SIAM J. SCI. COMPUT.
Vol. 23, No. 2, pp. 583–590

GENERALIZING EIGENVALUE THEOREMS TO PSEUDOSPECTRA THEOREMS*

MARK EMBREE† AND LLOYD N. TREFETHEN†

Abstract. Analysis of nonsymmetric matrix iterations based on eigenvalues can be misleading. In this paper, we discuss sixteen theorems involving ε-pseudospectra that each generalize a familiar eigenvalue theorem and may provide more descriptive information in some cases. Our organizing principle is that each pseudospectral theorem reduces precisely to the corresponding eigenvalue theorem when $\varepsilon = 0$.

Key words. eigenvalues, pseudospectra, matrix iterations

AMS subject classifications. 15A18, 15A60, 65F10, 65F15

PII. S1064827500373012

1. Introduction.

Though we speak of linear algebra, matrix iterative methods belong to the realm of linear analysis. Convergence of errors or residuals to zero is the concern, and this process has meaning because the algebraic problem is embedded in a normed space. Except for questions concerning finite termination, the appropriate tools for analyzing convergence are not the tools of algebra, such as eigenvalues, which are basis-independent, but those of analysis, such as singular values, which are defined via norms and necessarily change with the basis.

In this paper we consider the particular tools of linear analysis known as pseudospectra, which were invented to give information about matrices that lack a well-conditioned basis of eigenvectors. For simplicity, our norm $\|\cdot\|$ will always be the vector 2-norm and the matrix 2-norm that it induces. With this choice of norm, the matrices of interest are those that are far from normal in the sense that their eigenvectors, if a complete set exists, are far from orthogonal. Many of our results can be extended to other norms and also to operators as well as matrices, but we will not discuss these generalizations.

Throughout the article, A is an $N \times N$ matrix, and $\Lambda(A)$ denotes its spectrum, i.e., its set of eigenvalues, a subset of the complex plane $\mathbb{C}$. The pseudospectra of A are nested subsets of $\mathbb{C}$ that expand to fill the plane as $\varepsilon \to \infty$.

DEFINITION. *For each $\varepsilon \geq 0$, the ε-pseudospectrum $\Lambda_\varepsilon(A)$ of A is the set of numbers $z \in \mathbb{C}$ satisfying any of the following equivalent conditions:*

(i) $\|(z - A)^{-1}\| \geq \varepsilon^{-1}$;
(ii) $\sigma_{\min}(z - A) \leq \varepsilon$;
(iii) $\|Au - zu\| \leq \varepsilon$ *for some vector* u *with* $\|u\| = 1$;
(iv) z *is an eigenvalue of* $A + E$ *for some matrix* E *with* $\|E\| \leq \varepsilon$.

Here $\sigma_{\min}$ denotes the smallest singular value, and we employ the convention that $\|(z - A)^{-1}\| = \infty$ for $z \in \Lambda(A)$.

Pseudospectra were introduced as early as 1975 [12] and became a popular tool during the 1990s. We will not give detailed references here, but we refer the reader to [23] and [24] for examples, to [25] for algorithms and a list of applications, and to [26]

*Received by the editors June 5, 2000; accepted for publication (in revised form) September 6, 2000; published electronically July 10, 2001. This work was supported by UK Engineering and Physical Sciences Research Council grant GR/M12414.
http://www.siam.org/journals/sisc/23-2/37301.html
†Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QZ, UK (embree@comlab.ox.ac.uk, LNT@comlab.ox.ac.uk).

for history. For extensive online information about pseudospectra, including examples and a bibliography of papers by many authors, see the Pseudospectra Gateway [3].

This brief article is devoted to a simple idea:

Many theorems about eigenvalues are special cases $\varepsilon = 0$ of theorems about ε-pseudospectra.

Our whole content consists of the presentation of sixteen examples of theorems of this kind. These theorems are for the most part neither mathematically deep nor even new, though in some cases they have not been stated in the language of pseudospectra before. Nevertheless, for practical applications involving highly nonnormal matrices, they may sometimes be more useful than their eigenvalue special cases. This will tend to be so in situations where the eigenvalues of A are misleading, filling a region of $\mathbb{C}$ smaller than where A actually "lives." For an example illustrating the limitations of eigenvalue analysis for Krylov subspace methods for linear systems of algebraic equations, see [6]. Here is another extreme example. If A is nilpotent, with $A^K = 0$ for some $K \geq 1$, then $\Lambda(A) = \{0\}$. Some such matrices will have norms $\|A^k\|$ that diminish steadily toward 0 as $k \to K$, while for others, there may be no reduction until $k = K$ or great transient growth before the eventual decay. Eigenvalues alone cannot distinguish between these behaviors, but pseudospectra can.

Our presentation will adhere to a fixed pattern. In each case, we first list a theorem about eigenvalues, without proof, that is either elementary or well known. We follow this with a generalized theorem for pseudospectra together with an outline of a proof. Some pointers to the literature are included along the way, but we do not aim to be exhaustive, as it is often hard with this essentially elementary material to track down the first appearance of a result in print.

We hope that this article may provide a useful compendium for those concerned with nonnormal matrices and associated iterations, but we emphasize that this collection does not include all potentially useful theorems involving pseudospectra. By confining our attention to theorems that reduce for $\varepsilon = 0$ to valid statements about eigenvalues, we exclude some of the subtler estimates that may be obtained from pseudospectra, notably those based on contour integrals. One example is the Kreiss matrix theorem, which contains a constant eN that does not reduce cleanly to 1 as $\varepsilon \to 0$ [11, 19]. Another is the bound on a polynomial norm $\|p(A)\|$, of immediate relevance to iterations such as GMRES, that can be obtained by integrating $p(z)$ over the boundary contour(s) of $\Lambda_\varepsilon(A)$ [22]. For new results comparing such contour integral techniques to other approaches, see [5].

2. Sixteen theorems. Our first theorem indicates the connection between the ill-conditioning of solving a linear system with A and the existence of a pseudoeigenvalue near the origin. This result has been attributed to Gastinel (see [21, pp. 120, 133], [28, p. 248]).

THEOREM 1. *A is singular $\iff 0 \in \Lambda(A)$.*

THEOREM 1ε. *$\|A^{-1}\| \geq \varepsilon^{-1} \iff 0 \in \Lambda_\varepsilon(A)$.*

Proof. The proof is immediate from the definitions. ∎

Pseudospectra possess the satisfying property that every connected component of the ε-pseudospectrum must contain at least one eigenvalue. This property forms the basis for the following result.

THEOREM 2. *A has N distinct eigenvalues $\implies$ A is diagonalizable.*

THEOREM 2ε. *$\Lambda_\varepsilon(A)$ has N distinct components $\implies$ A is diagonalizable.*

Proof. From definition (iv) of pseudospectra it is clear that for any $\delta > 0$, $\Lambda_\varepsilon(A)$ is contained in the interior of $\Lambda_{\varepsilon+\delta}(A)$. By the continuity of matrix eigenvalues with

respect to perturbations, the same condition (iv) implies that if $\Lambda_\varepsilon(A)$ has N distinct components, so does $\Lambda_{\varepsilon+\delta}(A)$ for sufficiently small $\delta > 0$. Thus we see that $\|(z - A)^{-1}\|$ must achieve local maxima strictly in the interior of each of the N components of $\Lambda_{\varepsilon+\delta}(A)$. Now $\log\|(z - A)^{-1}\|$ is a subharmonic function of z throughout the complex plane except at the eigenvalues of A (see, e.g., [7, Thm. 3.13.1], [4]), and thus $\log\|(z-A)^{-1}\|$ and likewise $\|(z-A)^{-1}\|$ satisfy the maximum principle away from the eigenvalues of A. Putting these facts together, we see that each component of $\Lambda_\varepsilon(A)$ must contain an eigenvalue of A, which implies that A has N distinct eigenvalues and thus is diagonalizable. □

Gallestey has developed an algorithm for computing pseudospectra based on the maximum principle property used in the above proof [4]. A simpler exclusion algorithm, recently proposed by Koutis and Gallopoulos [10], is based upon the next result.

THEOREM 3. $\|(z - A)^{-1}\| \ \geq \ \frac{1}{\mathrm{dist}(z,\Lambda(A))}$.

THEOREM 3ε. $\|(z - A)^{-1}\| \ \geq \ \frac{1}{\mathrm{dist}(z,\Lambda_\varepsilon(A))+\varepsilon}$.

Proof. A perturbation of A of norm $\mathrm{dist}(z, \Lambda_\varepsilon(A)) + \varepsilon$ could make z an eigenvalue. □

The Koutis–Gallopoulos algorithm utilizes Theorem 3ε rewritten in the form

$$\mathrm{dist}(z, \Lambda_\varepsilon(A)) \ \geq \ \frac{1}{\|(z - A)^{-1}\|} - \varepsilon.$$

In our next theorem, S is an arbitrary nonsingular matrix and $\kappa(S)$ is its condition number, $\kappa(S) \equiv \|S\|\|S^{-1}\|$. Though the theorem is stated as an inclusion in one direction only, it applies in the other direction too, and in that sense Theorem 4 maintains our usual pattern of being the special case $\varepsilon = 0$ of Theorem 4ε. The result demonstrates that pseudospectra are invariant under unitary transformations, and also reflects the extent to which an ill-conditioned similarity transformation can alter pseudospectra. When B is diagonal, so that SBS^{-1} represents a diagonalization of A, Theorem 4ε is equivalent to the most familiar version of the Bauer–Fike theorem [1].

THEOREM 4. $A = SBS^{-1} \implies \Lambda(A) = \Lambda(B)$.

THEOREM 4ε. $A = SBS^{-1} \implies \Lambda_\varepsilon(A) \subseteq \Lambda_{\kappa(S)\varepsilon}(B)$.

Proof. Since $(z - A)^{-1} = S(z - B)^{-1}S^{-1}$, $\|(z - A)^{-1}\| \leq \kappa(S)\|(z - B)^{-1}\|$. Therefore if $\|(z - A)^{-1}\| \geq \varepsilon^{-1}$, then $\|(z - B)^{-1}\| \geq (\kappa(S)\varepsilon)^{-1}$. □

The following theorem makes use of the idea of the "average pseudoeigenvalue" of a matrix, $\mathrm{mean}_{\lambda_\varepsilon\in\Lambda_\varepsilon(A)}\lambda_\varepsilon$. Of course, this quantity needs to be defined. We could be very specific and make use of, say, Haar measure (isotropy in $\mathbb{C}^N$) on the space of $N \times N$ matrices, but for the purposes of this theorem it is enough to say that $\mathrm{mean}_{\lambda_\varepsilon\in\Lambda_\varepsilon(A)}\lambda_\varepsilon$ is the mean of the eigenvalues of $A + E$ averaged over any fixed distribution on the matrices E with $\|E\| \leq \varepsilon$ with the property that each matrix entry e_{ij} has mean 0.

THEOREM 5. $\mathrm{tr}(A) \ = \ N \cdot \mathrm{mean}_{\lambda\in\Lambda(A)}\lambda$.

THEOREM 5ε. $\mathrm{tr}(A) \ = \ N \cdot \mathrm{mean}_{\lambda_\varepsilon\in\Lambda_\varepsilon(A)}\lambda_\varepsilon$.

Proof. The theorem looks deep but is elementary. All we need to do is consider traces of perturbed matrices. Since each e_{jj} has mean 0 by assumption, so does their sum, and thus $\mathrm{tr}(A) = \mathrm{mean}_{\|E\|\leq\varepsilon}\mathrm{tr}(A + E) = N \cdot \mathrm{mean}_{\lambda_\varepsilon\in\Lambda_\varepsilon(A)}\lambda_\varepsilon$. □

Our next pair of results requires a definition of the *condition number* $\kappa_A(\Sigma(A))$ of a compact set $\Sigma = \Sigma(A) \subset \mathbb{C}$ depending on A with respect to perturbations of A. If Σ_1 and Σ_2 are compact subsets of $\mathbb{C}$, let $d(\Sigma_1, \Sigma_2)$ denote the Hausdorff

distance $d(\Sigma_1, \Sigma_2) = \max\{\max_{s\in\Sigma_1} d(s, \Sigma_2), \max_{s\in\Sigma_2} d(s, \Sigma_1)\}$, where $d(s, \Sigma)$ is the usual distance of a point s to a set Σ. Then

$$\kappa_A(\Sigma(A)) \equiv \limsup_{\delta\to 0} \left(\delta^{-1} \sup_{\|D\|\le\delta} d(\Sigma(A+D), \Sigma(A)) \right).$$

THEOREM 6. *$\Lambda(A)$ depends continuously on A, with condition number $\kappa_A(\Lambda(A)) = 1$ if A is normal.*

THEOREM 6ε. *$\Lambda_\varepsilon(A)$ depends continuously on A, with condition number $\kappa_A(\Lambda_\varepsilon(A)) = 1$ if A is normal.*

Proof. The continuity of $\Lambda_\varepsilon(A)$ in the Hausdorff metric follows from the analogous continuity of $\Lambda(A)$. Suppose now that A is normal, so that its ε-pseudospectrum is the union of ε-disks centered at each eigenvalue. For any $\delta \ge 0$ and $D \in \mathbb{C}^{N\times N}$ with $\|D\| \le \delta$, we have $\max_{s\in\Lambda_\varepsilon(A+D)} d(s, \Lambda_\varepsilon(A)) \le \delta$ and similarly $\max_{s\in\Lambda_\varepsilon(A)} d(s, \Lambda_\varepsilon(A+D)) \le \delta$. Thus,

$$\sup_{\|D\|\le\delta} d(\Lambda_\varepsilon(A+D), \Lambda_\varepsilon(A)) \le \delta.$$

Since $\cup_{\|D\|\le\delta}\Lambda_\varepsilon(A+D) = \Lambda_{\varepsilon+\delta}(A)$, there always exists some D with $\|D\| \le \delta$ such that $\max_{s\in\Lambda_\varepsilon(A+D)} d(s, \Lambda_\varepsilon(A)) = \delta$, and for such a D we must have $\|D\| = \delta$, since the pseudospectra are strictly nested sets. It follows that

$$\kappa_A(\Lambda_\varepsilon(A)) \equiv \limsup_{\delta\to 0} \left(\delta^{-1} \sup_{\|D\|\le\delta} d(\Lambda_\varepsilon(A+D), \Lambda_\varepsilon(A)) \right) = 1. \qquad \square$$

Eigenvalues can change dramatically with small perturbations, a warning that analysis based on them can be misleading. The following theorem hints that pseudospectra may be more robust.

THEOREM 7. $\Lambda(A+E) \subseteq \Lambda_{\|E\|}(A)$.

THEOREM 7ε. $\Lambda_\varepsilon(A+E) \subseteq \Lambda_{\varepsilon+\|E\|}(A)$.

Proof. If $z \in \Lambda_\varepsilon(A+E)$, then there exists a matrix F with $\|F\| \le \varepsilon$ such that $(A+E+F)u = zu$ for some $u \ne 0$. Since $\|E+F\| \le \varepsilon + \|E\|$, $z \in \Lambda_{\varepsilon+\|E\|}(A)$. $\square$

We now turn to the problems of estimating the behavior of a matrix from its spectrum and pseudospectra.

THEOREM 8. $\lambda \in \Lambda(A) \implies \|A\| \ge |\lambda|$.

THEOREM 8ε. $\lambda_\varepsilon \in \Lambda_\varepsilon(A) \implies \|A\| \ge |\lambda_\varepsilon| - \varepsilon$.

Proof. If $\lambda_\varepsilon \in \Lambda_\varepsilon(A)$, then $Au = \lambda_\varepsilon u + \varepsilon v$ for some vectors $u, v \in \mathbb{C}$ with $\|u\| = \|v\| = 1$. It follows that $\|Au\| \ge |\lambda_\varepsilon| - \varepsilon$. $\square$

The convergence analysis of stationary iterative methods is based on the behavior of powers of the iteration matrix. It has long been known that transient growth can occur even when the spectral radius of the iteration matrix is less than one (see, e.g., [27, p. 63]). The following two theorems use pseudospectra to describe this transient growth. The first is the "easy half of the Kreiss matrix theorem," that is, the half of that theorem that does not depend on N and whose proof is elementary [11].

THEOREM 9. $\max_{\lambda\in\Lambda(A)} |\lambda| > 1 \implies \sup_{k>0} \|A^k\| = \infty$.

THEOREM 9ε. $\max_{\lambda_\varepsilon\in\Lambda_\varepsilon(A)} |\lambda_\varepsilon| > 1 + C\varepsilon \implies \sup_{k\ge 0} \|A^k\| > C$.

Proof. Since $\|A^0\| = 1$, the result is trivial for $C < 1$, so assume $C \ge 1$. If $\max_{\lambda\in\Lambda(A)} |\lambda| > 1$, then the conclusion certainly holds, so assume $\max_{\lambda\in\Lambda(A)} |\lambda| \le 1$, in which case we have the convergent series representation

$$(z-A)^{-1} = z^{-1}(I + z^{-1}A + z^{-2}A^2 + \cdots),$$

which is valid for all z with $|z| > 1$. We now argue the contrapositive. If $\|A^k\| \le C$ for all $k \ge 0$, then

$$\|(z-A)^{-1}\| \;\le\; \frac{|z^{-1}|C}{1-|z^{-1}|} \;=\; \frac{C}{|z|-1}$$

for any z with $|z| > 1$. This implies that $\Lambda_\varepsilon(A)$ is contained in the disk about the origin of radius $1 + C\varepsilon$, i.e., $\max_{\lambda_\varepsilon \in \Lambda_\varepsilon(A)} |\lambda_\varepsilon| \le 1 + C\varepsilon$. ∎

THEOREM 10. $\lambda \in \Lambda(A) \implies \|A^k\| \ge |\lambda|^k$ *for all* k.

THEOREM 10ε. $\lambda_\varepsilon \in \Lambda_\varepsilon(A) \implies \|A^k\| \ge |\lambda_\varepsilon|^k - \frac{k\varepsilon\|A\|^{k-1}}{1-k\varepsilon/\|A\|}$ *for all* k *such that* $k\varepsilon < \|A\|$.

Proof. Pick E such that $\|E\| \le \varepsilon$ and $\lambda_\varepsilon \in \Lambda(A+E)$. Then $\|(A+E)^k\| \ge |\lambda_\varepsilon|^k$, which implies

$$\begin{aligned}\|A^k\| \;&\ge\; |\lambda_\varepsilon|^k - k\varepsilon\|A\|^{k-1} - \binom{k}{2}\varepsilon^2\|A\|^{k-2} - \cdots \\ &\ge\; |\lambda_\varepsilon|^k - k\varepsilon\|A\|^{k-1}\left(1 + k\varepsilon/\|A\| + (k\varepsilon)^2/\|A\|^2 + \cdots\right).\end{aligned}$$

Provided $k\varepsilon < \|A\|$, the series in this last equation converges, giving

$$\|A^k\| \ge |\lambda_\varepsilon|^k - \frac{k\varepsilon\|A\|^{k-1}}{1-k\varepsilon/\|A\|}.$$ ∎

Theorems 9 and 9ε have exact analogues for continuous time (see [14, 15]).

THEOREM 11. $\max_{\lambda\in\Lambda(A)} \mathrm{Re}\,\lambda > 0 \implies \sup_{t>0} \|e^{tA}\| = \infty$.

THEOREM 11ε. $\max_{\lambda_\varepsilon\in\Lambda_\varepsilon(A)} \mathrm{Re}\,\lambda_\varepsilon > C\varepsilon \implies \sup_{t>0} \|e^{tA}\| > C$.

Proof. As in the proof of Theorem 9ε, the conclusion is immediate if $C < 1$ or if $\max_{\lambda\in\Lambda(A)} \mathrm{Re}\,\lambda > 0$, so we assume that $C \ge 1$ and $\max_{\lambda\in\Lambda(A)} \mathrm{Re}\,\lambda \le 0$ and use the Laplace transform identity

$$(z-A)^{-1} \;=\; \int_0^\infty e^{-zt}e^{tA}dt,$$

which is valid for $\mathrm{Re}\,z > 0$. Again arguing the contrapositive, we note that if $\|e^{tA}\| \le C$ for all $t > 0$, then $\|(z-A)^{-1}\| \;\le\; C/\mathrm{Re}\,z$ for z with $\mathrm{Re}\,z > 0$, implying that $\Lambda_\varepsilon(A)$ is contained in the half-plane defined by $\mathrm{Re}\,z \le C\varepsilon$. ∎

Our next result is a pseudospectral generalization of Gerschgorin's theorem, which we believe to be new. It implies that if $\Lambda_\varepsilon(A)$ contains points distant from $\Lambda(A)$ for sufficiently small ε, then the bounds given by Gerschgorin's theorem will be more sharply descriptive of the pseudospectra than of the spectrum. Coupling this with Theorems 9ε and 10ε, one sees that Gerschgorin eigenvalue estimates may sometimes lead to more accurate predictions of transient behavior of iterative matrix processes than would be obtained from the exact eigenvalues! As has been pointed out in [13], this curious robustness phenomenon is of practical importance, for it sheds light on how it is that iterations such as GMRES may sometimes converge handily even when the associated Ritz values or harmonic Ritz values are far from accurate eigenvalue estimates. For these theorems, define $d_j = a_{jj}$ and $r_j = \sum_{k\ne j} |a_{jk}|$, and for any complex number z and real number $r \ge 0$, let $D(z,r)$ denote the closed disk about z of radius r.

THEOREM 12. $\Lambda(A) \;\subseteq\; \bigcup_j D(d_j, r_j)$.

THEOREM 12ε. $\Lambda_\varepsilon(A) \;\subseteq\; \bigcup_j D(d_j, r_j + \sqrt{N}\varepsilon)$.

Proof. Applying Gerschgorin's theorem to $A + E$ with $\|E\| \le \varepsilon$ yields inclusion disks centered at $d_j + e_{jj}$ with radius $\sum_{k\neq j} |a_{jk} + e_{jk}| \le r_j + \sum_{k\neq j} |e_{jk}|$. Each such disk is contained in the disk centered at d_j with radius $r_j + \sum_{k=1}^{N} |e_{jk}| = r_j + \|E_j\|_\infty$, where E_j denotes the matrix equal to E in the jth row and zero elsewhere. The term $\sqrt{N}\varepsilon$ comes from the inequality $\|E_j\|_\infty \le \sqrt{N}\|E_j\|_2 \le \sqrt{N}\|E\|_2$. ∎

The next result concerns the numerical range or field of values, which we denote by $W(A)$. In the context of iterative methods, the theorem indicates how analysis based on the field of values (see, e.g., [2]) relates to pseudospectral analysis. We write $\mathrm{conv}(S)$ for the convex hull in $\mathbb{C}$ of a set $S \subseteq \mathbb{C}$. The notation "$S \setminus \varepsilon$-border" also requires some explanation. By this we mean the set of points $z \in \mathbb{C}$ such that $D(z, \varepsilon) \subseteq S$. Perhaps Reddy, Schmid, and Henningson were the first to formulate this result in the language of pseudospectra [15, Thm. 2.2].

THEOREM 13. $W(A) \supseteq \mathrm{conv}(\Lambda(A))$.

THEOREM 13ε. $W(A) \supseteq \mathrm{conv}(\Lambda_\varepsilon(A)) \setminus \varepsilon$-border.

Proof. This result follows from a familiar result in functional analysis: that $W(A)$ is the intersection of all convex sets S that satisfy the condition

$$\|(z - A)^{-1}\| \le \frac{1}{\mathrm{dist}(z, S)}.$$

See, for example, Kato [9, p. 268]. ∎

The spectral mapping theorem (see, e.g., [9, p. 45]) is a jewel in the crown of eigenvalue theorems; it is theoretically appealing and practically relevant, forming the basis for rational transformation techniques for computing eigenvalues. The numerical range obeys a similar, though one-sided, mapping theorem [8]. Theorems 13 and 13ε suggest that a similar result might hold for pseudospectra. Our next theorem is a modest step in this direction, a precise mapping theorem for linear transformations [26, Thm. 2.4].

THEOREM 14. $\Lambda(\alpha + \beta A) = \alpha + \beta\Lambda(A)$ *for* $\alpha, \beta \in \mathbb{C}$.

THEOREM 14ε. $\Lambda_{\varepsilon|\beta|}(\alpha + \beta A) = \alpha + \beta\Lambda_\varepsilon(A)$ *for* $\alpha, \beta \in \mathbb{C}$.

Proof. The result is trivial when $\beta = 0$. Otherwise, note that

$$|\beta|\,\|(z - (\alpha + \beta A))^{-1}\| = \|(\beta^{-1}(z - \alpha) - A)^{-1}\|. \quad ∎$$

For Theorems 15 and 16, let V denote an $N \times k$ rectangular matrix with orthonormal columns for some $k \le N$, as might be obtained by Arnoldi or subspace iteration, and let H denote a $k \times k$ square matrix. In the Arnoldi iteration, H would have Hessenberg form, but this is not necessary for these theorems. First, we assume that the columns of V exactly span an invariant subspace of A. The resulting theorem forms the basis for algorithms that compute pseudospectra by projecting A onto a carefully chosen invariant subspace [15, 25, 29].

THEOREM 15. $AV = VH \implies \Lambda(H) \subseteq \Lambda(A)$.

THEOREM 15ε. $AV = VH \implies \Lambda_\varepsilon(H) \subseteq \Lambda_\varepsilon(A)$.

Proof. If $\|Hu - zu\| \le \varepsilon$ for some $u \in \mathbb{C}^N$ with $\|u\| = 1$, then $\|VHu - Vzu\| \le \varepsilon$ too, and this implies $\|AVu - zVu\| \le \varepsilon$. ∎

Practical algorithms such as the implicitly restarted Arnoldi method [18] or subspace iteration (see, e.g., [16, section V.1]) may not easily yield an exact basis for the invariant subspace. Rather, the columns of V form an orthonormal basis for some *approximate* invariant subspace of A. Let H denote the generalized Rayleigh quotient this basis forms, $H \equiv V^*AV$. With this notation, eigenvalue Theorem 15

has an alternative, more practical pseudospectral generalization. This theorem is a fundamental result in the perturbation theory of invariant subspaces; see [20] and references therein.

THEOREM 16. $AV = VH \implies \Lambda(H) \subseteq \Lambda(A)$.

THEOREM 16ε. $AV = VH + R \implies \Lambda(H) \subseteq \Lambda_\varepsilon(A)$ *for* $\varepsilon = \|R\|$.

Proof. Consider the square matrix $E = -RV^*$. Then $(A+E)V = AV - R = VH$, so by Theorem 15, the eigenvalues of H are eigenvalues of $A+E$ and hence ε-pseudoeigenvalues of A for $\varepsilon = \|-RV^*\| = \|R\|$. ∎

For an Arnoldi factorization with k basis vectors, $V \in \mathbb{C}^{n\times k}$, Theorem 16$\varepsilon$ reduces to a well-known result: $\varepsilon = \|R\| = |h_{k+1,k}|$, where $h_{k+1,k}$ is the $(k+1,k)$ entry in the extended upper Hessenberg matrix (see, e.g., [17, Lem. 2.1]).

Acknowledgments. We thank Anne Greenbaum for carefully reading an earlier version of this manuscript and pointing out several corrections. We also appreciate helpful comments from Albrecht Böttcher, Nick Higham, and the referees. We especially thank Yvan Notay for noting that the constant in Theorem 12ε could be improved from N to $\sqrt{N}$.

REFERENCES

[1] F. L. Bauer and C. T. Fike, *Norms and exclusion theorems*, Numer. Math., 2 (1960), pp. 137–141.

[2] M. Eiermann, *Fields of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.

[3] M. Embree and L. N. Trefethen, *Pseudospectra Gateway*, http://www.comlab.ox.ac.uk/pseudospectra, 2000.

[4] E. Gallestey, *Computing spectral value sets using the subharmonicity of the norm of rational matrices*, BIT, 38 (1998), pp. 22–33.

[5] A. Greenbaum, *Using the Cauchy integral formula and the partial fractions decomposition of the resolvent to estimate* $\|f(A)\|$, SIAM J. Matrix Anal., submitted, 2000.

[6] A. Greenbaum, V. Pták, and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.

[7] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, revised ed., AMS, Providence, RI, 1957.

[8] T. Kato, *Some mapping theorems for the numerical range*, Proc. Japan Acad., 41 (1965), pp. 652–655.

[9] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, New York, 1980.

[10] I. Koutis and E. Gallopoulos, *Exclusion Regions and Fast Estimation of Pseudospectra*, Tech. report, Department of Computer Engineering and Informatics, HPCLAB, University of Patras, Patras, Greece, 2000.

[11] H.-O. Kreiss, *Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differentialgleichungen approximieren*, BIT, 2 (1962), pp. 153–181.

[12] H. J. Landau, *On Szegő's eigenvalue distribution theory and non-Hermitian kernels*, J. Analyse Math., 28 (1975), pp. 335–357.

[13] 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. 796–825.

[14] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, New York, 1983.

[15] 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.

[16] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, UK, 1992.

[17] V. Simoncini and E. Gallopoulos, *Transfer functions and resolvent norm approximation of large matrices*, Electron. Trans. Numer. Anal., 7 (1998), pp. 190–201.

[18] D. C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[19] M. N. Spijker, *On a conjecture of LeVeque and Trefethen related to the Kreiss matrix theorem*, BIT, 31 (1991), pp. 551–555.

[20] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.
[21] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
[22] 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, pp. 336–360.
[23] 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, pp. 234–266.
[24] L. N. Trefethen, *Pseudospectra of linear operators*, SIAM Rev., 39 (1997), pp. 383–406.
[25] L. N. Trefethen, *Computation of pseudospectra*, in Acta Numer. 8, Cambridge University Press, Cambridge, UK, 1999, pp. 247–295.
[26] L. N. Trefethen, *Spectra and pseudospectra: The behavior of non-normal matrices and operators*, in The Graduate Student's Guide to Numerical Analysis, M. Ainsworth, J. Levesley, and M. Marletta, eds., Springer-Verlag, Berlin, 1999, pp. 217–250.
[27] R. S. Varga, *Matrix Iterative Analysis*, Prentice–Hall, Englewood Cliffs, NJ, 1962. Second edition published by Springer-Verlag, Berlin, 2000.
[28] J. H. Wilkinson, *Sensitivity of eigenvalues* II, Util. Math., 30 (1986), pp. 243–286.
[29] T. G. Wright and L. N. Trefethen, *Large-Scale Computation of Pseudospectra Using ARPACK and Eigs*, Tech. report 00/11, Oxford University Computing Laboratory Numerical Analysis Group, Oxford, UK, 2000, SIAM J. Sci. Comput., 23 (2001), pp. 591–605.