SIAM J. SCI. COMPUT.
Vol. 17, No. 1, pp. 1–15, January 1996

CALCULATION OF PSEUDOSPECTRA BY THE ARNOLDI ITERATION*

KIM-CHUAN TOH† AND LLOYD N. TREFETHEN‡

Abstract. The Arnoldi iteration, usually viewed as a method for calculating eigenvalues, can also be used to estimate pseudospectra. This possibility may be of practical importance, because in applications involving highly nonnormal matrices or operators, such as hydrodynamic stability, pseudospectra may be physically more significant than spectra.

Key words. Arnoldi, Lanczos, pseudospectra, numerical range, hydrodynamic stability

AMS subject classifications. 65F15, 47A12

1. Introduction.

Large-scale nonsymmetric matrix eigenvalue problems, which typically arise via discretization of non-self-adjoint differential or integral operators, are commonly solved numerically by the Arnoldi iteration and its variants [1], [8], [18], [25]. In this paper we explore the possibility that the Arnoldi iteration can also be used for the estimation of pseudospectra. Such an idea was first proposed by Nachtigal, Reichel, and Trefethen [17] and Freund [7], and methods much closer to those of the present paper have been presented by Ruhe in talks at the Householder (1993) and St. Girons (1994) symposia on linear algebra [23]. Recent developments indicate that in some applications, the pseudospectra of a matrix or operator may be more significant physically than its spectrum (see §7). Since calculation of pseudospectra is much more expensive than calculation of spectra, this suggests that it may be desirable to develop methods for determining them iteratively.

The idea investigated here is that the pseudospectra of a matrix A can be approximated by those of the Hessenberg matrices constructed by an Arnoldi iteration. In the version of this paper originally submitted for publication [30], we made use of the $n \times n$ Hessenberg matrices H_n (see the next section for definitions). However, Tom Manteuffel recommended to us that it might be advantageous to consider instead the $(n+1) \times n$ Hessenberg matrices $\tilde{H}_n$, as is done, for example, in [13]. Meanwhile, this is also the idea that Axel Ruhe has been investigating. In the end we have decided to present experiments here for both H_n and $\tilde{H}_n$, while giving greater attention to the latter. For the examples we have computed, the distinction between the two makes little difference in practice, but $\tilde{H}_n$ has theoretical advantages (monotonic convergence) and also conceptual appeal, because it bypasses the usual consideration of Ritz values or "Arnoldi eigenvalue estimates." Thus we find ourselves exploring a new way of interpreting approximations not only of pseudospectra but also of spectra, one that may be of interest even in the special case of the Lanczos iteration for symmetric matrices.

For any $\epsilon \geq 0$, the ϵ-pseudospectrum of a matrix A is defined by

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} : \ \|(zI - A)^{-1}\| \geq \epsilon^{-1}\}, \tag{1}$$

with the convention $\|(zI - A)^{-1}\| = \infty$ if $z \in \Lambda(A)$, the spectrum of A [32]. If $\|\cdot\|$ is the 2-norm, an equivalent definition is

*Received by the editors May 18, 1994; accepted for publication (in revised form) April 10, 1995. This work was supported by NSF grant DMS-9116110 and DOE grant DE-FG02-94ER25199.

†Center for Applied Mathematics, Cornell University, Ithaca, NY 14853 (kc@cam.cornell.edu).

‡Department of Computer Science, Cornell University, Ithaca, NY 14853 (LNT@cs.cornell.edu).

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} : \sigma_{\min}(zI - A) \le \epsilon\}, \tag{2}$$

where $\sigma_{\min}(zI - A)$ denotes the smallest singular value of $zI - A$. Either definition makes it clear that the pseudospectra of a matrix are a family of nested subsets of $\mathbf{C}$, with $\Lambda_0(A) = \Lambda(A)$. They can be computed most straightforwardly by evaluating $\sigma_{\min}(zI - A)$ on a grid of values in the complex z-plane, then sending the result to a contour plotter. S.-H. Lui has shown that this straightforward algorithm can be speeded up by a factor of five to ten by a preliminary reduction of A to Hessenberg form followed by inverse iteration combined with continuation [12]. Variants of this procedure involve triangular instead of Hessenberg reduction or Lanczos instead of inverse iteration. Such ideas apply to our matrices H_n and $\tilde{H}_n$ as well as A, and thus have little bearing on the relative speedup to be gained via Arnoldi iterations, so we shall not give details.

If A is normal (has a complete set of orthogonal eigenvectors), then $\Lambda_\epsilon(A)$ is just the closed ϵ-neighborhood of $\Lambda(A)$, but if A is far from normal, $\Lambda_\epsilon(A)$ may be much larger. These are the cases where difficulties are likely to arise if one tries to use the spectrum $\Lambda(A)$ to estimate quantities such as $\|A^n\|$, $\|e^{tA}\|$, or $\|f(A)\|$. Better estimates can often be obtained from the pseudospectra using methods such as Cauchy integrals, the Laplace transform, or the Kreiss matrix theorem [21], [32]–[34].

2. The Arnoldi iteration. Let A be a real or complex $m \times m$ matrix, and let $\|\cdot\|$ denote the 2-norm. A complete unitary reduction of A to upper Hessenberg form might be written $A = QHQ^*$ or $AQ = QH$. The idea of the Arnoldi iteration is to compute the successive steps of this reduction columnwise, starting from the condition that the first column of Q is a prescribed vector q_1 with $\|q_1\| = 1$. Let Q_n be the $m \times n$ matrix whose columns are the first n columns of Q,

$$Q_n = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_n \end{array}\right], \tag{3}$$

and let $\tilde{H}_n$ be the $(n+1) \times n$ upper-left section of H,

$$\tilde{H}_n = \begin{pmatrix} h_{11} & & \cdots & h_{1n} \\ h_{21} & h_{22} & & \\ & \ddots & \ddots & \vdots \\ & & h_{n,n-1} & h_{nn} \\ & & & h_{n+1,n} \end{pmatrix}. \tag{4}$$

Then we have

$$AQ_n = Q_{n+1}\tilde{H}_n, \tag{5}$$

and the nth column of this equation can be written $Aq_n = h_{1n}q_1 + \cdots + h_{nn}q_n + h_{n+1,n}q_{n+1}$. The Arnoldi iteration is the modified Gram–Schmidt iteration that implements this $(n+1)$-term recurrence relation:

$q_1 = \text{arbitrary } (\|q_1\| = 1)$
for $n = 1, 2, 3, \ldots$
  $v = Aq_n$
  for $j = 1$ **to** n
    $h_{jn} = q_j^* v$
    $v = v - h_{jn} q_j$
  $h_{n+1,n} = \|v\|$
  $q_{n+1} = v / h_{n+1,n}$

The vectors $\{q_j\}$ form orthogonal bases of the successive *Krylov subspaces* generated by A and q_1,

$$\mathcal{K}_n = \langle q_1, Aq_1, \ldots, A^{n-1}q_1 \rangle = \langle q_1, q_2, \ldots, q_n \rangle \subseteq \mathbf{C}^m.$$

As a practical matter the iteration can be implemented with the aid of a "black box" procedure for computing the matrix–vector products Aq_n, which can be designed to take advantage of sparsity or other structure of A.

The product $Q_n^* Q_{n+1}$ is equal to the $n \times (n+1)$ identity, i.e., the $n \times (n+1)$ matrix with 1 on the main diagonal and 0 elsewhere. Therefore $Q_n^* Q_{n+1} \tilde{H}_n$ is the $n \times n$ Hessenberg matrix obtained by removing the last row of $\tilde{H}_n$,

$$H_n = \begin{pmatrix} h_{11} & \cdots & h_{1n} \\ h_{21} & h_{22} & \\ & \ddots & \ddots & \vdots \\ & & h_{n,n-1} & h_{nn} \end{pmatrix}. \tag{6}$$

From (5) we can accordingly derive the formula

$$H_n = Q_n^* A Q_n. \tag{7}$$

Though they differ in only one row, H_n and $\tilde{H}_n$ are entirely different objects. To highlight the difference, it is helpful to extend $\tilde{H}_n$ to the $m \times n$ matrix $\hat{H}_n$ consisting of the first n columns of H:

$$\hat{H}_n = \begin{pmatrix} h_{11} & \cdots & h_{1n} \\ h_{21} & h_{22} & \\ & \ddots & \ddots & \vdots \\ & & h_{n,n-1} & h_{nn} \\ & & & h_{n+1,n} \\ 0 & & & 0 \\ \vdots & & & \vdots \\ 0 & & & 0 \end{pmatrix}. \tag{8}$$

The matrices $\tilde{H}_n$ and $\hat{H}_n$ are identical except for the presence of $m - (n+1)$ additional rows of zeros in the latter, and thus they have, for example, the same rank, norm, and singular values.

We can interpret H_n and $\hat{H}_n$ as follows. With respect to the basis $\{q_1, q_2, \ldots\}$, $\hat{H}_n$ represents the action of A on $\mathcal{K}_n$, whereas H_n represents the same operation followed by an orthogonal projection back into $\mathcal{K}_n$. Since the domain and range of H_n are the same, it makes sense to speak of the eigenvalues of H_n; but since $\hat{H}_n$ has distinct domain and range, it does not make sense to speak of its eigenvalues.

3. Estimation of pseudospectra and numerical range. We propose that in many cases, for sufficiently large n, some of the pseudospectra of A can be reasonably approximated by the corresponding pseudospectra of H_n or $\tilde{H}_n$:

$$\Lambda_\epsilon(A) \approx \Lambda_\epsilon(H_n) \approx \Lambda_\epsilon(\tilde{H}_n). \tag{9}$$

For $n \ll m$, the computation of $\Lambda_\epsilon(H_n)$ or $\Lambda_\epsilon(\tilde{H}_n)$ will be $O((m/n)^3)$ times faster than that of $\Lambda_\epsilon(A)$. Note that in considering $\Lambda_\epsilon(\tilde{H}_n)$, we are dealing with the ϵ-pseudospectrum of a rectangular matrix. This set can be defined just as for square matrices by (1), with I now denoting a rectangular version of the identity and $(zI - A)^{-1}$ denoting the pseudoinverse. Equivalently, it can be defined by (2). As far as we know, pseudospectra of rectangular matrices have not been discussed before. However, the smallest singular value of $\tilde{H}_n$ has been considered in work by Meza, for example, on iterative solution of ill-conditioned systems of equations [16].

We are not aware of very satisfactory theorems to justify the approximation $\Lambda_\epsilon(A) \approx \Lambda_\epsilon(H_n)$. On the other hand the approximations $\Lambda_\epsilon(A) \approx \Lambda_\epsilon(\tilde{H}_n)$ converge monotonically (cf. (3.21) of [16] and Thm. 3.1 of [13]).

THEOREM 1. *Let an $m \times m$ matrix A be unitarily similar to a Hessenberg matrix H, and let $\tilde{H}_n$ denote the $(n+1) \times n$ section* (4). *(In particular, $\tilde{H}_n$ might be computed by an Arnoldi iteration, with arbitrary restarts in case a zero subdiagonal element $h_{n+1,n}$ is encountered.) Then for any z we have*

$$\sigma_{\min}(zI - \tilde{H}_1) \geq \sigma_{\min}(zI - \tilde{H}_2) \geq \sigma_{\min}(zI - \tilde{H}_3) \geq \cdots \geq \sigma_{\min}(zI - A), \tag{10}$$

and, consequently, for any $\epsilon \geq 0$,

$$\Lambda_\epsilon(\tilde{H}_1) \subseteq \Lambda_\epsilon(\tilde{H}_2) \subseteq \Lambda_\epsilon(\tilde{H}_3) \subseteq \cdots \subseteq \Lambda_\epsilon(A). \tag{11}$$

Proof. Since $zI - \tilde{H}_n$ and $zI - \hat{H}_n$ differ only by rows of zeros, they have the same singular values, and we may replace $\tilde{H}_n$ in (10) by $\hat{H}_n$. Since $zI - A$ and $zI - H$ are unitarily similar, they too have the same singular values, and we may replace A in (10) by H. Since $\hat{H}_n$ is simply the first n columns of H, (10) now follows directly from the characterization $\sigma_{\min}(A) = \min_{\|x\|=1} \|Ax\|$. By (2), this implies (11). □

It is interesting to compare Theorem 1 with the more familiar interpretation of Arnoldi and Lanczos iterations. Conventionally, a set of n Ritz values are considered at step n, and one is faced with the problem of estimating how close they may be to eigenvalues of A. In Theorem 1, there are no Ritz values. However, it may be noted that Ritz values can be defined as the points z at which $\sigma_{\min}(zI - H_n)$ achieves a local minimum (namely, zero). An analogue for the rectangular case would be to consider the points at which $\sigma_{\min}(zI - \tilde{H}_n)$ achieves a local minimum. It follows from (10) that this minimum value will be equal to zero if and only if z is an eigenvalue of A corresponding to an eigenvector that lies in the Krylov subspace $\mathcal{K}_n$.

Besides pseudospectra, it is well known that an Arnoldi iteration also may provide estimates of the numerical range (= field of values) of A, which we denote by $W(A)$. Now it is H_n that we most naturally make use of:

$$W(A) \approx W(H_n). \tag{12}$$

gain we have monotonic convergence (cf. Thm. 3.1 of [13]).

THEOREM 2. *Let A and H be as in Theorem* 1, *and let H_n denote the $n \times n$ section* (6). *Then*

$$W(H_1) \subseteq W(H_2) \subseteq W(H_3) \subseteq \cdots \subseteq W(A). \tag{13}$$

Proof. This is an easy consequence of the definition $W(A) = \{x^*Ax : \|x\| = 1\}$. □

4. Numerical experiments. We now turn to numerical examples computed in MATLAB. Consider first the $m \times m$ "Kahan matrix"

$$A = \begin{pmatrix} 1 & -c & -c & -c & -c \\ & s & -sc & -sc & -sc \\ & & s^2 & -s^2c & -s^2c \\ & & & \ddots & \vdots \\ & & & & s^{m-1} \end{pmatrix}, \tag{14}$$

where $s^{m-1} = 0.1$ and $s^2 + c^2 = 1$. Matrices of this type were proposed by Kahan to illustrate that QR factorization with column pivoting is not a fail-safe method of numerical rank determination [8], [11]. Pseudospectra of this matrix, for $m = 32$, were plotted in [32].

In Fig. 1, we take $m = 64$ and consider Arnoldi approximations $\Lambda_\epsilon(H_n)$ to $\Lambda_\epsilon(A)$ with $n = 5, 10, 15, 20$. At each of these steps, the figure shows the approximate numerical range and the ϵ-pseudospectra for $\epsilon = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$. In this and all of our numerical experiments, the initial vector q_1 is random (independent normally distributed entries).

To the eye, at least, the convergence of $\Lambda_\epsilon(H_n)$ to $\Lambda_\epsilon(A)$ in Fig. 1 is compelling. At $n = 5$, the Arnoldi iteration has not learned much of value, but by $n = 10$, recognizable approximations have begun to emerge. At $n = 20$ the approximations are excellent. Since the cost of an SVD grows cubically with the dimension of the matrix, and 1000 or more SVDs are involved in making one of these plots, calculations of pseudospectra even for these small matrices can be time-consuming, and if $n = 64$ can be replaced by $n = 20$, the savings will be a factor of around 30.

Note that, as is typical in cases of extreme nonnormality, the convergence of the eigenvalues of H_n to those of A in Fig. 1 is slow. The eigenvalues are too ill conditioned to be easily resolved. This is just the sort of problem where eigenvalues are likely to be of limited physical significance and where pseudospectra may provide a useful alternative. For a discussion of the physical significance of pseudospectra, including transient evolution phenomena, the effect of small perturbations, and the notion of "pseudo-resonance" in highly nonnormal systems, see [34].

Figure 2 shows the same computation, for the same matrix, except now the approximations are based on the rectangular Hessenberg matrix $\tilde{H}_n$. Broadly speaking, the approximations are about as good as in Fig. 1. Notice that there are no longer any Ritz values in the plot, which might be considered conceptually advantageous. On the other hand, in dealing with a matrix whose behavior was close to normal in some parts of the complex plane and far from normal in others, it might certainly be convenient to have Ritz values.

Not every matrix behaves as nicely as in Figs. 1 and 2. In Fig. 3 we consider the 64×64 "Grcar matrix," a Toeplitz matrix with -1 on the subdiagonal and 1 on the main diagonal and on the first three superdiagonals. This time, Arnoldi approximations based on $\tilde{H}_n$ at steps 10, 20, 30, and 40 are plotted, and only at $n = 40$ is reasonable convergence of the pseudospectra beginning to be evident. Although $(40/64)^3 \approx 0.24$ is still substantially less than 1, this is

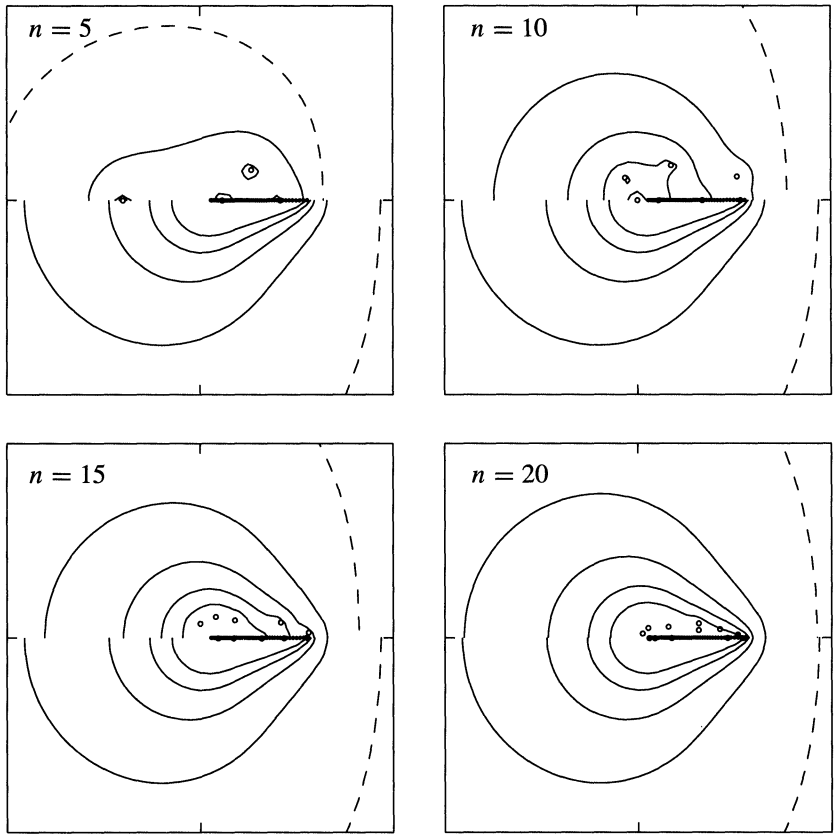

FIG. 1. *ϵ-pseudospectra of the* 64 × 64 *Kahan matrix* (14) *compared with those of four Arnoldi approximations* H_n ($\epsilon = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$). *The upper half of each plot corresponds to* $\Lambda_\epsilon(H_n)$, *and the lower half to* $\Lambda_\epsilon(A)$. *The dashed curves represent an analogous comparison of the numerical ranges* $W(H_n)$ *and* $W(A)$. *The small dots are the eigenvalues of A in the lower half-plane (hard to distinguish; they appear like a solid interval on the real axis), and the small circles are the eigenvalues of* H_n *in the upper half-plane (Ritz values). The real and imaginary axes are marked by ticks in each plot; the axis limits are* $-1.8 \le \Re z \le 1.8$, $-1.8 \le \Im z \le 1.8$.

a case where there is probably little to be gained in approximating $\Lambda_\epsilon(A)$ by $\Lambda_\epsilon(\tilde{H}_n)$ (or $\Lambda_\epsilon(H_n)$, whose performance is about the same).

Figures 2 and 3 represent two of the thirteen examples of highly nonnormal matrices considered in [32]. We have plotted Arnoldi approximations to pseudospectra for all of these and find that the Kahan matrix exhibits the best convergence and the Grcar matrix among the worst. Based on these examples alone, one would probably conclude that the Arnoldi approach to calculation of pseudospectra holds some promise but is not completely convincing.

However, the matrices of [32] are not typical of the large-scale problems that arise in practice. For these special matrices, selected for their dramatic pseudospectra, the nonnormality is such that all m eigenvalues are strongly coupled to one another. In applications, it is more typical for a small number of eigenvalues to be dominant and not strongly coupled to the others, with the behavior of the pseudospectra in the vicinity of these eigenvalues being of primary interest. In such cases Arnoldi approximations may perform better.

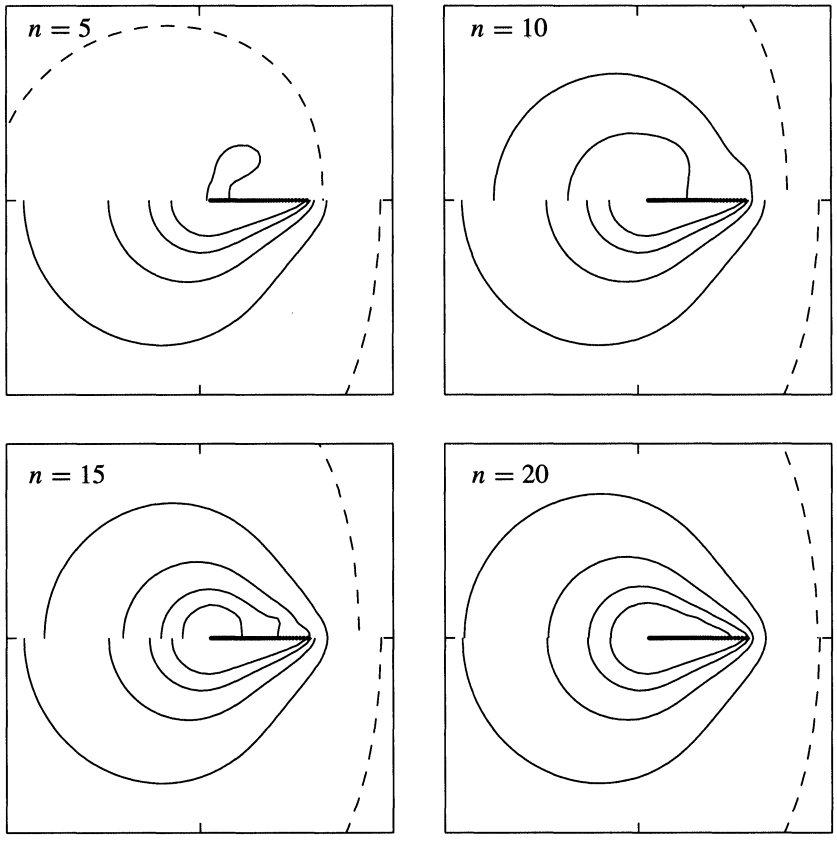

FIG. 2. *Same as Fig. 1, but now the curves in the upper half of each plot correspond to* $\Lambda_\epsilon(\check{H}_n)$ *instead of* $\Lambda_\epsilon(H_n)$. *Since the Ritz values have no simple connection to* $\check{H}_n$, *they are no longer shown. The dashed curves still correspond to* $W(H_n)$ *and* $W(A)$.

Figure 4 presents an example of this kind. Here A is the 64×64 bidiagonal matrix defined by

$$a_{k,k+1} = a_{k,k} = k^{-1/2}. \tag{15}$$

We can think of this as a prototype of a highly nonnormal compact operator and imagine that it is the behavior of the spectrum and pseudospectra away from the origin that is of interest. The convergence in this part of the complex plane is excellent. Moreover, it would be nearly as fast even if m were much larger than 64. For an example like this, the payoff in estimating pseudospectra iteratively could be huge.

5. Modified Arnoldi iterations. The large-scale matrices that arise most often in applications are of a kind different from all of our examples so far. These are discretizations of differential operators, which are not compact. As in Fig. 4, it is the behavior near a few leading eigenvalues that is of greatest interest, but the rest of the spectrum will typically extend to ∞ in the complex plane.

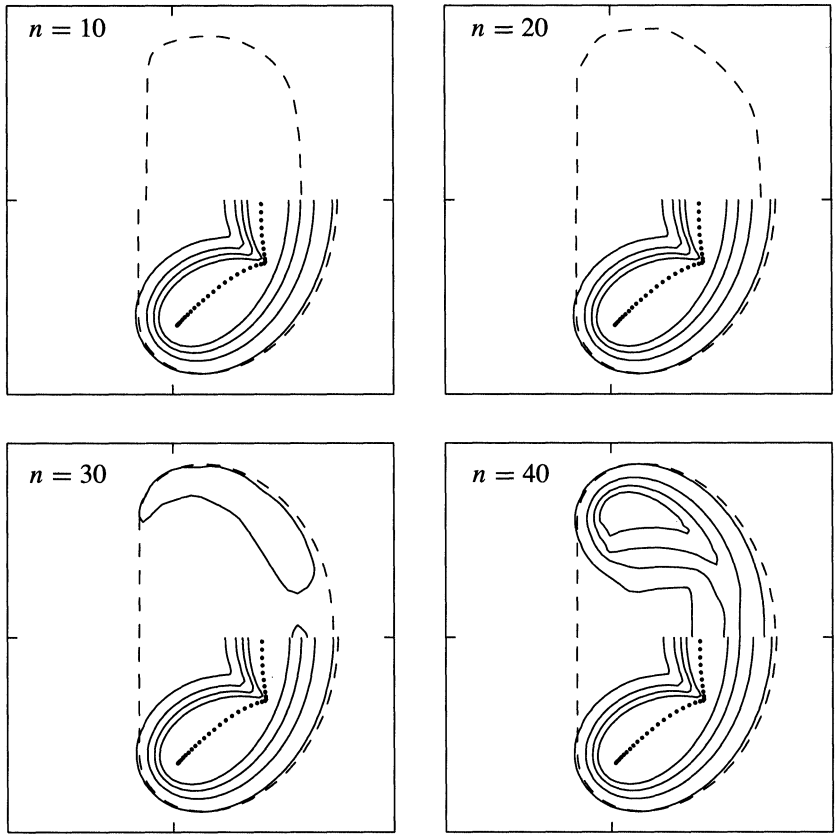

FIG. 3. *Same as Fig. 2, except for the pentadiagonal Grcar matrix of dimension* $m = 64$. *The axis limits are* $-3 \le \Re z \le 4$, $-3.5 \le \Im z \le 3.5$.

It is well known that a pure Arnoldi iteration may be ineffective in cases of this kind. As an example, consider another bidiagonal matrix defined by

$$a_{k,k} = -0.3k, \qquad a_{k,k+1} = 1. \tag{16}$$

Here the spectral and pseudospectral behavior near the origin should be largely unaffected by whether the dimension is 64 or 64,000. A pure Arnoldi iteration will have difficulty nonetheless, and the difficulty will increase with the dimension. This is illustrated in Fig. 5a, with $m = 64$, where we see quite disappointing convergence to the pseudospectra near $z = 0$.

Solutions to this problem have been proposed by a number of authors. One approach is to suppress the part of the spectrum far from the origin by an ancillary linear process such as a Chebyshev iteration or some other polynomial filter. Such ideas have been investigated by Chatelin and her colleagues and by Saad, Scott, Sorensen, and others [6], [9], [17], [24], [25], [27], [29]. A more powerful possibility, when it is feasible, is to modify the problem with the use of matrix inverses, in effect working with rational functions of A rather than just polynomials. Variations on this theme go by names such as shift-and-invert Arnoldi and rational Krylov iteration, and have been investigated by Ruhe, Saad, and Spence, among others [15], [22], [23], [25]. To be effective, such methods depend on the assumption that inverting A (i.e., solving a system $Ax = b$) is cheaper than the computation of main interest. This

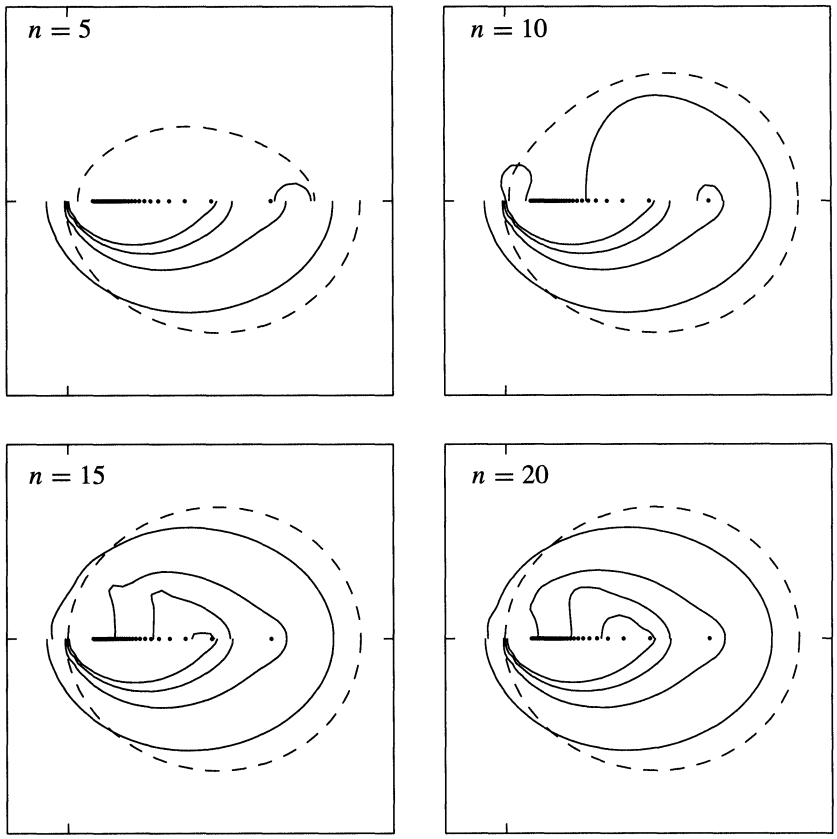

FIG. 4. *Same as Fig. 2 but for the bidiagonal matrix* (15), *a prototype of a compact operator. The axis limits are* $-0.3 \le \Re z \le 1.6$, $-0.95 \le \Im z \le 0.95$.

assumption is satisfied by many sparse eigenvalue problems, since sparsity can often be better taken advantage of for systems of equations than eigenvalues. It is amply satisfied in many calculations of pseudospectra, since these computations are even more expensive. Finally, an intermediate class of acceleration methods is based on preconditioning the eigenvalue problem by inverses not of A itself but of more easily inverted approximations $M \approx A$. This is the idea behind Davidson's method [4], [5] and related methods developed more recently by Meerbergen, Van der Vorst, and others [14], [25], [28].

We shall not attempt a systematic comparison of the uses for estimating pseudospectra of the various acceleration and preconditioning methods that have been proposed. Instead, we shall consider just the simplest modified Arnoldi process to give an idea of the great speedups that may be achieved by these methods. Figure 5b is a repetition of Fig. 5a in which the Arnoldi iteration has been replaced by an "inverse Arnoldi" iteration carried out with A^{-1} instead of A. This entails a solution of a system of equations involving A at each step, but this is a minor matter since A is bidiagonal. The result is a Hessenberg matrix H_n that approximates A^{-1}, and it is the pseudospectra of H_n^{-1} that we plot as approximations to those of A. Since H_n^{-1} is square, the plot also shows its eigenvalues.

Figure 5b shows excellent agreement of $\Lambda_\epsilon(H_n^{-1})$ and $\Lambda_\epsilon(A)$. Evidently the inverse-Arnoldi idea is highly effective for this problem. We take this as illustrative of the kind

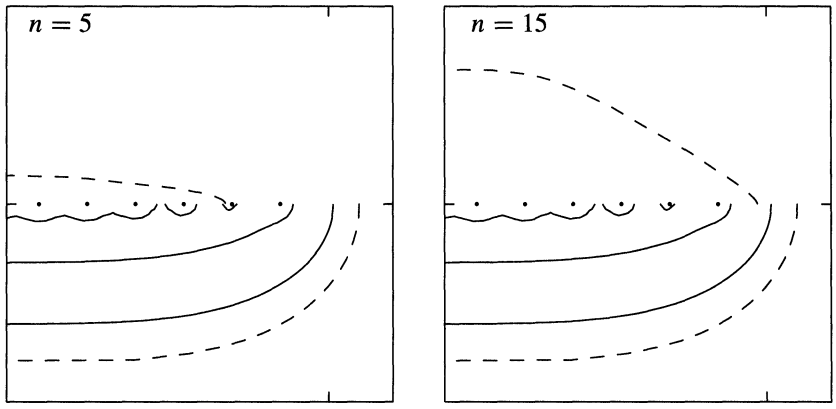

FIG. 5A. *Same as Fig.* 2 *but for the bidiagonal matrix* (16), *a prototype of an unbounded operator. The axis limits are* $-2 \le \Re z \le 0.4$, $-1.2 \le \Im z \le 1.2$.

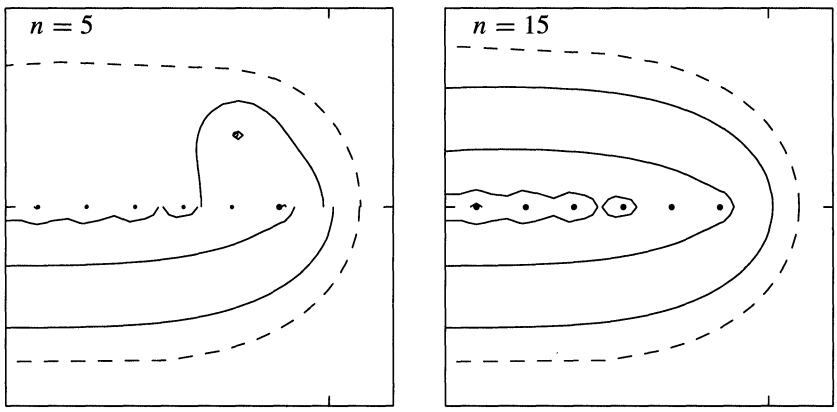

FIG. 5B. *Repetition of Fig.* 5a *with the Arnoldi iteration replaced by an inverse-Arnoldi iteration based on* A^{-1} *instead of A. The plot shows pseudospectra and eigenvalues of* H_n^{-1}.

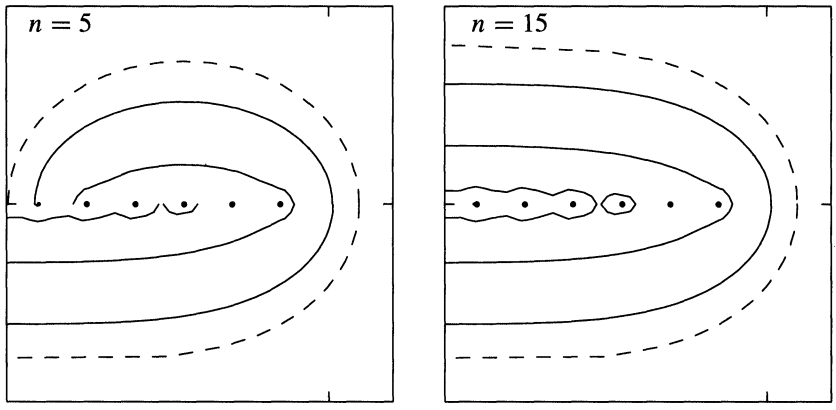

FIG. 5C. *Repetition of Fig.* 5a *with the Arnoldi iteration replaced by a projection of A onto the invariant subspace associated with the n eigenvalues of maximal real part.*

of gains that may be achieved by acceleration techniques in cases where A is invertible at reasonable cost. More sophisticated algorithms of this kind (based on $\tilde{H}_n$, not just H_n) are currently under development by Ruhe [23].

6. Projection onto an invariant subspace. There is another, more elementary, trick that we must not fail to mention. The calculation of pseudospectra is expensive, much more so than the calculation of a single eigenvalue decomposition. It follows that when the latter is affordable, much may be gained by simply calculating an eigenvalue decomposition of A, then projecting it onto the subspace of $\mathbf{C}^m$ spanned by certain eigenvectors. In typical applications these might be the eigenvectors associated with a subset of eigenvalues of A of maximal real part. This idea may be useful even when no Arnoldi iterations are in store; it is used, for example, in [20].

The mechanics of such a projection are as follows. We have already noted after (8) that if Q is an $m \times n$ matrix with orthonormal columns, then Q^*AQ represents the projection of A onto the column space of Q. Suppose now that we start with an $m \times n$ matrix V whose columns are selected eigenvectors of A, satisfying $AV = VD$ for some $n \times n$ diagonal eigenvalue matrix D. If $V = QR$ is a QR decomposition of V, with Q of dimension $m \times n$ and R of dimension $n \times n$, then we have $Q^*V = R$ and $Q = VR^{-1}$ and therefore

$$Q^*AQ = Q^*AVR^{-1} = Q^*VDR^{-1} = RDR^{-1}. \tag{17}$$

Thus RDR^{-1} (upper triangular) is the matrix representation of the projection of A onto the subspace spanned by the selected eigenvectors.

Figure 5c illustrates that for the example (16), this eigenvalue projection idea gives highly accurate pseudospectra. Despite its triviality, this trick can save a great deal of work. For example, an eigenvalue decomposition of a matrix of dimension 300 is a straightforward matter, whereas computing pseudospectra of such a matrix is a major project on most machines available today. If the dimension can be reduced to 30 by eigenvalue projection, the calculation of pseudospectra becomes easy.

Figure 6, following the format of Fig. 5, presents a less contrived example. Consider the convection-diffusion operator

$$\mathcal{L}u = u'' + u', \qquad u(0) = u(d) = 0 \tag{18}$$

acting on the interval $[0, d]$ (in the Hilbert space $L^2[0, d]$). The spectrum of this operator is a discrete, unbounded subset of the negative real axis, but, as discussed in [21], the pseudospectra are large regions in the left half-plane shaped approximately like parabolas. Taking $d = 40$, suppose we want to determine these pseudospectra in the neighborhood of the origin determined by the axis limits in Fig. 6. As discussed in [21], an efficient procedure is to construct a discretization matrix A based on Chebyshev spectral differentiation (we omit details). Unfortunately, for an accurate picture, the dimension of A has to be on the order of 100, making the calculation of the pseudospectra quite time-consuming, and if we wanted results in a larger region of the complex plane, matters would get worse. The figure shows that the inverse-Arnoldi idea works reasonably well here. Since A is dense, the eigenvalue projection idea is even better, and with $n = 40$ it produces a perfect picture with ten times less computing than would be involved in treating the full matrix with $m = 100$.

The dashed curves in Fig. 6, corresponding to the boundary of the numerical range, are worth a comment. Note that in Figs. 6b and 6c, we appear to have convincing convergence of the numerical range estimates in the upper half of the plot, but no dashed mirror image appears

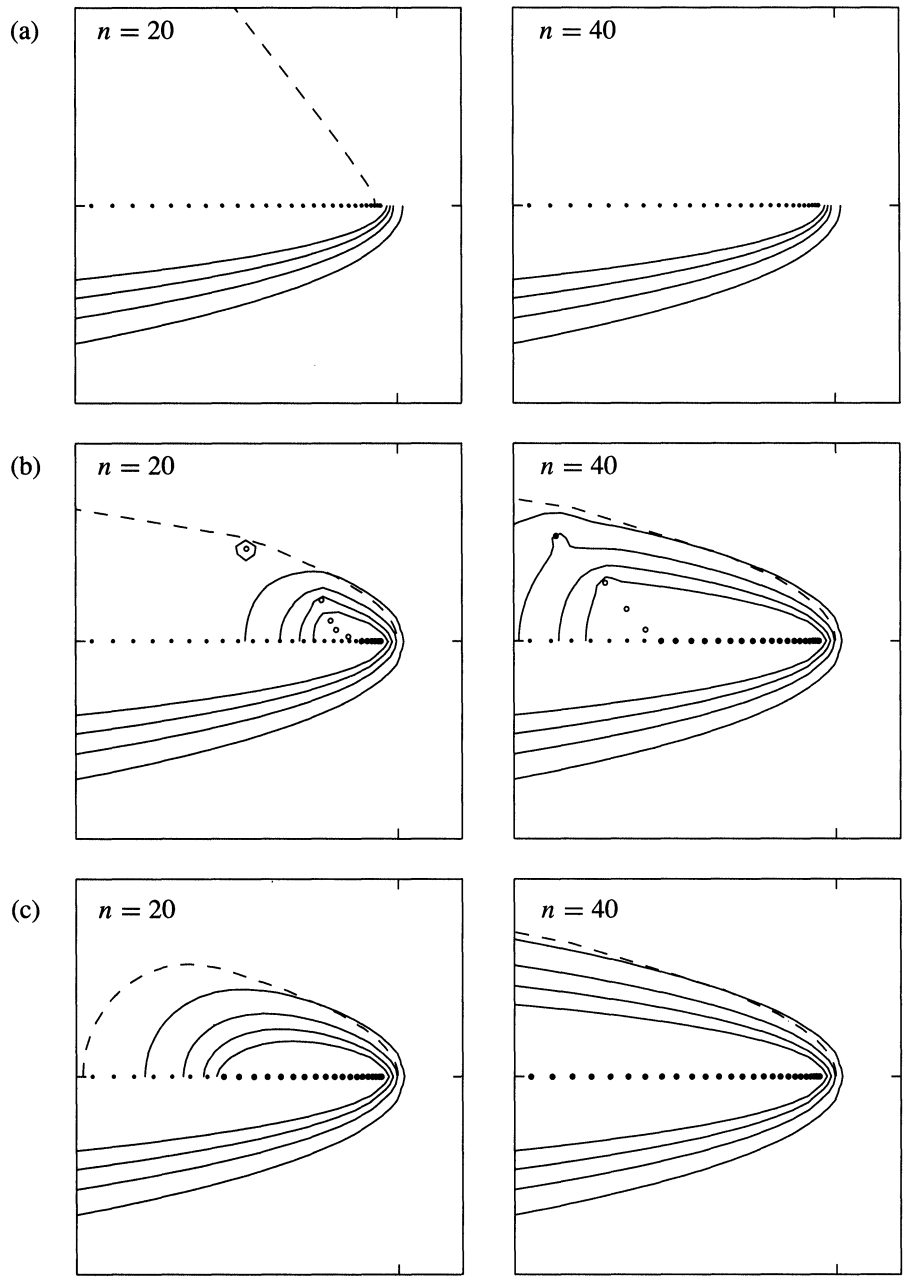

FIG. 6. *Same as Fig. 5, but for a* 100×100 *Chebyshev spectral approximation to the convection-diffusion operator* (18). *The axis limits are* $-5 \le \Re z \le 1$, $-3 \le \Im z \le 3$.

in the lower half. The explanation is that the actual 100×100 spectral differentiation matrix considered here has some huge "outlier" eigenvalues of size 1.2×10^4. These eigenvalues are artifacts of the discretization, with no relevance to the convection-diffusion operator $\mathcal{L}$, but they make the numerical range of A much larger than the axis scales of the figure. Thus the numerical range estimates in the upper half of Fig. 6 are no good at all, strictly speaking, for the matrix A being approximated. As it happens, however, they are excellent approximations

to the numerical range of the operator $\mathcal{L}$. This stroke of good fortune is more than just coincidence, but of course one would have to be cautious about counting on such effects in applications.

For larger problems than those illustrated in this paper, a combination of the Arnoldi iteration and eigenvalue projections might be advantageous. A matrix of dimension 5000, for example, might be projected to dimension 100 by the Arnoldi iteration, whereupon an eigenvalue decomposition might be used to project further to a matrix of dimension 50, whose pseudospectra could then be plotted quickly by the methods proposed by Lui [12]. In such a sequence, the speedup over a naive calculation of pseudospectra might be on the order of many thousands.

7. Discussion. In this paper we have proposed that the iterative algorithms that have been developed for calculating spectra of large matrices may also be useful for estimating pseudospectra. If the matrix is a sparse approximation of a differential operator, the gains to be achieved by rational variants of the Arnoldi algorithm may be very great. If the matrix is dense, other acceleration devices may play a role, and surprisingly good results can be achieved by the simple method of computing the eigenvalue decomposition of A, then projecting onto an invariant subspace associated with a subset of the eigenvalues.

We have made no attempt to explain why our methods approximate pseudospectra as well as they do, beyond the lower bound on $\Lambda_\epsilon(A)$ of Theorem 1. An upper bound on $\Lambda_\epsilon(A)$ has been developed by Ruhe [23], but it appears to be far from sharp in practice.

It should be emphasized that the idea of using Arnoldi iterations for purposes more general than the calculation of eigenvalues is not new. The Arnoldi iteration potentially has relevance in all kinds of matrix problems where A is too big to deal with directly, but where there is reason to expect that the essential behavior can be captured by a low-dimensional projection. The example that has received the most attention is the use of Krylov subspaces to approximate e^{tA} [6], [10], [26]. Of course, the approximation of pseudospectra is not unrelated to the approximation of e^{tA}, since the ultimate purpose of estimating pseudospectra is often to obtain better insight into the behavior of $\|e^{tA}\|$ than the spectrum alone provides.

In closing, we would like to make two remarks.

Our first observation concerns the uses of Arnoldi iterations and the uses of spectra. The examples of this paper have shown that in cases of pronounced nonnormality, the behavior of a Krylov subspace iteration may be more closely tied to the pseudospectra of a matrix or operator than to its spectrum. A curious parallel of this statement is the recent discovery that in applications involving pronounced nonnormality, what is ultimately of physical interest may also be tied more closely to the pseudospectra than to the spectrum. In particular this is true of problems of hydrodynamic instability of fluid flows in a pipe or a channel, where traditional eigenvalue methods fail to explain the instabilities that are observed in practice, but pseudospectra do much better [2], [3], [19], [31], [34]. In our view, these two parallel statements about spectra and pseudospectra form a natural pair. In highly nonnormal problems, the Arnoldi iteration may indeed not be effective at determining eigenvalues, but we should not wish it to be. The information that it does acquire may be deeper and more valuable. It will be interesting to see whether this vision of broader uses of Arnoldi iterations in applications comes to fruition in upcoming years.

Our second remark is addressed to all those who use nonsymmetric Krylov subspace iterations or who produce software for such computations. For the present, due to the great variety of iterations and preconditioners that have been found to be useful, computations of this kind almost invariably involve an element of exploration. We believe that such explorations can be carried out far more effectively if the user habitually produces plots, not just numbers. Such plots might show Ritz values, lemniscates, pseudospectra of H_n, pseudospectra of $\tilde{H}_n$,

eigenvalues of perturbed matrices, or other things—tastes differ, and at this time no one choice seems clearly superior to all others. But we firmly believe it is a mistake to plot nothing at all, or to plot nothing but Ritz values. The habit of looking at plots leads almost unconsciously to new questions and new understanding. Matrices and operators have personalities, which may be revealed with surprising economy by a few curves on the computer screen.

Acknowledgments. We are greatly indebted to Tom Manteuffel for his good advice, which has had a fundamental impact on this paper, and to Axel Ruhe for sharing with us his unpublished report [23]. We are also grateful to Anne Greenbaum for numerous technical suggestions.

REFERENCES

[1] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
[2] L. Boberg and U. Brosa, *Onset of turbulence in a pipe*, Z. Naturforschung, 43a (1988), pp. 697–726.
[3] K. M. Butler and B. F. Farrell, *Three-dimensional optimal perturbations in viscous shear flow*, Phys. Fluids A, 4 (1992), pp. 1637–1650.
[4] M. Crouzeix, B. Philippe, and M. Sadkane, *The Davidson method*, SIAM J. Sci. Comput., 15 (1994), pp. 62–76.
[5] E. R. Davidson, *The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices*, J. Comput. Phys., 17 (1975), pp. 87–94.
[6] W. S. Edwards, L. S. Tuckerman, R. A. Friesner, and D. C. Sorensen, *Krylov methods for the incompressible Navier-Stokes equations*, J. Comput. Phys., 110 (1994), pp. 82–102.
[7] R. W. Freund, *Quasi-kernel polynomials and their use in non-hermitian matrix iterations*, J. Comput. Appl. Math., 43 (1992), pp. 135–158.
[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, MD, 1989.
[9] D. Ho, *Tchebychev acceleration technique for large scale nonsymmetric matrices*, Numer. Math., 56 (1990), pp. 721–734.
[10] M. Hochbruck and C. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., submitted.
[11] W. Kahan, *Numerical linear algebra*, Canad. Math. Bull., 9 (1966), pp. 757–801.
[12] S.-H. Lui, *Computation of pseudospectra by continuation*, SIAM J. Sci. Comput., 17 (1996), to appear.
[13] T. A. Manteuffel and G. Starke, *On hybrid iterative methods for nonsymmetric systems of linear equations*, Numer. Math., to appear.
[14] K. Meerbergen and D. Roose, *Matrix transformations for computing rightmost eigenvalues of large sparse nonsymmetric matrices*, Report TW206 (revised), Dept. of Comp. Sci., K.U. Leuven, April 1995.
[15] K. Meerbergen, A. Spence, and D. Roose, *Shift-invert and Cayley transforms for detection of rightmost eigenvalues of nonsymmetric matrices*, BIT, 34 (1994), pp. 409–423.
[16] J. C. Meza, *A Modification to the GMRES Method for Ill-Conditioned Linear Systems*, Report SAND95-8220, Sandia National Laboratories, Albuquerque, NM, April 1995.
[17] N. M. Nachtigal, L. Reichel, and L. N. Trefethen, *A hybrid GMRES algorithm for nonsymmetric linear systems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795.
[18] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.
[19] S. C. Reddy and D. S. Henningson, *Energy growth in viscous channel flows*, J. Fluid Mech., 252 (1993), pp. 209–238.
[20] S. C. Reddy, P. J. Schmid, and D. S. Henningson, *Pseudospectra of the Orr–Sommerfeld operator*, SIAM J. Appl. Math., 53 (1993), pp. 15–47.
[21] S. C. Reddy and L. N. Trefethen, *Pseudospectra of the convection-diffusion operator*, SIAM J. Appl. Math., 54 (1994), pp. 1634–1649.
[22] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems*, II: *Matrix pairs*, Linear Algebra Appl., 197/198 (1994), pp. 283–296.
[23] ———, *The Rational Krylov Algorithm for Large Nonsymmetric Eigenvalues—Mapping the Resolvent Norms (Pseudospectrum)*, Draft report based on talk delivered at Sparse Days at St. Girons symposium (July, 1994), unpublished but available from `http://http.cs.berkeley.edu/~ruhe`, March, 1995.
[24] Y. Saad, *Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems*, Math. Comp., 42 (1984), pp. 567–588.
[25] ———, *Numerical Methods for Large Eigenvalue Problems*, Manchester U. Press, Manchester, UK, 1992.

[26] Y. SAAD, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.

[27] J. A. SCOTT, *An Arnoldi code for computing selected eigenvalues of sparse real unsymmetric matrices*, Rep. RAL-93-097, Rutherford Appleton Lab., Oxon, England, December, 1993.

[28] G. L. G. SLEIJPEN AND H. A. VAN DER VORST, *A generalized Jacobi–Davidson iteration method for linear eigenvalue problems*, SIAM. J. Matrix Anal. Appl., 17 (1996), to appear.

[29] D. C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 375–385.

[30] K.-C. TOH AND L. N. TREFETHEN, *Calculation of pseudospectra by the Arnoldi iteration*, Report CTC94TR179, Cornell Theory Center, Cornell University, Ithaca, NY, May, 1994.

[31] A. E. TREFETHEN, L. N. TREFETHEN, AND P. J. SCHMID, *Spectra and pseudospectra for pipe Poiseuille flow*, Report 94-16, Interdisciplinary Project Center for Supercomputing, Swiss Federal Inst. of Technology, Zurich, Switzerland, August, 1994.

[32] L. N. TREFETHEN, *Pseudospectra of matrices*, in Numerical Analysis 1991, D. F. Griffiths and G. A. Watson, eds., Longman Scientific and Technical, Harlow, UK, 1992.

[33] ———, *Spectra and Pseudospectra: The Behavior of Non-Normal Matrices and Operators*, manuscript.

[34] L. N. TREFETHEN, A. E. TREFETHEN, S. C. REDDY, AND T. A. DRISCOLL, *Hydrodynamic stability without eigenvalues*, Science, 261 (1993), pp. 578–584.