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 \(\varepsilon\)-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

1. Introduction. Though we speak of linear algebra, 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 \(\varepsilon\)-pseudospectrum \(\Lambda_{\varepsilon}(A)\) of \(A\) is the set of numbers \(z \in \mathbb{C}\) satisfying any of the equivalent conditions:

(i) \(\| (z - A)^{-1} \| \geq \varepsilon^{-1}\);
(ii) \(\sigma_{\min}(z - A) \leq \varepsilon\);
(iii) \(\| Au - z u \| \leq \varepsilon\) 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 [11] and became a popular tool during the 1990s. We will not give detailed references here but refer the reader to [20] and [21] for examples, to [22] for algorithms and a list of applications, and to [23] for history. For extensive online information about pseudospectra, including examples and a bibliography of papers by many authors, see the Web site [4].

This brief article is devoted to a simple idea:
Many theorems about eigenvalues are special cases $\varepsilon = 0$ of theorems about $\varepsilon$-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 non-normal 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 $C$ smaller than where $A$ actually "lives." For an example illustrating the limitations of eigenvalue analysis for Krylov subspace methods for linear systems, see [7]. Consider the following simpler but 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 non-normal 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 $\varepsilon N$ that does not reduce cleanly to 1 as $\varepsilon \to 0$ [10, 17]. Another is the bound on a polynomial norm $\|p(A)\|$, of immediate relevance to iterations such as GMRES, that can be obtained by integrating $p(\varepsilon)$ over the boundary contour(s) of $\Lambda_\varepsilon(A)$ [19]. For new results comparing such contour integral techniques to other approaches, see [6].

2. Sixteen Theorems. Our first theorem states that ill-conditioning is equivalent to the existence of a pseudoeigenvalue near the origin. The result has been attributed to Gastinel (see [18, pp. 120, 133]).

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

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

**Proof.** Immediate from the definitions. $\square$

Pseudospectra possess the satisfying property that every connected component of the $\varepsilon$-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.** The function $\log \|(z - A)^{-1}\|$ is subharmonic, which implies that it satisfies a maximum principle, and therefore so does the resolvent norm $\|(z - A)^{-1}\|$. This implies that each component of $\Lambda_\varepsilon(A)$ contains at least one eigenvalue of $A$. $\square$
Gallestey has developed an algorithm for computing pseudospectra based on the
maximum principle property used in the above proof [5]. A simpler exclusion algo-
rithm, recently proposed by Gallopoulos, is based upon the next result [2].

**Theorem 3.** \( \| (z - A)^{-1} \| \geq \frac{1}{\text{dist}(z, \Lambda(A))} \).

**Theorem 3\( \varepsilon \).** \( \| (z - A)^{-1} \| \geq \frac{1}{\text{dist}(z, \Lambda_\varepsilon(A)) + \varepsilon} \).

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

Gallopoulos’s algorithm utilizes Theorem 3\( \varepsilon \) rewritten in the form

\[
\text{dist}(z, \Lambda_\varepsilon(A)) \geq \frac{1}{\| (z - A)^{-1} \|} - \varepsilon.
\]

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 4\( \varepsilon \). 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 \( SBS^{-1} \) is a diagonalization of \( A \), Theorem 4\( \varepsilon \) is equivalent to
the most familiar version of the Bauer–Fike Theorem [1].

**Theorem 4.** \( A = SBS^{-1} \Rightarrow \Lambda(A) = \Lambda(B) \).

**Theorem 4\( \varepsilon \).** \( A = SBS^{-1} \Rightarrow \Lambda_\varepsilon(A) \subseteq \Lambda_\varepsilon(S)(B) \).

**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, \( \text{mean}_{\lambda \in \Lambda_\varepsilon(A)} \lambda \). 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
\( \text{mean}_{\lambda \in \Lambda_\varepsilon(A)} \lambda \) 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
column entry \( e_{ij} \) has mean 0.

**Theorem 5.** \( \text{tr}(A) = N \cdot \text{mean}_{\lambda \in \Lambda_\varepsilon(A)} \lambda \).

**Theorem 5\( \varepsilon \).** \( \text{tr}(A) = N \cdot \text{mean}_{\lambda \in \Lambda_\varepsilon(A)} \lambda \).

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

Our next pair of results require a definition of the *condition number* \( \kappa_A(\Sigma) \)
of a set \( \Sigma = \Sigma(A) \) depending on \( A \) with respect to perturbations of \( A \). If \( \Sigma_1 \)
and \( \Sigma_2 \) are generic sets, let \( d(\Sigma_1, \Sigma_2) \) denote the Hausdorff distance, \( d(\Sigma_1, \Sigma_2) = \max_{s \in \Sigma_1} \max_{s \in \Sigma_2} d(s, \Sigma_1) \), where \( d(s, \Sigma) \) is the usual distance of a point \( s \) to a
set \( \Sigma \). Then

\[
\kappa_A(\Sigma) \equiv \limsup_{\varepsilon \to 0} \left( \varepsilon^{-1} \sup_{\| E \| = \varepsilon} d(\Sigma(A + E), \Sigma(A)) \right).
\]
Theorem 6. \( \Lambda(A) \) depends continuously on \( A \), with condition number 1 if \( A \) is normal.

Theorem 6\( \epsilon \). \( \Lambda_{\epsilon}(A) \) depends continuously on \( A \), with condition number 1 if \( A \) is normal.

Proof. This follows from the definitions of \( \Lambda_{\epsilon}(A) \) and the condition number, together with the continuity of eigenvalues with respect to perturbations of the entries.

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 7\( \epsilon \). \( \Lambda_{\epsilon}(A + E) \subseteq \Lambda_{\epsilon + \|E\|}(A) \).

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

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

Theorem 8. \( \lambda \in \Lambda(A) \implies \|A\| \geq |\lambda| \).

Theorem 8\( \epsilon \). \( \lambda_{\epsilon} \in \Lambda_{\epsilon}(A) \implies \|A\| \geq |\lambda_{\epsilon}| - \epsilon \).

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

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., [24]). 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 [10].

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

Theorem 9\( \epsilon \). \( \max_{\lambda \in \Lambda_{\epsilon}(A)} |\lambda_{\epsilon}| > 1 + C\epsilon \implies \sup_{k \geq 0} \|A^k\| > C \).

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

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

valid for all \( z \) with \( |z| > 1 \). We now argue the contrapositive. If \( \|A^k\| \leq C \) for all \( k \geq 0 \), then

\[
\|(z - A)^{-1}\| \leq \frac{|z^{-1}|C}{1 - |z^{-1}|} = \frac{C}{|z| - 1}
\]

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

Theorem 10. \( \lambda \in \Lambda(A) \implies \|A^k\| \geq |\lambda|^k \) for all \( k \).

Theorem 10\( \epsilon \). \( \lambda_{\epsilon} \in \Lambda_{\epsilon}(A) \implies \|A^k\| \geq |\lambda_{\epsilon}|^k - \frac{k\epsilon \|A\|^{k-1}}{1 - k\epsilon \|A\|} \) for all \( k \) such that \( k\epsilon < \|A\| \).
Proof. Pick $E$ such that $\|E\| \leq \varepsilon$ and $\lambda_\varepsilon \in \Lambda_\varepsilon(A + E)$. Then $\|(A + E)^k\| \geq |\lambda_\varepsilon|^k$, which implies

$$\begin{align*}
\|A^k\| &\geq |\lambda_\varepsilon|^k - k\varepsilon\|A\|^k - \left(\frac{k}{2}\right)\varepsilon^2\|A\|^{k-1} - \cdots \\
&\geq |\lambda_\varepsilon|^k - k\varepsilon\|A\|^k - (1 + k\varepsilon/\|A\| + (k\varepsilon^2/\|A\|^2 + \cdots).
\end{align*}$$

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

$$\|A^k\| \geq |\lambda_\varepsilon|^k - \frac{k\varepsilon\|A\|^k}{1 - k\varepsilon/\|A\|}.$$
write \( \text{conv}(S) \) 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 Hennington were the first to formulate this result in the language of pseudospectra [13, Thm. 2.2].

**Theorem 13.** \( W(A) \supseteq \text{conv}(\Lambda(A)) \).

**Theorem 13c.** \( W(A) \supseteq \text{conv}(\Lambda_\varepsilon(A)) \setminus \varepsilon\) border.

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

\[
||z - A||^{-1} \leq \frac{1}{\text{dist}(z, S)}.
\]

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 13c 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 [23, Thm. 2.4].

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

**Theorem 14c.** \( \Lambda_\varepsilon(\alpha + \beta A) = \alpha + \beta \Lambda_\varepsilon(A) \) for \( \alpha, \beta \in \mathbb{C} \).

**Proof.** The result is trivial when \( \beta = 0 \). Otherwise, note that

\[
|\beta|||z - (\alpha + \beta A)|^{-1}|| = ||\beta^{-1}(z - \alpha) - A|^{-1}||.
\]

For Theorems 15 and 16, let \( V \) denote an \( N \times k \) rectangular matrix with orthonormal columns, for some \( k \leq N \), as might be obtained in 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 [13, 22].

**Theorem 15.** \( AV = VH \implies \Lambda(H) \subseteq \Lambda(A) \).

**Theorem 15c.** \( AV = VH \implies \Lambda_\varepsilon(H) \subseteq \Lambda_\varepsilon(A) \).

**Proof.** If \( ||Hu - zu|| \leq \varepsilon \) for some \( u \in \mathbb{C}^N \) with \( ||u|| = 1 \), then \( ||VHu - Vzu|| \leq \varepsilon \) too, and this implies \( ||AVu - \varepsilon Vzu|| \leq \varepsilon \).

Practical algorithms such as the implicitly restarted Arnoldi method [16] or subspace iteration (see, e.g., [14, §V.1]) rarely 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.

**Theorem 16.** \( AV = VH \implies \Lambda(H) \subseteq \Lambda(A) \).

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

**Proof.** Consider the square matrix \( E = AV - VH \), so by Theorem 15, the eigenvalues of \( H \) are eigenvalues of \( A + E \) and hence \( \varepsilon\)-pseudoeigenvalues of \( A \) for \( \varepsilon = ||R|| = ||E|| \).
PSEUDOSPECTRA THEOREMS

For an Arnoldi factorization with \( k \) basis vectors, \( V \in \mathbb{C}^{n \times k} \), Theorem 16ε 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., [15, Lem. 2.1]).

Acknowledgment. We thank Anne Greenbaum for carefully reading an earlier version of this manuscript and pointing out several corrections.

REFERENCES