}H^{-1}e_1.$$

Therefore (2.7) can be written as

$$\tilde{H}H^{-1}e_1 = ce_1. \tag{2.8}$$

Clearly, if $\tilde{H}$ is of the form RH, where R is a nonsingular upper triangular matrix, then $\tilde{H}$ is an unreduced upper Hessenberg matrix satisfying (2.8). Conversely, if $\tilde{H}$ is an unreduced upper Hessenberg matrix satisfying (2.8), and if we write $\tilde{H}$ in the form XH, then the first column of X must be a scalar multiple of e_1. But the requirement that XH be upper Hessenberg then implies that the elements below the diagonal in subsequent columns of X are also zero, and the requirement that XH be unreduced implies that the diagonal elements of X are nonzero. Thus, the only matrices B satisfying (2.4) are of the form $B = W\tilde{H}W^*$, where $\tilde{H} = RH$. □

The characterization of Theorem 1 will prove useful in later sections, where we consider normal or near normal matrices of the form RH, for various nonsingular upper triangular matrices R.

There are many equivalent ways to characterize the matrices B that generate the same Krylov residual spaces. We give two such characterizations which will prove useful in later sections. The following theorem can be derived from Theorem 1, but it is just as easy to prove it directly, which we do below.

THEOREM 2.2. *Using the above notation, suppose B is of the form*

$$B = W\hat{R}\hat{H}W^* \tag{2.9}$$

where $\hat{R}$ is any nonsingular upper triangular matrix and

$$\hat{H} = \begin{pmatrix} 0 & . & . & 0 & 1/ < v, w^n > \\ 1 & & & . & - < v, w^1 > / < v, w^n > \\ & . & & . & . \\ & & . & . & . \\ 0 & . & . & 1 & - < v, w^{n-1} > / < v, w^n > \end{pmatrix}. \tag{2.10}$$

Then (2.4) holds, and, conversely, any matrix B that satisfies (2.4) is of the form (2.9, 2.10).

Proof. Let V_A be the matrix with columns $(Av, ..., A^n v)$. The condition (2.4) is equivalent to

$$V_A R_1 = B(v, w^1, ..., w^{n-1}), \tag{2.11}$$

for some nonsingular upper triangular matrix R_1. To see that this is so, note that (2.11) implies that Bv is a nonzero multiple of Av (and hence of w^1), that Bw^1 (and hence B^2v) is a linear combination of A^2v and Av, with

the coefficient of A^2v being nonzero, etc. We can also write $V_A = WR_2$, for some nonsingular upper triangular matrix R_2, and

$$(2.12) \qquad B(v, w^1, ..., w^{n-1}) = BW \begin{pmatrix} < v, w^1 > & 1 & 0 & . & 0 \\ < v, w^2 > & 0 & 1 & & . \\ . & . & & . & . \\ . & . & & & 1 \\ < v, w^n > & 0 & . & . & 0 \end{pmatrix}$$

Substituting these expressions into (2.11) and solving for B gives

$$B = W\hat{R}\hat{H}W^*,$$

where $\hat{R} = R_2R_1$ and $\hat{H}$ is the inverse of the rightmost matrix in (2.12). It is easy to check that this is the matrix $\hat{H}$ defined in (2.10). Since, by assumption, the vectors $Av, ..., A^nv$ are linearly independent, it is also the case that $v, Av, ..., A^{n-1}v$ are linearly independent and hence that $< v, w^n >$ is nonzero. Thus, the matrix $\hat{H}$ is well-defined and the theorem is proved. □

This theorem will be used later to show that any sequence of Krylov residual spaces can be generated by a matrix having any nonzero eigenvalues.

Another characterization of the class of matrices described in Theorems 1 and 2 is given in the following theorem. Let $\mathcal{L}(X)$ denote the lower triangle (including diagonal) of a matrix X, and let $\hat{\mathcal{L}}(X)$ denote the strict lower triangle of X. The following theorem characterizes the matrices X for which HX is upper triangular.

THEOREM 2.3. *If H is an n by n nonsingular unreduced upper Hessenberg matrix, then HX is upper triangular if and only if the elements of the lower triangle of X satisfy*

$$(2.13) \qquad X_{i-1,j} = -\frac{1}{H_{i,i-1}} \sum_{k=i}^{n} H_{ik}X_{kj}, \quad i = n, n-1, ..., j+1, \quad j = 1, ..., n,$$

where the bottom row of elements X_{nj}, $j = 1, ..., n$ is arbitrary. These equations are satisfied by any matrix X whose lower triangle is of the form

$$(2.14) \qquad \mathcal{L}(X) = \mathcal{L}(H^{-1})\, D, \qquad D \text{ diagonal},$$

and if $\mathcal{L}(H^{-1})$ has no zero columns, then all solutions to (2.13) are of the form (2.14).

Proof. The condition that HX be upper triangular means that the elements of the lower triangle of X must satisfy

$$(HX)_{ij} = \sum_{k=i-1}^{n} H_{ik}X_{kj} = 0, \quad j < i. \tag{2.15}$$

Since $H_{i,i-1}$ is nonzero for all i, we can solve these equations for $X_{i-1,j}$, $i = n, n-1, ..., j+1$, in terms of X_{nj}, to obtain (2.13). If the elements X_{nj} are taken to be of the form $d_j H_{nj}^{-1}$, for some scalars d_j, then, since $H^{-1}D$ also satisfies equations (2.15) when $D = diag(d_1, ..., d_n)$, the lower triangle of X must be equal to the lower triangle of $H^{-1}D$. Moreover, all solutions are of this form unless H^{-1} has a zero element in its last row, and, in this case, the equations (2.15) imply that all elements of that column of $\mathcal{L}(H^{-1})$ must be zero. □

According to Theorem 3, matrices $\tilde{H}$ of the form RH, as described in Theorem 1, are essentially those matrices whose inverses have lower triangles of the form $\mathcal{L}(H^{-1})D$, or, more generally, of the form (2.13). The strict upper triangle of the inverse of such a matrix is arbitrary. In a later section we will consider the Hermitian matrix obtained by taking the strict upper triangle to be the Hermitian transpose of the strict lower triangle.

3. Decompositions of the form $H = RN$. For a given matrix A and vector v, let W and H be defined as in the previous section. In light of Theorem 1, it will be instructive to consider matrices $\tilde{H}$ of the form RH, for some upper triangular matrices R; for if we then define the matrix B to be $W\tilde{H}W^*$, then the Krylov residual spaces generated by B with initial vector v will be the same as those generated by A with initial vector v. Since the GMRES algorithm minimizes the 2-norm of the residual vector over the linear variety consisting of the initial residual plus the Krylov residual space, it follows that if the GMRES method is applied to a linear system $Ax = b$, with initial residual $b - Ax^0 = v$, then it will generate the same residual vectors at each step as the GMRES method applied to a linear system $By = f$, with initial residual $f - By^0 = v$. We will express this by writing

$$GMRES(A, v) \equiv GMRES(B, v). \tag{3.1}$$

If, for each vector v, we can find a matrix B of the given form, for which we can analyze the behavior of the GMRES method applied to B, then we can also analyze the behavior of the GMRES method applied to A. Since the behavior of the GMRES method for normal matrices is well-understood in terms of the eigenvalues of the matrix (cf. equation (1.7)), it is desirable to find an upper triangular matrix R such that RH is normal, or, equivalently, to write H in the form $H = \tilde{R}N$, where $\tilde{R}$ is upper triangular (the inverse of the upper triangular matrix in Theorem 1) and N is normal. Alternatively, using Theorem 3, we will look for normal

matrices N such that the lower triangle of N^{-1} is equal to the product of the lower triangle of H^{-1} and a diagonal matrix.

There are many decompositions of this form, a few of which will be described in this section. If such a decomposition could be found for which the eigenvalues of N could be related to eigenvalues or pseudo-eigenvalues or other simple properties of the matrix A, then the convergence rate of the GMRES method applied to A could be explained in terms of these properties. In general, we have not been able to find such a decomposition, but in special cases this can be done.

3.1. Equivalent Unitary Matrices. Any upper Hessenberg matrix H can be written in the form

$$H = RQ, \tag{3.2}$$

where R is upper triangular and Q is unitary (and also upper Hessenberg). Thus, any behavior that can be seen with the GMRES algorithm applied to any matrix can be seen with the GMRES algorithm applied to a unitary matrix! Which unitary matrix will depend on the initial residual as well as the matrix, but for any matrix A and any initial residual r^0, if W is the matrix whose columns are the orthonormal basis vectors and H the upper Hessenberg matrix generated after n steps of $GMRES(A, r^0)$, and if $H = RQ$ is the RQ-decomposition of H, then

$$GMRES(A, r^0) \equiv GMRES(WQW^*, r^0). \tag{3.3}$$

We have not been able to establish any interesting relationships between the eigenvalues of the unitary matrix $B = WQW^*$ and special properties of A, although it appears that such relationships exist, at least in certain cases. Fig. 2a shows the eigenvalues of the unitary matrix B obtained from a GMRES computation with a random 19 by 19 matrix A and a random initial residual. Note that the eigenvalues are fairly uniformly distributed around the unit circle. The solid line in Fig. 2b shows the convergence of the GMRES algorithm, applied to either A or B, with this same initial residual, while the dotted line in the figure shows the convergence of the GMRES algorithm applied to B with a different random initial residual, uncorrelated with B. Note that the GMRES algorithm behaves similarly, when applied to either A or B, with an arbitrary initial vector, indicating that the *typical* behavior of the algorithm applied to B is similar to that of the algorithm applied to A. In both cases, we see slow convergence, which is characteristic of the GMRES algorithm applied to a random matrix or to a unitary matrix with eigenvalues all around the unit circle.

Also shown in Fig. 2b are sharp upper bounds for the residual at each step of the GMRES algorithm applied to A (+) and to B (x). These

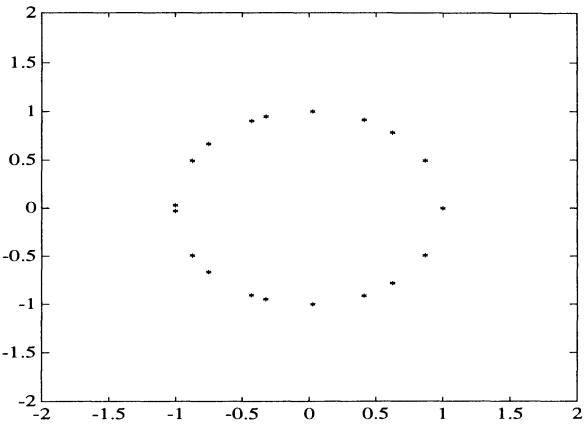

Fig. 2a. Eigenvalues of B. A=rand(19,19)

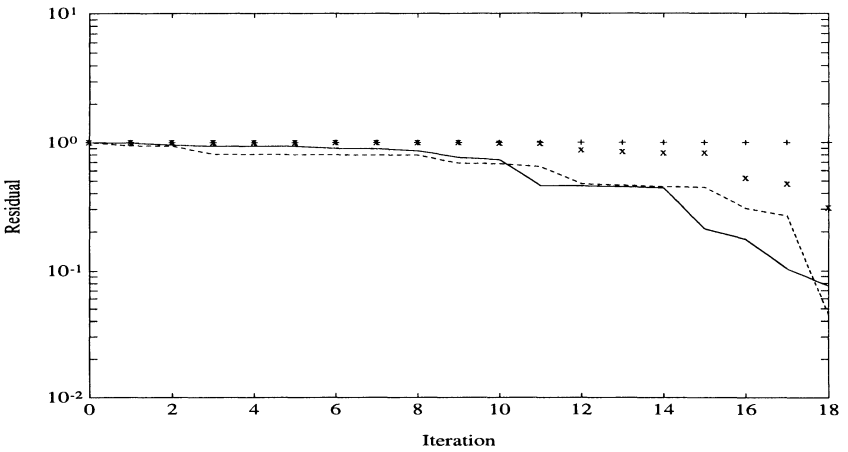

Fig. 2b. GMRMES(A or B,r0)(-) and GMRES(B,random)(–)

were computed numerically and then proven correct, as described earlier. (The initial vector r^0 for which $||r^k||/||r^0||$ is maximal was computed using an optimization code or other technique, and then it was determined that $||r^k|| = ||P_k(A)|| \cdot ||r^0||$, where P_k is the GMRES polynomial.) It is interesting to note that the computed sharp bound for A is 1.0000 for steps 1 through 18 – indicating that there is an initial residual for which the GMRES algorithm applied to A makes no progress at all until step $n = 19$! This is frequently observed to be the case for random matrices. Had we used this initial residual for our first experiment, the equivalent unitary matrix WQW^* would have had its eigenvalues uniformly distributed around the unit circle. This can be seen from the characterization of Theorem 2. In this case, $< r^0, w^j >$ is zero for $j = 1, ..., n-1$ and $| < r^0, w^n > | = ||r^0||$,

which we can assume without loss of generality is 1. Then the matrix $\hat{H}$ defined in Theorem 2 is a unitary shift matrix, and since the RQ decomposition is unique, up to multiplication by a unitary diagonal matrix, the matrix Q in (3.2) must be the product of a unitary diagonal matrix with $\hat{H}$. Matrices of this form have eigenvalues uniformly distributed about the unit circle.

In contrast, Fig. 3a shows the eigenvalues of the unitary matrix B obtained from a GMRES computation with a diagonal matrix A of order $n = 19$ whose eigenvalues are uniformly distributed between 1 and 10, again using a random initial residual. Note the large gaps in the spectrum of this unitary matrix. The solid line in Fig. 3b shows the convergence of the GMRES algorithm, applied to either A or B, with this same initial residual, while the dotted line in the figure shows the convergence of the GMRES algorithm applied to B with a different random initial residual, uncorrelated with B. Again, the GMRES algorithm behaves similarly, when applied to either A or B, with an arbitrary initial vector. Now the convergence is faster, which is characteristic of the GMRES algorithm applied to either a well-conditioned symmetric positive definite matrix or a unitary matrix with large gaps in its spectrum.

Also shown in Fig. 3b are sharp upper bounds for the residual at each step of the GMRES algorithm applied to A (+) and to B (x). Again, the sharp bounds for A and B are similar, and each is not much worse than the typical behavior of the algorithm for that matrix.

Although it seems surprising at first, it is perhaps not totally unexpected that any behavior that can be seen with the GMRES method can be seen with the method applied to a unitary matrix. It is known that the worst possible behavior – no progress at all until step n – can occur with a unitary shift matrix [1,14], (this will be discussed further in section 4), and, of course, the best possible behavior – convergence in one step – occurs with the identity. We have shown that any convergence behavior between these two extremes can also occur with a unitary matrix.

3.2. Equivalent Hermitian Positive Definite Matrices. If zero is outside the field of values of A, then zero will also be outside the field of values of $H = W^*AW$. In this case, H can be factored in the form $H = UL$, where U and L are nonsingular upper and lower triangular matrices, respectively. This decomposition can also be written in the form

$$H = (UL^{-*})(L^*L), \tag{3.4}$$

where the first factor UL^{-*} is upper triangular and the second is Hermitian positive definite (and also tridiagonal). Thus, in this case, the GMRES method applied to A behaves just as it does when applied to a certain Hermitian positive definite matrix!

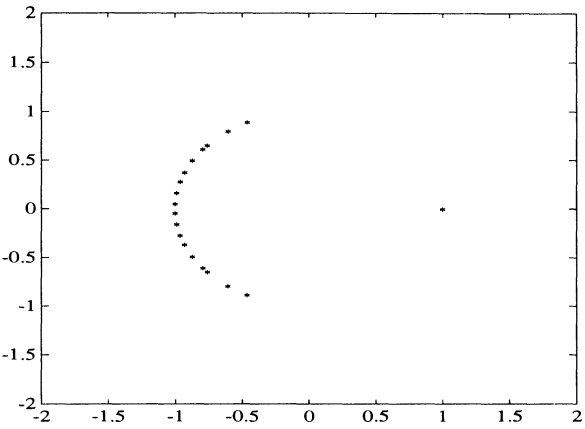

Fig. 3a. Eigenvalues of B. A=diag(1,1.5,2, ...,10)

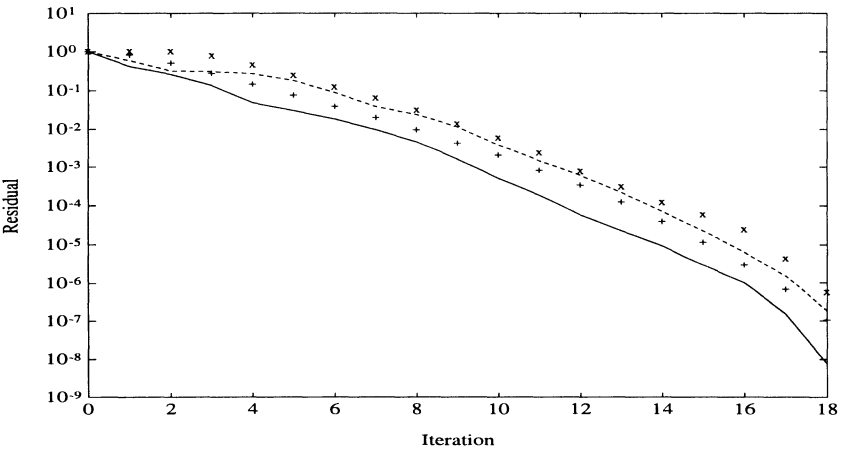

Fig. 3b. GMRMES(A or B,r0)(-) and GMRES(B,random)(–)

$$GMRES(A, r^0) \equiv GMRES(WL^*LW^*, r^0). \tag{3.5}$$

Since $B = WL^*LW^*$ is positive definite, the GMRES algorithm converges strictly monotonically. This gives a new proof of the known fact that if zero is outside the field of values of A, then the GMRES method converges strictly monotonically. (See [3] or [4] for a different proof.)

If, in addition to having its field of values separated from the origin, the matrix A is close to Hermitian, then $H = W^*AW$ will be close to Hermitian, and one might expect that if the matrices U and L in (3.4) are scaled properly, then the Hermitian matrix L^*L would be close to $H = UL$. In this case the eigenvalues of L^*L would be close to either

eigenvalues or singular values of A, indicating that either of these quantities would essentially determine the convergence rate of the GMRES algorithm. We will not go through a formal perturbation analysis here (for such an analysis, see [17]), but make only the following simple observation about the condition number κ of the matrix L^*L:

$$\kappa(L^*L) \leq \kappa(L^*U^{-1}) \cdot \kappa(H) \tag{3.6}$$

Any conditions on A that would ensure that, for any unitarily similar upper Hessenberg matrix H, the properly scaled factors of the UL decomposition of H satisfy $\kappa(L^*U^{-1}) < \kappa(H)$, would ensure that the condition number of $B = WL^*LW^*$ is less than that of A^*A. If the eigenvalues of A^*A are fairly uniformly distributed throughout the interval, this would then imply that the GMRES algorithm applied to A converges in fewer iterations than the algorithm applied to the normal equations A^*A (though, of course, the comparison of total work would depend on the number of steps taken and the cost of applying A and A^*). Unfortunately, we do not know of simple and interesting conditions on the matrix A that guarantee this property.

One can derive rough bounds on the size of the U and L factors separately, based on the distance from the field of values of A to the origin. Similar bounds on the size of U^{-1} and L^{-1} can be expressed in terms of the distance from the field of values of A^{-1} to the origin. If both of these distances are fairly large, then we believe that the typical or worst-case behavior of the GMRES algorithm applied to L^*L will be similar to that of the algorithm applied to A. This is in contrast to the case in which the field of values of A contains the origin. For such problems, it may still be possible to factor the matrix H in the form (3.4) but now the matrix L^*L may be arbitrarily ill-conditioned. In such cases, the typical behavior of the GMRES algorithm applied to L^*L may be much worse than that of the method applied to A, even though the behavior is identical for the particular initial vector r^0 used in generating L^*L.

As an example, we considered the matrix of Lenferink and Spijker [12]. This is a non-normal tridiagonal matrix of the form

$$\text{tridiag}[\ 1/(i+1),\ -3-2i,\ i+1\ ], \quad i = 1, ..., n.$$

It is strongly diagonally dominant and its field of values is well-separated from the origin [19]. Rather than using a random initial vector this time, for each step k, we determined an initial residual r^0 that gave rise to a residual of largest possible norm at step k. This was done using an optimization code, as described previously. For each of these initial vectors, we then computed the Hessenberg matrix H and the factorization (3.4), where U and L were scaled to have the same diagonal elements. The goal was to determine how ill-conditioned L^*L is for these "worst-case" initial vectors.

We used a matrix A of order $n = 16$. The condition number of A was 10.6. The condition number of L^*L ranged from 15.6 to 27.3 – greater than that of A, but less than that of A^*A.

Fig. 4 shows the residual bound

$$2\left[\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k\right]^{-1} \tag{3.7}$$

based on the most ill-conditioned matrix L^*L , with $\kappa = 27.3$ (solid line). From (3.5), this is also a bound on the convergence rate of the GMRES method applied to A. The dashed line in Fig. 4 shows the sharp error bound, which was computed numerically. Note the similarity between the two, at least in the early steps of the computation. This suggests that if a reasonable bound on the condition number of L^*L could be established a priori, then (3.7) would give a realistic bound on the convergence rate of the GMRES algorithm applied to A.

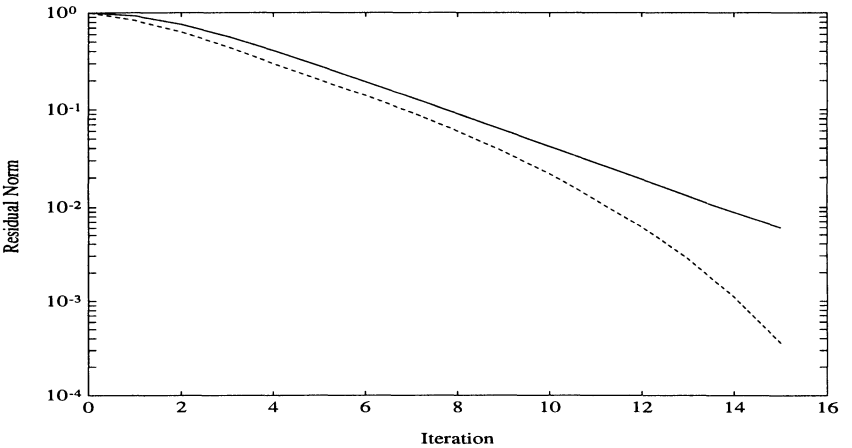

Fig. 4. Bound from HPD Matrix (-) and Sharp Bound (–)

3.3. Equivalent Hermitian Matrices. Using the characterization of Theorem 3, for most matrices A and initial residuals r^0, it is possible to define a Hermitian (not necessarily positive definite) matrix B such that $GMRES(A, r^0) \equiv GMRES(B, r^0)$. If the matrix H^{-1} of Theorem 3 has no zero columns in its lower triangle, then let the Hermitian matrix X be defined by

$$X = \mathcal{L}(H^{-1})D + (\hat{\mathcal{L}}(H^{-1})D)^*, \tag{3.8}$$

where D is any diagonal matrix such that $\mathcal{L}(H^{-1})D$ has real diagonal elements. (Recall that $\mathcal{L}(H^{-1})$ denotes the lower triangle of H^{-1} including

the diagonal, while $\hat{\mathcal{L}}(H^{-1})$ denotes the strict lower triangle of H^{-1}. We will also use the notation $\mathcal{U}(\cdot)$ and $\hat{\mathcal{U}}(\cdot)$ to denote the upper triangle and strict upper triangle of a matrix, respectively.) If H^{-1} has a column of zeros in its lower triangle, say, column j, then that column can be replaced by any column of the form (2.13) provided the diagonal element X_{jj} is real, and then the strict upper triangle of X can be taken to be the Hermitian transpose of the strict lower triangle. If the diagonal matrix D in (3.8), or, the elements X_{nj} in (2.13) can be chosen so that the Hermitian matrix X is nonsingular, then, according to Theorem 3, H can be written in the form $H = RX^{-1}$ for some nonsingular upper triangular matrix R, and so we have

$$GMRES(A, r^0) \equiv GMRES(WX^{-1}W^*, r^0). \tag{3.9}$$

If the matrix A^{-1} is close to Hermitian, then $H^{-1} = W^*A^{-1}W$ will be close to Hermitian, and one might expect that the Hermitian matrix X could be chosen to be close to H^{-1} and hence to have nearby eigenvalues. From this it would follow that the inverses of the eigenvalues of $B = WX^{-1}W^*$ would be close to the inverses of the eigenvalues of A. We will quantify this statement, limiting ourselves, for simplicity, to the case of real matrices A and real initial residuals r^0.

Suppose A^{-1} is real and close to symmetric, say,

$$A^{-1} = M + E$$

where M is symmetric and $\|E\|_F \leq \delta$. [We can take M to be the symmetric part of A^{-1}, $\frac{1}{2}(A^{-1} + A^{-T})$, since this is the closest symmetric matrix to A^{-1} in the Frobenius norm (see, e.g. [10]), and then E will be the skew-symmetric part, $\frac{1}{2}(A^{-1} - A^{-T})$]. Then H^{-1} is also close to symmetric:

$$H^{-1} = W^TMW + W^TEW = \tilde{M} + F$$

where $\tilde{M} = W^TMW$ is symmetric and $\|F\|_F = \|W^TEW\|_F \leq \delta$. If we define X using (3.8) with D equal to the identity, then the difference between X and H^{-1} is given by

$$X - H^{-1} = \mathcal{L}(\tilde{M}) + \mathcal{L}(F) + \hat{\mathcal{U}}(\tilde{M}) + (\hat{\mathcal{L}}(F))^T - \tilde{M} - F = (\hat{\mathcal{L}}(F))^T - \hat{\mathcal{U}}(F),$$

and this satisfies

$$\|X - H^{-1}\| \leq \|X - H^{-1}\|_F \leq \sqrt{2}\,\|F\|_F \leq \sqrt{2}\,\delta. \tag{3.10}$$

Since X is symmetric, this implies that the eigenvalues of X differ from those of H^{-1} by no more than $\sqrt{2}\delta$, or, equivalently, that the eigenvalues

of $B^{-1} = WXW^T$ differ from those of A^{-1} by no more than $\sqrt{2}\delta$. If we make the additional assumption that the absolute value of the smallest eigenvalue of H^{-1} is greater than $\sqrt{2}\delta$, then this will ensure that the matrix X is nonsingular. Since δ is a bound on the size of the skew-symmetric part of A, a sufficient assumption in terms of A and its eigenvalues $\lambda_1, ..., \lambda_n$ is

$$\min_{i=1,...,n} \frac{1}{|\lambda_i|} > \sqrt{2}\, \frac{\|A^{-1} - A^{-T}\|_F}{2}, \tag{3.11}$$

With the assumption (3.11), then, the residual bound (1.7) can be replaced by

$$\frac{\|r^k\|}{\|r^0\|} \leq \min_{p_k \epsilon \mathcal{P}_k} \max_{z^{-1} \epsilon \bigcup_{i=1}^{n} \mathcal{B}(\lambda_i^{-1}, \sqrt{2}\delta) \bigcap \mathbf{R}} |p_k(z)|, \tag{3.12}$$

where $\mathcal{B}(\lambda_i^{-1}, \sqrt{2}\delta)$ denotes the ball of radius $\sqrt{2}\delta$ centered at λ_i^{-1} and $\mathbf{R}$ denotes the real numbers.

If the eigenvalues of A^{-1} are sufficiently well-separated that the disks $\mathcal{B}(\lambda_i^{-1}, \sqrt{2}\delta)$ do not overlap, then X will have exactly one eigenvalue in each of these disks, and if the eigenvalues are ordered accordingly we can write

$$\frac{\|r^k\|}{\|r^0\|} \leq \max_{\{\mu_1,...,\mu_n:\ |\frac{1}{\mu_i} - \frac{1}{\lambda_i}| \leq \sqrt{2}\delta,\ \mu_i \epsilon \mathbf{R},\ i=1,...,n\}} \min_{p_k \epsilon \mathcal{P}_k} \max_{i=1,...,n} |p_k(\mu_i)|. \tag{3.13}$$

The bounds (3.12) and (3.13) are frequently overestimates of the actual residual, because the bound $\sqrt{2}\delta$ on the distance from each eigenvalue of X to that of A^{-1} is a fairly large overestimate. Still, for matrices that are close to symmetric but have ill-conditioned eigenvectors, the bounds (3.12) and (3.13) may be much smaller than (1.7). For such problems they are often comparable to the pseudo-spectral bound (1.11), for the best value of ϵ.

4. Eigenvalue information alone cannot ensure fast convergence. We have mentioned several cases in which the eigenvalues of A essentially determine the convergence rate of the GMRES algorithm:

1. If the matrix Z of eigenvectors is well-conditioned, then the bound (1.7) is nearly sharp.
2. If, for some fairly small value of ϵ, the length of the boundary of the ϵ-pseudo-spectrum is not much greater than $2\pi\epsilon$, then (1.11) gives a reasonable bound on the size of the residual.
3. If A is close to Hermitian in the sense of (3.11), then the bound (3.12) or (3.13) is a reasonable estimate.

In general, however, eigenvalue information alone cannot ensure fast convergence of the GMRES algorithm. Any behavior that can be seen with the method can be seen with the method applied to a matrix having *any* nonzero eigenvalues. This is most easily seen from Theorem 2. Let C be a companion matrix with the desired eigenvalues, say,

$$C = \begin{pmatrix} 0 & . & . & 0 & \alpha_0 \\ 1 & & & & \alpha_1 \\ & . & & & . \\ & & . & & . \\ & & & 1 & \alpha_{n-1} \end{pmatrix}. \tag{4.1}$$

(The eigenvalues of this matrix are the roots of the polynomial $z^n - \sum_{j=0}^{n-1} \alpha_j z^j$, and the constants α_j, $j = 0, ..., n-1$ can be chosen so that C has the desired eigenvalues.) Define the upper triangular matrix $\hat{R}$ in Theorem 2 by

$$\hat{R} = \begin{pmatrix} 1 & 0 & . & 0 & \alpha_1 + < v, w^1 > \alpha_0 \\ 0 & 1 & & . & \alpha_2 + < v, w^2 > \alpha_0 \\ . & & . & . & . \\ 0 & . & . & 1 & \alpha_{n-1} + < v, w^{n-1} > \alpha_0 \\ 0 & 0 & 0 & 0 & < v, w^n > \alpha_0 \end{pmatrix}.$$

Then it is easily seen that $\hat{H}\hat{R}$ is equal to C, and since the spectrum of $\hat{R}\hat{H}$ is the same as that of $\hat{H}\hat{R}$, the matrix $B = W\hat{R}\hat{H}W^*$ also has the desired spectrum. We have thus proved the following corollary to Theorem 2:

COROLLARY 4.1. *Given any nonzero values $\mu_1, ..., \mu_n$, there is a matrix B with eigenvalues $\mu_1, ..., \mu_n$ such that (2.4) holds.*

As an example, let us consider matrices that give rise to the worst possible behavior – no progress at all until step n – when the initial residual is the first unit vector e_1. The following unitary shift matrix was given in ([1]) and ([14]) to illustrate this phenomenon:

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & ... & 1 \\ 1 & 0 & 0 & ... & 0 \end{pmatrix}.$$

It is clear that the initial residual $e_1 = (1, 0, ..., 0)^T$ is already orthogonal to the space $[Ae_1, ..., A^{n-1}e_1]$, and so the residual is not changed until step n. The Hessenberg matrix H and the unitary matrix W generated by the GMRES algorithm with this initial vector are

$$H = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

According to Theorem 1, the class of matrices that generate the same Krylov residual spaces as A (with initial vector e_1) consists of all matrices of the form $WRHW^*$, where R is any nonsingular upper triangular matrix; i.e., all nonsingular matrices of the form

$$\begin{pmatrix} 0 & * & 0 & \dots & 0 \\ 0 & * & * & \dots & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & * & * & \dots & * \\ * & * & * & \dots & * \end{pmatrix},$$

where the $*$'s can be any values.

This class includes, for example, all companion matrices of the form C^*, where C is given in (4.1). As noted previously, the constants α_j, $j = 0, \dots, n-1$, can be chosen to make this matrix have any desired eigenvalues, say, all 1's. It follows that the worst-case behavior of the GMRES algorithm (i.e., no progress until step n) can occur for a matrix having what one might have expected to be the best possible eigenvalues!

5. Conclusions, open questions, and possible applications. We have pointed out several cases in which the eigenvalues of the matrix A essentially determine the behavior of the GMRES algorithm. We have shown, in general, however, that this is not the case. The same behavior can be seen for a matrix having any given eigenvalues.

We have introduced a technique for analyzing the behavior of the GMRES algorithm applied to non-normal matrices by defining an equivalence class containing normal matrices – the class of matrices that, with a given initial residual, produce the same GMRES residuals at every step. In this regard, we have raised more questions than we have answered. We have shown that there is always an equivalent unitary matrix. We have demonstrated numerically that certain properties of the eigenvalues of this unitary matrix are related to special properties of A, but we have not proved such a relationship. We have shown that if zero is outside the field of values of A then there is an equivalent Hermitian positive definite matrix, but we have not been able to bound the condition number of this Hermitian positive definite matrix in terms of simple properties of A.

Our analysis still has not explained the behavior of the first example given in Section 1 – a matrix A of the form $Z\Lambda Z^{-1}$, where Z and Λ are defined by (1.12). There we claimed that the eigenvalues of this matrix essentially determine the convergence of the GMRES iteration. Yet, for certain initial residuals, there can be no normal matrix in A's equivalence class whose eigenvalues are close to those of A. This is because the field of values of A contains the origin. It follows that there is an initial vector for which the GMRES iteration makes no progress at step one. Any normal matrix with this behavior must have eigenvalues on both sides of the origin, and so its spectrum cannot be close to the positive spectrum of A. Using several different random initial residuals and also the initial residuals that give rise to the largest possible residuals at steps one and two, we computed the equivalent unitary matrix WQW^* defined by (3.2) and the equivalent Hermitian matrix $WX^{-1}W^*$ defined by (3.8) (with $D = I$). In all cases, there were large gaps in the spectrum of the unitary matrix, similar to those shown in Fig. 3a. The equivalent Hermitian matrix had one large negative eigenvalue (which probably was not computed accurately due to rounding errors) and the remaining eigenvalues were close to those of A – all but one were in the interval $[1, 5]$ and the remaining one was less than 20. This indicates that for this problem as well, there are normal matrices in A's equivalence class for which the typical behavior of the GMRES algorithm is similar to that for A. It remains an open question how to show a priori that this is the case. Another possible approach is to consider near-normal matrices in the equivalence class – matrices with well-conditioned (but not necessarily unitary) eigenvector matrices. There may be a near-normal matrix in A's equivalence class whose eigenvalues are all close to those of A.

The equivalence class we have defined consists of matrices for which the residuals at each step of the GMRES algorithm are identical. In practice, one seldom runs a GMRES computation for a full n steps. Instead, one runs for some fixed number of steps, say, $m << n$, and then restarts. The class of matrices that generate the same residuals at steps 1 through m is broader than the class we have considered. Investigation of this larger class may lead to results about the convergence of the restarted GMRES iteration.

This technique of analysis – equating the behavior of the algorithm applied to a given problem to its behavior when applied to some other problem that is better understood – may prove useful for analyzing other iterative methods as well. The biconjugate gradient [5] and QMR [7,6] iterations are examples, and work along these lines has begun.

Finally, we would like to note that the results of Section 4 are somewhat discouraging, as far as Krylov space iterative methods are concerned. It is disappointing that one can determine, just from the sparsity pattern of a matrix, that, for some initial residual, every Krylov space method will converge poorly (or diverge)! Similar behavior can be expected with a

random initial residual, as our numerical experiments indicate. While it is sometimes possible to choose an initial guess for which the algorithm will display better convergence, we believe that *typically* the initial guess plays only a small role in determining the convergence rate of the algorithm, and so this behavior is to be expected for problems with such coefficient matrices. (And, of course, this same behavior would be observed for any matrix unitarily similar to one with a bad sparsity pattern.) This suggests that it is important to consider methods that choose approximate solutions from outside the Krylov space. An example is the GMRESR method of van der Vorst and Vuik [20]. Another approach is to look for preconditioners that make the preconditioned matrix close to normal, in some appropriate sense. Further research along these lines would seem an important next step.

Acknowledgment. We would like to thank Nick Trefethen for helpful discussions and advice about the subject matter of this paper.

REFERENCES

[1] P. Brown, "A Theoretical Comparison of the Arnoldi and GMRES Algorithms," SIAM Jour. Sci. Stat. Comput. 20 (1991), pp. 58-78.

[2] N. Dunford and J. T. Schwartz, *Linear Operators. Part I: General Theory*, Interscience Publishers, Inc., New York, 1958.

[3] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, "Variational Iterative Methods for Nonsymmetric Systems of Linear Equations," SIAM J. Num. Anal. 20 (1983), pp. 245-357.

[4] H. C. Elman, "Iterative Methods for Large, Sparse, Nonsymmetric Systems of Linear Equations," Ph.D. thesis and Res. Rep. No. 229, Dept. of Comp. Sci., Yale U., 1982.

[5] R. Fletcher, "Conjugate Gradient Methods for Indefinite Systems," in G. A. Watson, ed., Proc. of the Dundee Biennial Conference on Numerical Analysis, Springer-Verlag, New York, 1975.

[6] R. Freund, M. Gutknecht, and N. Nachtigal, "An Implementation of the Look-Ahead Lanczos Algorithm for Non-Hermitian Matrices," RIACS Tech. Report 91.09, April, 1991.

[7] R. Freund and N. Nachtigal, "QMR: a Quasi-Minimal Residual Method for Non-Hermitian Linear Systems," RIACS Tech. Report 91.09, April, 1991.

[8] A. Greenbaum and L. Gurvits, "Max-Min Properties of Matrix Factor Norms," to appear.

[9] A. Greenbaum and L. N. Trefethen, "GMRES-CR and Arnoldi-Lanczos as Matrix Approximation Problems," to appear.

[10] T. Hayden and J. Wells, "Approximation by Matrices Positive Semidefinite on a Subspace," Lin. Alg. Appl. 109 (1988), pp. 115-130.

[11] W. Joubert, "A Robust GMRES-Based Adaptive Polynomial Preconditioning Algorithm for Nonsymmetric Linear Systems," to appear.

[12] H. W. J. Lenferink and M. N. Spijker, "A Generalization of the Numerical Range of a Matrix," Lin. Alg. Appl. 140 (1990), pp. 251-266.

[13] C. Moler, J. Little, and S. Bangert, *PRO-MATLAB for Sun Workstations*, The MathWorks, Inc., Sherborn, MA, August, 1987.

[14] N. M. Nachtigal, S. Reddy, and L. N. Trefethen, "How Fast are Nonsymmetric Matrix Iterations?" Proceedings of the Copper Mountain Conference on Iterative Methods, Copper Mountain, CO, 1990.

[15] S. C. Reddy, "Pseudospectra of Operators and Discretization Matrices and an Application to Stability of the Method of Lines," Ph.D. thesis, MIT, 1991.

[16] Y. Saad and M. H. Schultz, "GMRES: A Generalized Minimum Residual Algorithm for Solving Nonsymmetric Linear Systems," SIAM J. Sci. Stat. Comput. 7 (1986), pp. 856-869.

[17] G. W. Stewart, "On the Perturbation of LU, Cholesky, and QR Factorizations," in preparation.

[18] L. N. Trefethen, "Approximation Theory and Numerical Linear Algebra," in *Algorithms for Approximation II*, J. C. Mason and M. G. Cox, eds., Chapman and Hall, London, 1990.

[19] L. N. Trefethen, "Pseudospectra of Matrices," in D. F. Griffiths and G. A. Watson, Numerical Analysis 1991, Longman, 1992.

[20] H. van der Vorst and C. Vuik, "GMRESR: A Family of Nested GMRES Methods," Report 91-80, Delft University of Technology, 1991.

[21] P. Vinsome, "Orthomin, An Iterative Method for Solving Sparse Sets of Simultaneous Linear Equations," SPE 5729, Fourth Symposium on Numerical Simulation of Reservoir Performance, SPE, Feb., 1976.

[22] D. M. Young and K. C. Jea, "Generalized Conjugate Gradient Acceleration of Nonsymmetrizable Iterative Methods," Lin. Alg. Appl. 34 (1980), pp. 159-194.