Spectral conditioning and pseudospectral growth

J. V. Burke · A. S. Lewis · M. L. Overton

Abstract Using the language of pseudospectra, we study the behavior of matrix eigenvalues under two scales of matrix perturbation. First, we relate Lidskii's analysis of small perturbations to a recent result of Karow on the growth rate of pseudospectra. Then, considering larger perturbations, we follow recent work of Alam and Bora in characterizing the distance from a given matrix to the set of matrices with multiple eigenvalues in terms of the number of connected components of pseudospectra.

Mathematics Subject Classification (2000) 15A18 · 65F15

1 Introduction

Consider a fixed n-by-n complex matrix A. We denote the smallest singular value of A by $\underline{\sigma}(A)$. An eigenvalue of A is *simple* if it has algebraic multiplicity one, and

J. V. Burke
Department of Mathematics, University of Washington,
Seattle, WA 98195, USA
e-mail: burke@math.washington.edu

A. S. Lewis (✉)
School of Operations Research and Industrial Engineering,
Cornell University, Ithaca, NY 14853, USA
e-mail: aslewis@orie.cornell.edu
URL: people.orie.cornell.edu/~aslewis

M. L. Overton
Courant Institute of Mathematical Sciences,
New York University, New York, NY 10012, USA
e-mail: overton@cs.nyu.edu

multiple otherwise; the eigenvalue is *nonderogatory* if its *geometric* multiplicity is one, and *derogatory* otherwise. An eigenvalue is *semisimple* if its geometric and algebraic multiplicities coincide.

For $\epsilon \geq 0$, the ϵ-*pseudospectrum* of A is the set

$$\Lambda_\epsilon(A) = \{w \in \mathbf{C} : \underline{\sigma}(A - wI) \leq \epsilon\}. \tag{1.1}$$

This definition is slightly different from that of [16], the canonical reference on pseudospectra: the definition there corresponds to what we call the "strict pseudospectrum" (see Sect. 3). Clearly $\Lambda_0(A)$ is just the spectrum of A: we denote it simply $\Lambda(A)$. Pseudospectra are just sets of all eigenvalues of matrices within a given distance:

$$\Lambda_\epsilon(A) = \bigcup_{\|X-A\| \leq \epsilon} \Lambda(X). \tag{1.2}$$

These definitions are for the spectral norm, to which we restrict attention.

A simple application of the maximum modulus principle shows that each connected component of $\Lambda_\epsilon(A)$ contains an eigenvalue of A. Consider a particular eigenvalue z of A, and denote the pseudospectral component containing z by $\Lambda_\epsilon^z(A)$. For small ϵ, distinct eigenvalues correspond to distinct components: in Wilkinson's terminology [18], each component then consists of all possible "successors" of the corresponding eigenvalue associated with perturbations to A of size no more than ϵ. As pointed out in [18], for larger ϵ, components coalesce, and the notion of "successor" is more subtle.

Any simple eigenvalue z is an analytic function, locally, of the underlying matrix A; the gradient of this function is vu^* for corresponding unit left and right eigenvectors u and v. Consequently, for small $\epsilon \geq 0$, the pseudospectral component $\Lambda_\epsilon^z(A)$ is approximately a ball centered at z of radius $|u^*v|^{-1}\epsilon$, and hence grows roughly linearly in ϵ at rate $|u^*v|^{-1}$, the classical "condition number" of the eigenvalue of z. Furthermore, for small ϵ, the component $\Lambda_\epsilon^z(A)$ is a convex neighborhood of z [14]. This convexity property holds more generally for nonderogatory eigenvalues, but may fail in general [4].

We study here two questions on the growth of pseudospectral components as ϵ increases.

- What is the growth rate of components corresponding to *multiple* eigenvalues?
- How far can pseudospectral components grow before coalescing?

We answer the first question using Lidskii's perturbation theory for eigenvalues with arbitrary Jordan structure [12], as reviewed in [13]. The latter work introduced a "Hölder condition number" for eigenvalues, which, following recent work of Karow [10], we re-interpret in terms of pseudospectral growth rate. For the second question, we follow the philosophy of recent work of Alam and Bora [1] in relating the coalescence of pseudospectral components to the problem of finding the distance from A to the set of matrices with multiple eigenvalues.

The results we describe here are not new, but we think they deserve a wide audience. Our aim here is to highlight their interest and relationships with a fresh and concise presentation.

2 Lidskii's perturbation theory

Consider an eigenvalue z of the matrix A. As we mentioned in the previous section, the classical condition number of this eigenvalue, namely

$$\kappa = \limsup_{A \neq W \to A, \; w \to z, \; w \in \Lambda(W)} \frac{|w - z|}{\|W - A\|},$$

describes the sensitivity of the eigenvalue to small changes in A. We can describe it alternatively as the infimum of those constants κ' such that

$$|w - z| \leq \kappa' \|W - A\| \;\; \text{for all } (W, w) \text{ near } (A, z) \text{ with } w \in \Lambda(W).$$

As is well-known, if the eigenvalue z is not semisimple, its classical condition number is infinite. In this case however, we can still quantify conditioning by switching to a logarithmic scale [5,11,13]. Let us define the *condition exponent* of z by

$$\beta = \liminf_{A \neq W \to A, \; w \to z, \; w \in \Lambda(W)} \frac{\log |w - z|}{\log \|W - A\|}.$$

Alternatively, β is the supremum of those constants β' such that

$$|w - z| \leq \|W - A\|^{\beta'} \;\; \text{for all } (W, w) \text{ near } (A, z) \text{ with } w \in \Lambda(W).$$

As essentially observed in [13] (an observation that follows from Theorem 2.4 below), fundamental work of Lidskii [12] implies the following characterization.

Theorem 2.1 (condition exponent) *The reciprocal of the condition exponent of an eigenvalue is its order as a root of the minimal polynomial.*

In particular, the condition exponent of a simple (or semisimple) eigenvalue is one.

Equipped with the idea of the condition exponent, we can then refine our measure of conditioning by defining the *condition coefficient*

$$\alpha = \limsup_{A \neq W \to A, \; w \to z, \; w \in \Lambda(W)} \frac{|w - z|^{1/\beta}}{\|W - A\|}.$$

Alternatively, α is the infimum of those constants α' such that

$$|w - z| \leq (\alpha' \|W - A\|)^{\beta} \;\; \text{for all } (W, w) \text{ near } (A, z) \text{ with } w \in \Lambda(W).$$

The pair (β, α) is very close to the notion of the *Hölder condition number* introduced in [13]. Indeed, we see below that it is precisely equivalent.

To explore these ideas further, let us fix the following notation.

Assumption 2.2 *The matrix* $A \in \mathbf{M}^n$ *has an eigenvalue* $z \in \mathbf{C}$, *and transforms to Jordan form via a matrix* $P \in \mathbf{M}^n$:

$$P^{-1}AP = \begin{bmatrix} J & & & & \\ & J & & & \\ & & \ddots & & \\ & & & J & \\ & & & & \bar{A} \end{bmatrix} \quad where \; J = \begin{bmatrix} z & 1 & 0 & \cdots & 0 \\ 0 & z & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & z \end{bmatrix} \in \mathbf{M}^m.$$

Furthermore, the Jordan block J *occurs* k *times, and each Jordan block in the matrix* $\bar{A}$ *is either smaller than* J *or corresponds to a different eigenvalue. The columns of* P *are* $p_1, p_2, \ldots, p_n$, *and the rows of* P^{-1} *are* $q_1, q_2, \ldots, q_n$, *and we define matrices*

$$\widetilde{Q} = \begin{bmatrix} q_m \\ q_{2m} \\ \vdots \\ q_{km} \end{bmatrix} \quad and \; \widetilde{P} = [p_1, p_{m+1}, \ldots, p_{(k-1)m+1}]$$

having rows and columns respectively the left and right eigenvectors corresponding to the k *m-by-m Jordan blocks for the eigenvalue* z, *and a quantity*

$$\widetilde{\alpha} = \|\widetilde{P}\widetilde{Q}\|.$$

Under this assumption, m is the order of the eigenvalue z as a root of the minimal polynomial for A. Hence, by Theorem 2.1, the condition exponent of z is $1/m$. Note that $\tilde{\alpha}$ is the norm of the projector onto the geometric eigenspace for z, not the norm of the projector onto the full invariant subspace. We will establish below that $\tilde{\alpha}$ equals α, the condition coefficient for the eigenvalue z. The basic tool for doing so is the following special case of the main result in [13].

Theorem 2.3 (eigenvalue perturbation) *Suppose Assumption 2.2 holds, and consider any sequences of numbers* $\epsilon_r \downarrow 0$ *and matrices* $A_r \in \mathbf{M}^n$ *satisfying*

$$A_r = A + \epsilon_r B + o(\epsilon_r) \quad as \; r \to \infty$$

for some matrix $B \in \mathbf{M}^n$. *Denote the eigenvalues of the matrix* $\widetilde{Q}B\widetilde{P} \in \mathbf{M}^k$ *by* $w_1, w_2, \ldots, w_k$ *(listed by algebraic multiplicity). Then* km *of the eigenvalues of* A_r *can be expressed in the form*

$$z + \gamma^s w_j^{1/m} \epsilon_r^{1/m} + o(\epsilon_r^{1/m}), \quad for \; s = 0, 1, \ldots, m-1, \quad and \; j = 1, 2, \ldots, k,$$

where $w_j^{1/m}$ *denotes any mth root of* w_j *and* γ *is a primitive mth root of unity, while the remaining* $n - km$ *eigenvalues are either within a distance* $o(\epsilon_r^{1/m})$ *of the eigenvalue* z *or remain uniformly bounded away from* z *as* $r \to \infty$.

As a consequence of this eigenvalue perturbation theorem, we derive a recent result concerning pseudospectral growth due to Karow [10]. Specifically, we show that small pseudospectral components are approximately disks centered at eigenvalues. We denote the closed unit disk in $\mathbf{C}$ by D. The distance from a point $s \in \mathbf{C}$ to a set $T \subset \mathbf{C}$ is

$$d_T(s) = \inf_{t \in T} |s - t|$$

and the *Hausdorff distance* between T and another set $S \subset \mathbf{C}$ is

$$h(S, T) = \max \left\{ \sup_{s \in S} d_T(s), \ \sup_{t \in T} d_S(t) \right\}.$$

Theorem 2.4 (pseudospectral growth) *Suppose Assumption 2.2 holds, so in particular, the eigenvalue z is a zero of multiplicity m for the minimal polynomial of the matrix A. Then the Hausdorff distance between the pseudospectral component $\Lambda_\epsilon^z(A)$ and the disk $z + (\widetilde{\alpha}\epsilon)^{1/m} D$ is $o(\epsilon^{1/m})$ as $\epsilon \downarrow 0$.*

Proof Without loss of generality we can suppose the eigenvalue z is zero. If the result fails to hold, there is a sequence of numbers $\epsilon_r \downarrow 0$ and a constant $\kappa > 0$ such that

$$h\left((\widetilde{\alpha}\epsilon_r)^{1/m} D, \Lambda_{\epsilon_r}^0(A)\right) > \kappa \epsilon_r^{1/m} \quad \text{for } r = 1, 2, \dots.$$

Consequently, we can assume the existence of a sequence of points $z_r \in \mathbf{C}$ such that either

$$z_r \in \Lambda_{\epsilon_r}^0(A) \quad \text{and} \quad |z_r| > (\widetilde{\alpha}^{1/m} + \kappa)\epsilon_r^{1/m} \tag{2.5}$$

or

$$|z_r| \le (\widetilde{\alpha}\epsilon_r)^{1/m} \quad \text{and} \quad d_{\Lambda_{\epsilon_r}^0(A)}(z_r) > \kappa \epsilon_r^{1/m}. \tag{2.6}$$

By taking a subsequence, we can suppose in fact that one of these conditions holds for all r.

Let us begin by supposing condition (2.5) holds for all r. Clearly $z_r \to 0$. Furthermore, there is a sequence of matrices $A_r \in \mathbf{M}^n$ satisfying $\|A_r - A\| \le \epsilon_r$ and $z_r \in \Lambda(A_r)$ for all r. Clearly $A_r \ne A$ for all large r, so we can suppose the sequence of numbers $\delta_r = \|A_r - A\|$ is never zero. Taking a subsequence, we can assume $\delta_r^{-1}(A_r - A)$ converges to a matrix B with norm one, so

$$A_r = A + \delta_r B + o(\delta_r) \quad \text{as } r \to \infty.$$

We can now apply Theorem 2.3 (eigenvalue perturbation). After taking a further subsequence, there is an eigenvalue w of the matrix $\widetilde{Q}B\widetilde{P}$ such that

$$z_r = \delta_r^{1/m} w^{1/m} + o(\delta_r^{1/m}),$$

where $w^{1/m}$ is an mth root of w. Since w is a nonzero eigenvalue of $\widetilde{Q}B\widetilde{P}$ (by the second condition in (2.5)), it is also an eigenvalue of the matrix $\widetilde{P}\widetilde{Q}B$, which has norm

no larger than $\widetilde{\alpha}$. Hence we obtain

$$|z_r| \leq (\epsilon_r \widetilde{\alpha})^{1/m} + o(\epsilon_r^{1/m}),$$

contradicting our assumption that condition (2.5) holds.

Suppose on the other hand that condition (2.6) holds for all r. Since $0 \in \Lambda_{\epsilon_r}^0(A)$, we know that each point z_r is nonzero, so after taking a subsequence we can suppose that $|z_r|^{-1} z_r$ approaches a point w on the unit circle.

Fix a pair of unit left and right singular vectors $u, v \in \mathbf{C}^n$ corresponding to the largest singular value $\widetilde{\alpha}$ of the matrix $\widetilde{P} \widetilde{Q}$. Since $u^* \widetilde{P} \widetilde{Q} v = \widetilde{\alpha}$, we deduce $\widetilde{\alpha} \in \Lambda(\widetilde{Q} v u^* \widetilde{P})$, and so $\widetilde{\alpha} w^m \in \Lambda(\widetilde{Q}(w^m v u^*) \widetilde{P})$. Theorem 2.3 (eigenvalue perturbation) now shows

$$(\widetilde{\alpha}\delta)^{1/m} w \in \Lambda(A + \delta w^m v u^*) + o(\delta^{1/m}) \subset \Lambda_\delta(A) + o(\delta^{1/m}) \tag{2.7}$$

as $\delta \downarrow 0$. In fact we can replace the pseudospectrum $\Lambda_\delta(A)$ by its component $\Lambda_\delta^0(A)$ on the right-hand side: otherwise there would exist a sequence $\delta_r \downarrow 0$ and a nonzero eigenvalue z' of the matrix A such that

$$(\widetilde{\alpha}\delta_r)^{1/m} w \in \Lambda_{\delta_r}^{z'}(A) + o(\delta_r^{1/m}) \quad \text{as } r \to \infty,$$

and this is impossible since the right-hand converges to the singleton $\{z'\}$ as $r \to \infty$.

In inclusion (2.7), let $\delta = \widetilde{\alpha}^{-1} |z_r|^m$. Since $\delta \leq \epsilon_r$ (by condition (2.6)), we deduce

$$|z_r| w \subset \Lambda_{\epsilon_r}^0(A) + o(z_r),$$

as $r \to \infty$, so

$$d_{\Lambda_{\epsilon_r}^0(A)}(z_r) \leq \left| |z_r| w + o(z_r) - z_r \right| = o(z_r) = o(\epsilon_r^{1/m}).$$

But this contradicts condition (2.6).   □

A more general version of Theorem 2.4 appears in an unpublished thesis of Karow [10, Theorems 2.6.6 and 5.4.4]. Whereas the argument presented above relies on a direct eigenvalue perturbation analysis using the pseudospectral representation (1.2), Karow's elegant proof instead relies on the representation (1.1). His argument depends on the following fact (see also the unpublished thesis of Ilahi [9], [5, Proposition 2.1], and [11, Remark 4.4]): whenever $z \in \mathbf{C}$ is a zero of multiplicity m for the minimal polynomial of the matrix A, we have

$$\underline{\sigma}(A - wI) = \frac{|w - z|^m}{\widetilde{\alpha}} + O(|w - z|^{m+1}) \tag{2.8}$$

for points $w \in \mathbf{C}$ near z (where $\widetilde{\alpha}$ is defined in Assumption 2.2). For comparison, we end this section with a short proof of Eq. (2.8): Theorem 2.4 (pseudospectral growth) follows immediately.

Without loss of generality, we can consider the eigenvalue $z = 0$ of a singular matrix $A \in \mathbf{M}^n$. Revising the notation of Assumption 2.2, we instead write

$$P^{-1}AP = \begin{bmatrix} N & 0 \\ 0 & \hat{A} \end{bmatrix},$$

where the matrices P and $\hat{A}$ are invertible and the nilpotent matrix N satisfies $N^m = 0$ but $N^{m-1} \neq 0$. Then for small $w \in \mathbf{C}$, we have

$$\frac{|w|^m}{\underline{\sigma}(A - wI)} = |w|^{m-1}\|(I - w^{-1}A)^{-1}\|$$

$$= |w|^{m-1}\left\|P\begin{bmatrix} (I - w^{-1}N)^{-1} & 0 \\ 0 & (I - w^{-1}\hat{A})^{-1} \end{bmatrix}P^{-1}\right\|$$

$$= \left\||w|^{m-1}P\begin{bmatrix} \sum_{j=0}^{m-1}(w^{-1}N)^j & 0 \\ 0 & (I - w^{-1}\hat{A})^{-1} \end{bmatrix}P^{-1}\right\|$$

$$= \left\|P\begin{bmatrix} \left(\frac{|w|}{w}N\right)^{m-1} + O(w) & 0 \\ 0 & O(w^m) \end{bmatrix}P^{-1}\right\|$$

$$= \left\|P\begin{bmatrix} N^{m-1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}\right\| + O(w) = \tilde{\alpha} + O(w).$$

Equation (2.8) now follows.

3 Nearest matrix with a multiple eigenvalue

A well-known problem studied by many authors concerns the distance from a given square matrix A to the set of matrices with multiple eigenvalues. The survey [8] summarizes early work on this question and its relationship to ill-conditioning. Several authors have approached the problem and generalizations by considering when components of the pseudospectrum $\Lambda_\epsilon(A)$ merge as ϵ grows. In particular, [18] considers the coalescence of components of the pseudospectrum (which is called there the "fundamental inclusion domain"), defined via Eq. (1.2) but using a general matrix norm. Merging of more general partitions of the pseudospectrum is studied in [6,7]. In related earlier work on "spectrum separation" [17], $\Lambda_\epsilon(A)$ is called the "ϵ-spectrum". More recently, Alam and Bora use the pseudospectral approach to prove the following result [1].

Theorem 3.1 (eigenvalue coalescence) *For any $n > 1$, the distance from a given n-by-n matrix A to the set of matrices with multiple eigenvalues is the minimum value of $\epsilon \geq 0$ such that the ϵ-pseudospectrum of A has less than n components.*

In this section we give a new proof of this result, following the philosophy of Alam and Bora in [1], but more topological in spirit, particularly in the derivation of the lower bound on the distance to the set of matrices with multiple eigenvalues. We also follow [1] in showing that the desired multiple eigenvalues are exactly the points lying in the closure of more than one component of the strict pseudospectrum for the critical value of ϵ.

For large ϵ, the pseudospectrum $\Lambda_\epsilon(A)$ is connected. For example, providing ϵ is larger than the finite number

$$\max\left\{\underline{\sigma}(A - zI) : z \in \operatorname{conv}\Lambda(A)\right\},$$

the connected set $\operatorname{conv}\Lambda(A)$ must be contained in $\Lambda_\epsilon(A)$. But the spectrum $\Lambda(A)$ intersects each component of $\Lambda_\epsilon(A)$, which must therefore be connected.

Denoting the number of components of any set $S \subset \mathbf{C}$ by $\operatorname{comp} S$, we can therefore define a number

$$\bar{\epsilon} = \inf\left\{\epsilon \geq 0 : \operatorname{comp}\Lambda_\epsilon(A) < n\right\}. \tag{3.2}$$

Lemma 3.3 *The infimum in Eq. (3.2) is attained.*

Proof If the result fails, then $\operatorname{comp}\Lambda_{\bar{\epsilon}}(A) = n$ and yet, for some sequence $\epsilon_r > \bar{\epsilon}$ decreasing to $\bar{\epsilon}$, we have $\Lambda_{\epsilon_r}(A) < n$. But this is a contradiction, because the compact sets $\Lambda_{\epsilon_r}(A)$ converge to the compact set $\Lambda_{\bar{\epsilon}}(A)$, and the function comp is lower semicontinuous with respect to convergence of compact sets [3, Proposition 4.1]. $\square$

If $\bar{\epsilon} = 0$, we have nothing more to prove, so we suppose $\bar{\epsilon} > 0$. We next consider the strict pseudospectrum

$$\Omega = \{z \in \mathbf{C} : \underline{\sigma}(A - zI) < \bar{\epsilon}\},$$

which is an open set with closure the pseudospectrum $\Lambda_{\bar{\epsilon}}(A)$ [2, Corollary 4.3].

Lemma 3.4 *The strict pseudospectrum Ω has n components.*

Proof Note that, just as for pseudospectra, each component of Ω contains an eigenvalue, so if the claim fails, two eigenvalues $z_0 \neq z_1$ of the matrix A lie in the same component of Ω. Hence there is a continuous path $t \mapsto z(t) \in \Omega$ (for $t \in [0, 1]$) satisfying $z(0) = z_0$ and $z(1) = z_1$. Consequently, the number

$$\epsilon' = \max_{t \in [0,1]} \underline{\sigma}(A - z(t)I)$$

is strictly less than $\bar{\epsilon}$. Furthermore, this path shows that z_0 and z_1 lie in the same component of the pseudospectrum $\Lambda_{\epsilon'}(A)$, so $\operatorname{comp}\Lambda_{\epsilon'}(A) < n$, contradicting the definition of $\bar{\epsilon}$. $\square$

The strict pseudospectrum Ω is therefore the disjoint union of nonempty connected open sets $\Omega_1, \Omega_2, \ldots, \Omega_n$.

Lemma 3.5 *For any matrix E with $\|E\| < \bar{\epsilon}$, the matrix $A + E$ has exactly one eigenvalue in each of the n components of the strict pseudospectrum Ω. Furthermore, each of these eigenvalues is simple.*

Proof Intuitively, the outline of the proof is straightforward. If we let $\epsilon = \|E\|$, then each of the n components of Ω must contain exactly one component of the pseudospectrum $\Lambda_\epsilon(A)$, and by using the continuity of the spectrum, each of these latter components must contain exactly one eigenvalue of $A + E$. Formalizing this argument takes a little care.

Denote the components of $\Lambda_\epsilon(A)$ by $C_1, C_2, \ldots, C_n$, each of which is compact because $\Lambda_\epsilon(A)$ is compact. Since $\Lambda_\epsilon(A) \subset \Omega$ and each C_j contains an eigenvalue of A, we can suppose $C_j \subset \Omega_j$ for each j. Now consider the number

$$\bar{t} = \sup \left\{ t' \in [0, 1] : \Lambda(A + tE) \cap C_1 \neq \emptyset \; \forall t \in [0, t'] \right\}.$$

We claim the supremum above is attained. If $\bar{t} = 0$, there is nothing to prove, so suppose $\bar{t} > 0$: we just need to show $\Lambda(A + \bar{t}E) \cap C_1 \neq \emptyset$. Choose any sequence $t_1, t_2, \ldots \in (0, \bar{t})$ approaching $\bar{t}$ and corresponding points $z_r \in \Lambda(A + t_r E) \cap C_1$. Since C_1 is compact, after taking a subsequence we can suppose z_r approaches some point $\bar{z} \in C_1$. But the set-valued map Λ taking a matrix to its spectrum is continuous, so $\bar{z} \in \Lambda(A + \bar{t}E)$, as required.

Now suppose $\bar{t} < 1$. We can then choose another sequence $t_1, t_2, \ldots \in (\bar{t}, 1)$ approaching $\bar{t}$ with $\Lambda(A + t_r E) \cap C_1 = \emptyset$. Since $\|t_r E\| \leq \epsilon$ for all r, we have $\Lambda(A + t_r E) \subset \Lambda_\epsilon(A)$, so in fact $\Lambda(A + t_r E) \subset \cup_{j>1} C_j$. Since the right hand side is closed, it also contains $\Lambda(A + \bar{t}E)$, by continuity of Λ, giving the contradiction $\Lambda(A + \bar{t}E) \cap C_1 = \emptyset$. Hence we must have $\bar{t} = 1$.

We have thus shown that the component C_1 (and hence Ω_1) contains an eigenvalue of the matrix $A + E$. But the same holds for each component, so the result follows. $\square$

Notice that Lemma 3.5 shows that the number $\bar{\epsilon}$ is a lower bound on the distance to the set of matrices with multiple eigenvalues. The last step in our proof relies on the following well-known tool for recognizing multiple eigenvalues (which follows quickly from the Jordan form, for example).

Proposition 3.6 (orthogonal eigenvectors) *Consider a matrix $B \in \mathbf{M}^n$ with an eigenvalue z corresponding to left and right eigenvectors $u, v \in \mathbf{C}^n$: in other words, u and v are nonzero and satisfy $Bv = zv$ and $u^*B = zu^*$. If $u^*v = 0$, then the eigenvalue z is multiple.*

Proceeding with our proof of Theorem 3.1, observe that each component Ω_j has connected closure, and the union of these closures is the set $\mathrm{cl}\,\Omega = \Lambda_{\bar{\epsilon}}(A)$, which has less than n components. Hence there must exist at least one point $\bar{z}$ lying in the closure of two distinct components of Ω. We conclude our proof by showing that any such $\bar{z}$ is our required multiple eigenvalue.

Lemma 3.7 *The following properties of a point $\bar{z} \in \mathbf{C}$ are equivalent:*

(i) *$\bar{z}$ lies in the closure of two distinct components of Ω;*

(ii) *there exists a matrix E such that $\|E\| = \bar{\epsilon}$ and $\bar{z}$ is a multiple eigenvalue of $A + E$.*

Proof Suppose property (i) holds. Since the components of Ω are disjoint, we immediately deduce $\underline{\sigma}(A - \bar{z}I) = \bar{\epsilon}$. If this singular value is multiple, then a standard application of the singular value decomposition gives a matrix E such that $\|E\| = \bar{\epsilon}$ and $A - \bar{z}I + E$ has rank no more than $n - 2$. In this case, $\bar{z}$ must be a multiple eigenvalue of $A + E$, as required.

We can therefore assume the singular value $\underline{\sigma}(A - \bar{z}I)$ is simple. Identify the space $\mathbf{C}$ as a Euclidean space, with inner product $\langle w, z \rangle = \mathrm{Re}\,(w^*z)$. In this setting, the function $g : \mathbf{C} \to \mathbf{R}$ defined by $g(z) = \underline{\sigma}(A - zI)$ is continuously differentiable around $\bar{z}$, with gradient $\nabla g(\bar{z}) = -u^*v$, for any corresponding normalized left and right singular vectors u and v [15]. If this gradient is nonzero, then the set $\Omega = \{z \in \mathbf{C} : g(z) < \epsilon\}$ is bounded near $\bar{z}$ by a smooth curve through $\bar{z}$, using the implicit function theorem, so is locally connected. But this contradicts the fact that $\bar{z}$ lies in the closure of two distinct components of Ω.

The only remaining possibility is $u^*v = 0$. But a quick calculation shows that if we set $E = -\bar{\epsilon}uv^*$, then $\|E\| = \bar{\epsilon}$, and u and v are respectively left and right eigenvectors for the matrix $A + E$, corresponding to the eigenvalue $\bar{z}$. Property (ii) then follows from Proposition 3.6.

Conversely, suppose property (ii) holds. If property (i) fails, then $\bar{z}$ lies in the closure of a unique component of Ω, say Ω_1. Hence some open neighborhood U of $\bar{z}$ is disjoint from each component Ω_j for $j > 1$. Consider any sequence $t_r \in (0, 1)$ converging to 1. For each r, the matrix $A + t_r E$ has a simple eigenvalue, say z_r^j, in each component Ω_j, by Lemma 3.5. After taking a subsequence, we can suppose $z_r^j \to z^j \in \mathrm{cl}\,\Omega_j$ as $r \to \infty$, and the numbers z^j are just the eigenvalues of $A + E$, listed by algebraic multiplicity, since

$$\det(zI - (A + E)) = \lim_{r \to \infty} \det(zI - (A + t_r E)) = \lim_{r \to \infty} \prod_{j=1}^{n} (z - z_r^j) = \prod_{j=1}^{n} (z - z^j).$$

But by assumption, if $j > 1$, then $z_r^j \notin U$ for all large r, and hence $z^j \notin U$. In particular, $z^j \neq \bar{z}$ for all $j > 1$, so $\bar{z}$ cannot be a multiple eigenvalue of $A + E$. This is a contradiction. $\qquad\square$

This concludes the proof of Theorem 3.1, and the claim following it.

In Lemma 3.7, the multiple eigenvalue $\bar{z}$ may vary in character, depending on the case. Consider, for example, the two matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}.$$

In the first case, the nearest matrix with a double eigenvalue is the zero matrix, whereas in the second, the nearest such matrix is similar to a two-by-two Jordan block [4]

The techniques of Lemmas 3.3, 3.4 and 3.5 extend easily to an arbitrary matrix A whose distinct eigenvalues $z_1, z_2, \ldots, z_k$ have algebraic multiplicities $m_1, m_2, \ldots, m_k$ respectively. In the definition of the number $\bar{\epsilon}$, Eq. (3.2), if we replace the number of components n by k, Lemmas 3.3 and 3.4 (with n replaced by k) hold with unchanged proofs. A result analogous to Lemma 3.5 also holds: specifically, for any matrix E satisfying $\|E\| < \bar{\epsilon}$, the matrix $A + E$ has exactly m_j eigenvalues (counted by multiplicity) in the component Ω_j of the strict pseudospectrum Ω containing z_j. Again, the proof is largely unchanged. We simply replace the condition $\Lambda(A + tE) \cap C_1 \neq \emptyset$ in the definition of the number $\bar{t}$ by the condition that the matrix $A + tE$ has m_1 eigenvalues (counted by multiplicity) in the set C_1, and then again use the continuity of the mapping from a matrix to the set of all vectors obtained by listing the eigenvalues by multiplicity in all possible orders. Maintaining this setting and turning to Lemma 3.7, the same proof still shows that property (i) implies property (ii). On the other hand, if property (ii) holds, then certainly the eigenvalue $\bar{z}$ must lie in the closure of at least one component of Ω, say Ω_1, and if the multiplicity m_1 is one, then the same argument as for Lemma 3.7 shows that $\bar{z}$ must also lie in the closure of another component.

References

1. Alam, R., Bora, S.: On sensitivity of eigenvalues and eigendecompositions of matrices. Linear Algebra Appl. **396**, 273–301 (2005)
2. Burke, J.V., Lewis, A.S., Overton, M.L.: Optimization and pseudospectra, with applications to robust stability. SIAM J. Matrix Anal. Appl. **25**, 80–104 (2003) Corrigendum: www.cs.nyu.edu/cs/faculty/overton/papers/pseudo_corrigendum.html
3. Burke, J.V., Lewis, A.S., Overton, M.L.: Pseudospectral components and the distance to uncontrollability. SIAM J. Matrix Anal. Appl. **26**, 350–361 (2004)
4. Burke, J.V., Lewis, A.S., Overton M.L.: Convexity and Lipschitz behavior of small pseudospectra. SIAM J. Matrix Anal. Appl. (2007, in press)
5. Chaitin-Chatelin, F., Harrabi, A., Ilahi, A.: About Hölder condition numbers and the stratification diagram for defective eigenvalues. Math. Comput. Simulation **54**, 397–402 (2000)
6. Demmel, J.W.: A Numerical Analyst's Jordan Canonical Form. PhD thesis, University of California, Berkeley (1983)
7. Demmel, J.W.: Computing stable eigendecompositions of matrices. Linear Algebra Appl. **79**, 163–193 (1986)
8. Demmel, J.W.: Nearest defective matrices and the geometry of ill-conditioning. In: Cox, M.G., Hammerling, S. (eds.) Reliable Numerical Computation, pp. 35–55. Oxford University Press, Oxford (1990)
9. Ilahi, A.: Validation du Calcul sur Ordinateur: Application de la Théorie des Singularités. PhD thesis, University of Toulouse I (1998)
10. Karow, M.: Geometry of Spectral Value Sets. PhD thesis, University of Bremen (2003)
11. Karow, M.: Eigenvalue condition numbers and a formula of Burke, Lewis and Overton. Electron. J. Linear Algebra **15**, 143–153 (2006)
12. Lidskii, V.B.: Perturbation theory of non-conjugate operators. USSR Comput. Mathematics Math. Phys. **6**(1), 73–85 (1966)
13. Moro, J., Burke, J.V., Overton, M.L.: On the Lidskii-Vishik-Lyusternik perturbation theory for the eigenvalues of matrices with arbitrary Jordan structure. SIAM J. Matrix Anal. Appl. **18**, 793–817 (1997)
14. Polyak, B.T.: Convexity of nonlinear image of a small ball with applications to optimization. Set-Valued Anal. **9**, 159–168 (2001)
15. Sun, J.G.: A note on simple non-zero singular values. J. Comput. Math. **6**, 258–266 (1988)
16. Trefethen, L.N., Embree, M.: Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators. Princeton University Press, Princeton (2005)
17. Varah, J.M.: On the separation of two matrices. SIAM J. Numer. Anal. **16**, 216–222 (1979)
18. Wilkinson, J.H.: Sensitivity of eigenvalues II. Utilitas Mathematica **30**, 243–286 (1986)