Submitted 5/18/92 to the SIAM Journal on Matrix Analysis and Applications

GMRES/CR and Arnoldi/Lanczos as Matrix Approximation Problems

Anne Greenbaum*
Courant Institute of Mathematical Sciences
251 Mercer St.
New York, NY 10012
greenbau@cims.nyu.edu

Lloyd N. Trefethen†
Department of Computer Science
Cornell University
Ithaca, NY 14853
lnt@cs.cornell.edu

Abstract

The GMRES and Arnoldi algorithms, which reduce to the CR and Lanczos algorithms in the symmetric case, both minimize $\| p(A)\, b \|$ over polynomials p of degree n. The difference is that p is normalized at $z = 0$ for GMRES and at $z = \infty$ for Arnoldi. Analogous "ideal GMRES" and "ideal Arnoldi" problems are obtained if one removes b from the discussion and minimizes $\| p(A) \|$ instead. Investigation of these true and ideal approximation problems gives insight into how fast GMRES converges and how the Arnoldi iteration locates eigenvalues.

*Supported by DOE Contract DEFG0288ER25053.

† Supported by NSF Grant DMS-9116110. Part of the work of both authors was performed while visiting the Institute for Mathematics and its Applications at the University of Minnesota.

1. Introduction

Since the 1950s it has been recognized that matrix iterative methods are naturally connected with approximation theory. The most familiar connections are between polynomial approximation and the numerous iterative methods that make use of Krylov subspaces, including the Richardson, Chebyshev, conjugate gradient, biconjugate gradient, CGNR, GMRES, CGS, Bi-CGSTAB, and QMR iterations. Sometimes rational approximation problems also arise, notably in the analysis of ADI iterations, circulant-preconditioned Toeplitz iterations, and Krylov subspace algorithms via Padé approximation. Recent references on these matters include [4,8,15,21].

The approximation problems that are discussed in the linear algebra literature almost invariably involve scalar functions defined on subsets of the complex plane, or, if the matrix A is symmetric, the real axis. The set in question is the spectrum $\Lambda(A)$ or an estimate of the spectrum. If A is normal, such reductions are sometimes exact in the sense that the behavior of the matrix iteration is determined exactly by the properties of the approximation problem. If A is not normal, however, they are always approximate. GMRES, for example, does not exactly solve any known approximation problem in the complex plane, when A is not normal. Between the approximation problem and the convergence of the matrix iteration there is a gap of size $\kappa(V)$, the condition number of a matrix of eigenvectors of A. When $\kappa(V)$ is large, predictions based on the approximation problem may have little bearing on the actual convergence of the matrix algorithm [15].

The purpose of this paper is to explore a different kind of approximation problem that can also be associated with iterative linear algebra, involving matrices instead of scalars. Instead of asking how small a polynomial $p(z)$ can be on the set $\Lambda(A)$, we ask how small the norm $\|p(A)\|$ can be. Matrix approximation questions are implicit in much of the literature of matrix iterations; we certainly do not claim to be the first to consider them. However, they have received no discussion in print that we are aware of. We believe it is important to investigate these problems if one's goal is an understanding of matrix iterations that does not depend upon hidden assumptions of near-normality.

We shall concentrate on two algorithms for nonsymmetric matrix problems: GMRES, which solves systems of equations $Ax = b$, and Arnoldi, which computes eigenvalues of A. Our matrix approximation analogues of these processes are called the "ideal GMRES" and "ideal Arnoldi" problems. Mathematically, the new result presented here is a proof of the existence and uniqueness of ideal GMRES and

Arnoldi approximants. (Existence is trivial, but uniqueness is surprisingly tricky.) In the final section we propose four questions, as yet unresolved, whose solution might further advance our understanding of matrix iterations.

2. GMRES and Arnoldi

Throughout this paper N and $n < N$ are integers, A is an $N \times N$ matrix,* b is an N-vector, $\| \cdot \|$ is the 2-norm, and

$$P_n = \{\text{polynomials of degree} \le n \text{ with } p(0) = 1\},$$
$$P^n = \{\text{monic polynomials of degree } n\}.$$

The difference between P_n and P^n is that P_n is normalized at $z = 0$ and P^n at $z = \infty$.

GMRES [4,19] is an algorithm that solves the following approximation problem† successively for $n = 1, 2, 3 \ldots$:

GMRES approximation problem. *Find $p_* \in P_n$ such that*

$$\| p_*(A)\, b \| = \textit{minimum}. \tag{1}$$

An equivalent statement is

$$b \approx \left\langle Ab, A^2 b, \ldots, A^n b \right\rangle, \tag{1'}$$

where "$y \approx V$" denotes the problem of finding the best approximation with respect to $\| \cdot \|$ to the point y in the space V. This characterization of GMRES is well known. To explain it one notes that GMRES finds a vector x_n in the Krylov subspace $\mathcal{K}_n = \langle b, Ab, \ldots, A^{n-1}b \rangle$ such that the residual $r_n = b - Ax_n$ has minimal norm over all $x \in \mathcal{K}_n$. This vector x_n can be represented in the form $x_n = q(A)b$ for some polynomial $q(z)$ of degree $n-1$, and (1) comes upon writing $r_n = p_*(A)b$ with $p_*(z) = 1 - zq(z) \in P_n$.

The Arnoldi iteration [1,4] is an algorithm that solves the analogous problem involving P^n instead of P_n:

---

* Nothing essential changes if we take $N = \infty$ and let A be a bounded operator.

† We have assumed that the initial guess for the iteration is $x_0 = 0$. Arbitrary initial guesses can be handled by an easy modification.

Arnoldi approximation problem. *Find $p^* \in P^n$ such that*

$$\| p^*(A)\, b \| = \textit{minimum}. \tag{2}$$

Equivalently,

$$A^n b \approx \left\langle b, Ab, \ldots, A^{n-1}b \right\rangle. \tag{2'}$$

The vector b is no longer the right-hand side of a system of equations, but an arbitrary initial vector. This characterization is also known, but not so widely known.* It can be readily proved as a consequence of the usual formulation of the Arnoldi process in terms of orthogonality. The Arnoldi iteration "finds p^*" in the sense that it constructs a Hessenberg matrix H_n of which p^* is the characteristic polynomial.

If A is real and symmetric, the GMRES and Arnoldi iterations reduce to the conjugate residual (CR) and Lanczos iterations, respectively. Everything said in this paper about GMRES applies also to CR, and everything said about Arnoldi applies also to Lanczos.

A comparison of (1) and (2) suggests that from an approximation point of view, the difference between the GMRES and Arnoldi algorithms is slight. This analogy is rarely brought out in accounts of these algorithms, partly for historical reasons and partly because the usual applications of the two algorithms are different. Whereas GMRES is applied to solve systems of equations $Ax = b$, so that (1) comes quickly to mind as a description of it, the Arnoldi iteration is traditionally thought of as a method for estimating eigenvalues of A. The "Arnoldi eigenvalue estimates"† at step n are the eigenvalues of H_n, that is, the roots of p^*. But what the Arnoldi iteration actually does is solve (2); its connection with eigenvalues is indirect and approximate.

The formulations (1) and (2) provide elegant proofs of certain well-known properties of the GMRES and Arnoldi iterations. For example, one sees immediately from (1) and (2) that both of these iterations are essentially invariant under changes of scale ($A \to \alpha A$, $\alpha \in \mathbb{C}$) and under unitary similarity transformations ($A \to UAU^*$, $U^* = U^{-1}$). The Arnoldi iteration is also translation-invariant

* In the symmetric (Lanczos) case (2) appears as Corollary 12-3-7 of [17], unfortunately with a typographical error.

† The roots of p^* are Ritz values of A with respect to the Krylov subspace (2′). By analogy, the roots of p_* are sometimes called "pseudo-Ritz values" [4]. The Ritz values lie in the field of values of A, and the pseudo-Ritz values lie in the inverse of the field of values of A^{-1}; see [13].

($A \to A + \alpha I$, $\alpha \in \mathbb{C}$), since ∞ is translation-invariant, but GMRES is not, since 0 is not. In these statements and throughout this paper, we ignore the effects of rounding error.

3. Ideal GMRES and Ideal Arnoldi

The GMRES and Arnoldi iterations depend on the starting vector b. However, one may remove b from the discussion and pose the following "ideal" approximation problems:

Ideal GMRES approximation problem. *Find $q_* \in P_n$ such that*

$$\| q_*(A) \| = \textit{minimum}, \tag{3}$$

or equivalently,

$$I \approx \left\langle A, A^2, \dots, A^n \right\rangle. \tag{3'}$$

Ideal Arnoldi approximation problem. *Find $q^* \in P^n$ such that*

$$\| q^*(A) \| = \textit{minimum}, \tag{4}$$

that is,

$$A^n \approx \left\langle I, A, \dots, A^{n-1} \right\rangle. \tag{4'}$$

Whereas (1) and (2) are vector approximation problems, (3) and (4) involve matrices. Procedures for computing these polynomials, either actual or in our imaginations, might be called ideal GMRES and ideal Arnoldi algorithms. Some computations of this kind are discussed in §6.

We believe that studying these idealized problems is a fruitful way to gain insight into the properties of Krylov subspace iterations in linear algebra. Our reasoning is as follows. The behavior of a GMRES or Arnoldi iteration is determined by two things: A and b. However, though the special properties of b are occasionally important, more often the features that one cares about do not differ very much from one choice of b to another. It is the properties of A that usually decide between an iteration that converges in 10 steps and one that requires 100 or 1000 (which in practice means it is time to look for a better preconditioner). By passing from (1)–(2) to (3)–(4) we disentangle this matrix essence of the process

from the distracting effects of the initial vector, and end up with a pair of elegant mathematical problems in the bargain.

The solutions to (1)–(2) and (3)–(4) are related by the following bounds. The proof of this theorem is easy; the four inequalities follow from the minimality properties (1), (3), (2), and (4), respectively.

THEOREM 1. *The true and ideal GMRES polynomials are related by*

$$\frac{\| p_*(A)\, b \|}{\|b\|} \le \| q_*(A) \| \le \|p_*(A)\|, \tag{5}$$

and the true and ideal Arnoldi polynomials are related by

$$\frac{\| p^*(A)\, b \|}{\|b\|} \le \| q^*(A) \| \le \|p^*(A)\|. \tag{6}$$

These two pairs of bounds are identical in form, but from the point of view of applications, the nature of the relationship between (4) and (2) is quite different from that between (3) and (1). The purpose of GMRES is to solve (1), not (3). The relevance of (3) is that it gives an upper bound on how slow the convergence may be, thanks to (5), and if the right-hand-side b is "random enough" one may expect that this bound will be close to sharp. For an Arnoldi iteration aimed at estimating eigenvalues, the logic is reversed. It is our opinion that the essence of the process by which an Arnoldi iteration locates eigenvalues is the solution not of (2) but of (4). The iteration solves (2) becase that is what is computationally tractable, but the implicit hope is that if b is "random enough," the solution to (2) will be a good approximation to the solution to (4). See [23].

The ideal Arnoldi polynomial q^* might be called the degree n *Chebyshev polynomial of* A, in analogy to the notion in approximation theory of a Chebyshev polynomial of a subset of the complex plane, which is a monic polynomial that achieves minimal sup-norm on that set.* Another way to view q^* is as a *pseudo-annihilating polynomial* for A, i.e., a monic polynomial that maps A to a matrix of norm $\epsilon \approx 0$. According to the usual definition, an annihilating polynomial is a polynomial that annihilates A exactly. This is a fragile concept, however, ill-posed with respect to perturbations of A and with little quantitative force. Krylov subspace iterations in numerical linear algebra are founded on the observation that

*See Chapter 16 of [9]. This analogy becomes an identity if A is normal; see the next section.

for practical purposes, a pseudo-annihilating polynomial with parameter $\epsilon = 10^{-10}$ or 10^{-20} is as useful as an exact annihilating polynomial and may be of vastly lower degree.

4. The special case when A is normal

When A is normal, problems (1)–(4) reduce to standard problems of approximation theory. For any vector b we have

$$p(A)b = \sum_{j=1}^{N} \omega_j p(\lambda_j) v_j, \qquad \omega_j = v_j^T b,$$

where $\{\lambda_j\}$ and $\{v_j\}$ are a set of eigenvalues and corresponding orthonormal eigenvectors of A, and therefore

$$\|p(A)b\| = \left(\sum_{j=1}^{N} |\omega_j|^2 |p(\lambda_j)|^2 \right)^{1/2}.$$

Thus the Arnoldi and GMRES problems (1)–(2) are equivalent to weighted least-squares approximation problems in the complex plane: find an appropriately normalized polynomial $p(z)$ that has minimal weighted 2-norm on $\Lambda(A)$ with respect to the discrete weight function $\{|\omega_j|^2\}$. As for the ideal Arnoldi and GMRES problems (3)–(4), the identity

$$\|p(A)\| = \sup_{\lambda \in \Lambda(A)} |p(\lambda)|$$

(for normal matrices only) shows that they are are equivalent to Chebyshev approximation problems in the complex plane: find a polynomial that has minimal supremum norm on $\Lambda(A)$. In particular, the ideal Arnoldi polynomial q_* is exactly the same as the Chebyshev polynomial for the set $\Lambda(A)$, mentioned in the last section.

Both weighted least-squares and Chebyshev approximation problems in the complex plane are well understood and discussed in many books. Existence and uniqueness of best approximations are easily proved (we defer a precise statement to the next section). In the Chebyshev case the computation of best approximations can be carried out by various methods such as linear programming, variations of the Remes algorithm, Lawson's algorithm, or other ideas; see [20] and the references therein. In the least-squares case the computation of best approximants is

a matter of routine linear algebra. One can use a QR decomposition, for example, and that is exactly what GMRES does.

There are two reasons to pay special attention to the case in which A is normal. First, some matrices are normal or close enough to normal that the results one obtains are sometimes applicable in practice. In particular, the conjugate residual and Lanczos iterations fall in this category since symmetric matrices are normal. Second, most people's intuitions about the behavior of matrices are based on the normal case. By studying how the normal case differs from the general one we obtain a valuable check on our intuitions.

5. Existence and uniqueness

We return now to problems (1)–(4) for matrices A that are arbitrary, i.e., not necessarily normal. The most fundamental questions to be asked about (1)–(4) are those of existence and uniqueness. For (1) and (2) the answers to both are straightforward and well known. For (3) and (4), existence is straightforward but uniqueness is not. So far as we are aware, though this seems surprising, the uniqueness of the solutions to (3) and (4) is a new result.

THEOREM 2. *The optimal polynomials p_*, p^*, q_*, q^* all exist. Provided that the minima in (1)–(4) are nonzero, and provided in the cases of GMRES and ideal GMRES that A is nonsingular, they are unique.*

Proof. Consider the formulations (1′)–(4′). In each case we have a problem of the form $y \approx V$, where V is a finite-dimensional subspace of a vector space W and $y \in W$. (For (1) and (2), V and W are spaces of N-vectors, whereas for (3) and (4) they are spaces of $N \times N$ matrices; generically we can speak of "vectors" in either case.) Existence of a closest point $v \in V$ to y follows by a standard compactness argument that can be found in any book on approximation theory. See for example p. 20 of [2] or p. 17 of [12].

The question of uniqueness of the polynomials p_*, p^*, q_*, q^* can be divided into two parts:

(a) Is the closest point $v \in V$ to y unique?

(b) Does it have a unique representation as a linear combination of the n vectors indicated in (1′)–(4′)?

Part *(b)* can be dispatched as follows. What we have to show is that the n vectors in question are linearly independent. Consider first the ideal Arnoldi problem

$(4')$, and suppose to the contrary that $I, A, \ldots, A^{n-1}$ are linearly dependent. Then by multiplying by a power of A if necessary we can find a linear combination of them that is zero and in which the coefficient of A^{n-1} is nonzero. Thus $A^{n-1} \in \langle I, A, \ldots, A^{n-2} \rangle$, which implies $A^n \in \langle A, A^2, \ldots, A^{n-1} \rangle$ and therefore $\|q^*(A)\| = 0$, contradicting the assumption that the minimum in (4) is nonzero. An analogous argument applies to $(2')$. For the GMRES problem $(3')$, suppose $A, A^2, \ldots, A^n$ are linearly dependent, which by a similar argument implies $A^n \in \langle A, A^2, \ldots, A^{n-1} \rangle$ and thus again $\|q^*(A)\| = 0$. If the constant term of q^* is nonzero, then dividing by that constant term yields a properly normalized GMRES polynomial q_* with $\|q_*(A)\| = 0$. On the other hand if the constant term is zero, then since A is nonsingular we can multiply by A^{-1} one or more times until it becomes nonzero. An analogous argument applies to $(1')$.

This brings us to part *(a)* of the proof of uniqueness. For problems $(1')$ and $(2')$ the uniqueness of v follows by another standard result in approximation theory, since the vector norm $\|\cdot\|$ is strictly convex. See, for example, p. 23 of [2] or p. 17 of [12]. The *matrix* norm $\|\cdot\|$, however, is not strictly convex, so a different argument is required for $(3')$ and $(4')$. One must wet one's hands in the linear algebra.

Consider first the ideal Arnoldi problem (4). Suppose that q_1 and q_2 are two distinct solutions to (4), and let the minimal norm they attain be

$$\|q_1(A)\| = \|q_2(A)\| = C.$$

If we define $q(z) = \frac{1}{2}(q_1(z) + q_2(z))$, then $\|q(A)\| \le C$, so we must have $\|q(A)\| = C$ since q_1 and q_2 are minimal. Let $w_1, \ldots, w_J$ be a set of maximal right singular vectors for $q(A)$, i.e., a set of orthonormal vectors with

$$\|q(A)w_j\| = C, \qquad 1 \le j \le J,$$

with J as large as possible. For each w_j we must have

$$\|q_1(A)w_j\| = \|q_2(A)w_j\| = C$$

and

$$q_1(A)w_j = q_2(A)w_j,$$

for otherwise, by the strict convexity of the vector norm $\|\cdot\|$, we would have $\|q(A)w_j\| < C$. Thus

$$(q_2 - q_1)(A)w_j = 0, \qquad 1 \le j \le J.$$

Now since $(q_2 - q_2)(z)$ is not identically zero, we can multiply it by a scalar and a suitable power of z to obtain a monic polynomial $\Delta q \in P^n$ such that

$$\Delta q(A) w_j = 0, \qquad 1 \le j \le J.$$

For $\epsilon \in (0,1)$, consider now the polynomial $q_\epsilon \in P^n$ defined by the convex linear combination

$$q_\epsilon(z) = (1-\epsilon) q(z) + \epsilon \Delta q(z).$$

If $w_{J+1}, \ldots, w_N$ denote the remainder of a set of N singular vectors of $q(A)$, with corresponding singular values $C > \sigma_{J+1} \ge \cdots \ge \sigma_N \ge 0$, then we have

$$\|q_\epsilon(A) w_j\| \le \begin{cases} (1-\epsilon) C & (1 \le j \le J), \\ (1-\epsilon)\sigma_{J+1} + \epsilon \|\Delta q(A)\| & (J+1 \le j \le N). \end{cases}$$

The first row is $< C$ for arbitrary ϵ, and the second row is $< C$ for sufficiently small ϵ, since $\sigma_{J+1} < C$. Since the singular vectors $w_1, \ldots, w_N$ form an orthonormal basis for $\mathbb{R}^N$, this implies that $\|q_\epsilon(A)\| < C$ for sufficiently small ϵ, contradicting the assumption that q_1 and q_2 are minimal.

For the ideal GMRES problem (3) the argument is the same except that from $q_2 - q_1$ we need to construct $\Delta q \in P_n$ rather than $\Delta q \in P^n$. If the constant term of $q_2 - q_1$ is nonzero, we do this by dividing by that constant term. If it is zero, we make use of the assumption that A is nonsingular and multiply by a suitable power of z^{-1}. ∎

6. Computations

If A is normal, the ideal Arnoldi and GMRES polynomials q^* and q_* are simply Chebyshev polynomials for the set $\Lambda(A)$, as noted in §4, and can be computed by various algorithms. If A is not normal, however, we know of no simple algorithm that is guaranteed to compute q^* and q_*. For simplicity, from now on we shall consider the ideal GMRES polynomial q_*; our remarks carry over straightforwardly to the ideal Arnoldi problem.

We have found that in many cases, q_* can be computed by using an optimization code to determine an initial vector b, with $\|b\| = 1$, for which $\|p_*(A)\, b\|$ is maximal at the prescribed step n. From Theorem 1 we have

$$\|p_*(A)\, b\| \le \|q_*(A)\| \le \|p_*(A)\| \tag{7}$$

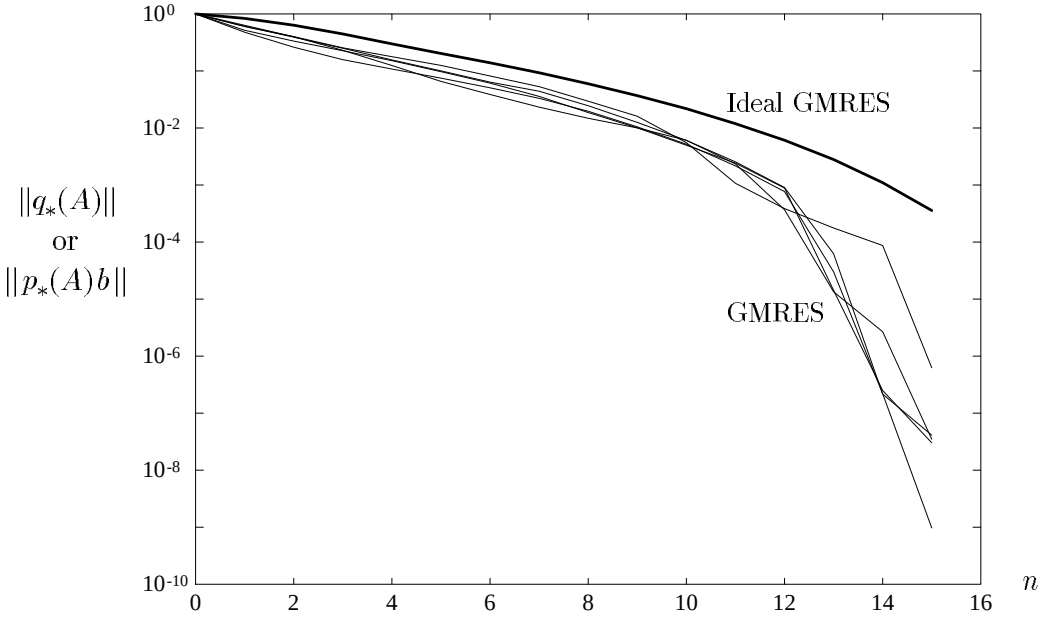

Figure 1. Convergence curves for the 16×16 Lenferink-Spijker matrix. The upper curve corresponds to ideal GMRES and the lower curves to standard GMRES with five different random initial vectors b.

for any b with $\|b\| = 1$. If a choice of b can be found for which $\| p_*(A)\, b \| = \| p_*(A) \|$, it follows that $p_* = q_*$. It is not known whether such a b always exists, but we conjecture that it does (see the first question of the next section). Maximizing the left-hand side of (7) seems to be easier in practice than minimizing the right-hand side, presumably because the latter problem is non-smooth. We have carried out our computations using the Matlab optimization routine *fminu* [14] coupled with a GMRES subprogram to compute $\| p_*(A)\, b \|$ for a given vector b. Although several attempts with different initial guesses b are often required for the optimization code to succeed, it usually does so eventually.

As an alternative to the use of a general optimization program such as *fminu*, we have found that one can also maximize the left-hand side of (7) by means of an algorithm modeled on Lawson's algorithm, which is a method of iteratively reweighted least-squares best known in the context of L^p approximation [18]. We have obtained good results this way in many cases. However, we have not yet found a method of this kind that converges consistently, and will therefore not give details here.

Figure 1 illustrates the behavior of the ideal vs. true GMRES polynomials by a very simple example—a matrix of Lenferink and Spijker [11,22]. This is a non-

normal tridiagonal matrix of the form $tridiag((i+1)^{-1}, -3-2i, i+1)$, $i = 1, \ldots, N$. For the Lenferink-Spijker matrix of order $N = 16$, we computed the ideal GMRES polynomials q_* of degree 1 through 15 as well as the true GMRES polynomials p_* for five different random initial vectors. The thick curve in the figure represents the norms $\|q_*(A)\|$ as a function of n, and the thinner curves represent $\|p_*(A)\,b\|$ for the various vectors b. The GMRES curves lie below the ideal GMRES curve, as they must, but exhibit qualitatively the same shape. Our experiments have not been sufficiently extensive to draw conclusions about how GMRES and ideal GMRES convergence curves compare in general.

It is interesting to note that while the norm of the ideal GMRES polynomial for this problem decreases strictly monotonically, this does not always happen. For many problems, $\|q_*(A)\|$ remains exactly equal to 1 for a number of steps before it begins to decrease. The case $n = 1$ of this phenomenon is fully understood: one can show that $\|q_*(A)\| < 1$ at step 1 if and only if the field of values of A lies in an open half-plane with respect to the origin in $\mathbb{C}$ (see [3] and §6 of [15]). For some problems, $\|q_*(A)\| = 1$ for steps 1 through $N-1$. This happens frequently with random matrices, for example. For such problems there is a right-hand side vector b for which the GMRES algorithm makes no progress whatsoever until step N.

A difference we have observed between the ideal GMRES polynomials for normal and non-normal matrices is the following. If A is normal, then $q_*(A)$ must have at least $n+1$ equal maximal singular values. This follows from the fact that the degree n Chebyshev polynomial for a set in the complex plane always takes on its maximum absolute value in at least $n+1$ distinct points. In contrast, when A is not normal, our experiments indicate that after the initial phase with $\|q_*(A)\| = 1$ just mentioned, $q_*(A)$ usually has only one maximal singular value.

7. Open questions

Our work on ideal GMRES and Arnoldi approximations has raised more questions than it has answered. We shall close with a list of four questions that we consider particularly interesting. In each of the following A is a matrix, b is a vector normalized by $\|b\| = 1$, and the *convergence curve* is the curve of $\|p_*(A)\,b\|$ or $\|q_*(A)\|$ as a function of the step number n. The questions are posed for GMRES, but they all have Arnoldi analogues.

1. Is the envelope attained? Theorem 1 asserts that the GMRES convergence curve lies below the ideal GMRES convergence curve, as illustrated in Figure 1. Given n, does there exist an initial vector b such that these two curves intersect

at step n, i.e., such that $\|p_*(A)\,b\| = \|q_*(A)\|$? The answer is known to be yes for symmetric matrices [5] and more generally for normal matrices [7,10], and for arbitrary matrices at step $n = 1$ [7,10]. It is also yes in the "generic" non-normal case in which the maximal singular value of $q_*(A)$ is simple. Whether it is yes in all cases is not known. If it is, then the ideal GMRES convergence curve can be described precisely as the upper envelope of the GMRES convergence curves corresponding to all initial vectors b.

2. What convergence curves are possible? At step n, $\|q_*(A)\|$ must be at least as small as $\min_{1\le k\le n-1} \|q_k(A)\|\|q_{n-k}(A)\|$, where q_k denotes the ideal GMRES polynomial for A at step k. Geometrically this means that the ideal GMRES convergence curve is convex when plotted on a logarithmic scale. Are all convergence curves that satisfy this convexity constraint possible? If not, how can one characterize those convergence curves that are possible?

3. Can any matrix be simulated by a normal matrix? Short of a full characterization of convergence curves, one may naturally ask if the possibilities for normal matrices are more restricted than for non-normal matrices. In other words, are there convergence curves that can only be generated by a non-normal matrix? For standard GMRES the answer has recently been proved to be no; any sequence $\|p_*(A)\,b\|$ as a function of n can be duplicated by another sequence $\|p_*(\tilde{A})\,b\|$ where $\tilde{A}$ is normal—in fact, unitary [6]. For ideal GMRES, the answer is unknown.

4. Do the pseudospectra determine the convergence curve? Our final question returns to the topic of whether matrix iterations are equivalent to approximation problems in the complex plane. No such equivalence has yet been found in the case where A is not normal. In particular, the spectrum of A does not determine the ideal GMRES convergence curve, and neither do the Jordan canonical form or the field of values. However, what about the ϵ-pseudospectra of A, that is, the sets $\Lambda_\epsilon(A) = \{z \in \mathbb{C} : \|(zI - A)^{-1}\| \ge \epsilon^{-1}\}$ [15,21,22]? Can two matrices have identical ϵ-pseudospectra for all $\epsilon > 0$ but distinct ideal GMRES convergence curves? This is a special case of the following more general question. If $\Lambda_\epsilon(A) = \Lambda_\epsilon(B)$ for all $\epsilon > 0$, where A and B are two matrices (not necessarily of the same dimension), does this imply $\|f(A)\| = \|f(B)\|$ for all analytic functions f? (Without loss of generality we can be take f to be a polynomial.) In other words, if "behavior" is measured by norms of functions of matrices, do the pseudospectra of a matrix determine its behavior?

Acknowledgments

We are happy to acknowedge several discussions with Michael Overton, who showed us how the ideal Arnoldi and GMRES problems relate to more general problems of minimization of singular values of functions of matrices [16].

References

[1] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem,* Quart. Appl. Math. 9 (1951), 17–29.

[2] E. W. Cheney, *Introduction to Approximation Theory,* McGraw-Hill, New York, 1966.

[3] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, *Variational iterative methods for nonsymmetric systems of linear equations,* SIAM J. Numer. Anal. 20 (1983), 245–357.

[4] R. W. Freund, G. H. Golub, and N. M. Nachtigal, *Iterative solution of linear systems,* in Acta Numerica 1992, Cambridge U. Press, 1992.

[5] A. Greenbaum, *Comparison of splittings used with the conjugate gradient algorithm,* Numer. Math. 33 (1979), 181–194.

[6] A. Greenbaum, *Matrices that generate the same Krylov varieties,* in preparation.

[7] A. Greenbaum and L. Gurvits, *Max-Min properties of matrix factor norms,* in preparation.

[8] M. H. Gutknecht, *The unsymmetric Lanczos algorithms and their relations to Padé approximation, continued fractions, and the qd algorithm,* Preliminary Proceedings of the Copper Mountain Conference on Iterative Methods, 1990 (preliminary version).

[9] E. Hille, *Analytic Function Theory, v. 2,* Chelsea, New York, 1973.

[10] W. A. Joubert, *A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems,* submitted to SIAM J. Matrix Anal. Applics.

[11] H. W. J. Lenferink and M. N. Spijker, *On the use of stability regions in the numerical analysis of initial value problems,* Math. Comp. 57 (1991), 221–237.

[12] G. G. Lorentz, *Approximation of Functions,* Chelsea, New York, 1986.

[13] T. A. Manteuffel and J. Otto, *On the roots of the orthogonal polynomials and residual polynomials associated with the conjugate gradient method,* submitted to J. Numer. Lin. Alg..

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

[15] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?,* SIAM J. Matrix Anal. Applics., to appear.

[16] M. L. Overton, *Large-scale optimization of eigenvalues,* SIAM J. Optim., to appear.

[17] B. N. Parlett, *The Symmetric Eigenvalue Problem,* Prentice-Hall, Englewood Cliffs, NJ, 1980.

[18] J. R. Rice and K. H. Usow, *The Lawson algorithm and extensions,* Math. Comp. 22 (1968), 118–127.

[19] Y. Saad and M. H. Schultz, *GMRES: A generalized minimum residual algorithm for solving nonsymmetric linear systems,* SIAM J. Sci. Stat. Comp. 7 (1986), 856–869.

[20] P. T. P. Tang, *A fast algorithm for linear complex Chebyshev polynomial approximation,* Math. Comp. 51 (1988), 721–739.

[21] L. N. Trefethen, *Approximation theory and numerical linear algebra,* in J. C. Mason and M. G. Cox, eds., Algorithms for Approximation II, Chapman and Hall, London, 1990.

[22] L. N. Trefethen, *Pseudospectra of matrices,* in D. F. Griffiths and G. A. Watson, eds., Numerical Analysis 1991, Longman, 1992.

[23] L. N. Trefethen, *Arnoldi and GMRES lemniscates,* in preparation.