

16.7 Eigenvalues Move Continuously

Chapter 12 §01 proved that every complex matrix has diagonalizable matrices arbitrarily close to it (Corollary 12.1.6), then admitted that the word “close” was doing work no definition supported, and handed the repair to this chapter. The repair is now in place: a norm on $M_n(\mathbb{C})$ says how close two matrices are, and Corollary 16.2.4 says that it does not matter which norm. This section asks the question that could not even be posed before that point. If the matrix moves a little, how far do its eigenvalues move?

The answer has two halves, and the second is the one worth remembering. The eigenvalues do move only a little — provided they are read as an unordered list, which is the only way they can be read. But “a little” is not proportional to the movement of the matrix, and for a matrix with a large Jordan block it is spectacularly larger.

Throughout, matrices are complex, as in Section 4, so that an $n \times n$ matrix always has exactly n eigenvalues counted with algebraic multiplicity (Corollary 6.5.5).

16.7.1 From matrices to polynomials

The eigenvalues of $\mathbf{A}$ are the roots of $p_{\mathbf{A}}$ (Theorem 9.2.1), with algebraic multiplicity meaning multiplicity as a root (Definition 9.3.1). So the question splits in two: do the coefficients of $p_{\mathbf{A}}$ move a little when $\mathbf{A}$ does, and do the roots of a polynomial move a little when its coefficients do? The first half is easy and we do it now; the whole difficulty is in the second.

The reason the first half is easy is that everything in sight is a polynomial in the entries. It is worth isolating that observation once, because it will be used four times in this section.

Lemma 16.7.1. Polynomial Maps of the Entries Are Continuous

Let $\Phi: M_{p \times q}(\mathbb{C}) \rightarrow M_{r \times s}(\mathbb{C})$ be a map each of whose output entries is given by a fixed polynomial, with complex coefficients, in the pq entries of the input. Fix any norm on the source and any norm on the target. Then Φ is **continuous**: whenever $\|\mathbf{A}_k - \mathbf{A}\| \rightarrow 0$ we have $\|\Phi(\mathbf{A}_k) - \Phi(\mathbf{A})\| \rightarrow 0$, and equivalently, for every $\mathbf{A}$ and every $\varepsilon > 0$ there is $\delta > 0$ such that $\|\mathbf{B} - \mathbf{A}\| < \delta$ forces $\|\Phi(\mathbf{B}) - \Phi(\mathbf{A})\| < \varepsilon$. The same holds for a map into $\mathbb{C}$.

Idea. Two reductions and one piece of arithmetic. Convergence in any norm is entrywise convergence, by Corollary 16.2.4, in the source and in the target alike; so the claim is about pq complex sequences producing rs complex sequences. That reduces it to the statement that a polynomial in finitely many complex variables respects limits, which is the algebra of limits.

Proof. Suppose $\|\mathbf{A}_k - \mathbf{A}\| \rightarrow 0$. By Corollary 16.2.4, each entry $(\mathbf{A}_k)_{ij} \rightarrow a_{ij}$. Sums and products of convergent complex sequences converge to the sum and the product of the limits: for sums this is Theorem 0.6.4, and for products it follows from

$$|z_k w_k - zw| \leq |z_k| |w_k - w| + |w| |z_k - z|$$

together with the fact that a convergent sequence is bounded. A polynomial is built from its variables and constants by finitely many sums and products, so each entry of $\Phi(\mathbf{A}_k)$ converges to the corresponding entry of $\Phi(\mathbf{A})$. Applying Corollary 16.2.4 in $M_{r \times s}(\mathbb{C})$ gives

$$\|\Phi(\mathbf{A}_k) - \Phi(\mathbf{A})\| \rightarrow 0.$$

For the equivalence of the two formulations, suppose the ε - δ statement failed at some $\mathbf{A}$ for some $\varepsilon > 0$. Then for each $k \geq 1$ the value $\delta = 1/k$ is not good enough, so there is $\mathbf{B}_k$ with $\|\mathbf{B}_k - \mathbf{A}\| < 1/k$ and $\|\Phi(\mathbf{B}_k) - \Phi(\mathbf{A})\| \geq \varepsilon$. The sequence $(\mathbf{B}_k)$ contradicts the sequential statement just proved. This proves the lemma. ■

Now the first half of the question is answered. Write $p_{\mathbf{A}}(x) = x^n + c_{n-1}(\mathbf{A})x^{n-1} + \cdots + c_0(\mathbf{A})$, which is legitimate because $p_{\mathbf{A}}$ is monic of degree n (Theorem 7.10.2). Each c_j is a polynomial in the entries of $\mathbf{A}$: the Leibniz sum of Definition 7.10.1 expresses $\det(x\mathbf{I} - \mathbf{A})$ as a signed sum of products of the entries of $x\mathbf{I} - \mathbf{A}$, each of which is $x - a_{jj}$ or $-a_{ij}$, and collecting the powers of x leaves every coefficient a polynomial with integer coefficients in the a_{ij} . So $\mathbf{A} \mapsto (c_0(\mathbf{A}), \dots, c_{n-1}(\mathbf{A}))$ is continuous by Lemma 16.7.1.

That leaves the second half: roots against coefficients. Before stating what is true there, we settle what cannot be true.

16.7.2 Why the answer must be about the multiset

The statement a reader would write down first is “each eigenvalue is a continuous function of the matrix”. It is not merely unproved; it is meaningless as written, because the eigenvalues come with no numbering — and the natural repair, “there is a way of numbering them that makes each one continuous”, is false. One two-by-two family shows it, and the failure is the ordinary square root.

Proposition 16.7.2. There Is No Continuous Choice of a Single Eigenvalue

For $t \in \mathbb{C}$ put

$$\mathbf{A}(t) = \begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix},$$

so that $p_{\mathbf{A}(t)}(x) = x^2 - t$ and the eigenvalues of $\mathbf{A}(t)$ are the two square roots of t . Then there is **no** continuous function $\lambda: D \rightarrow \mathbb{C}$ on the closed unit disc $D = \{t \in \mathbb{C} : |t| \leq 1\}$ such that $\lambda(t)$ is an eigenvalue of $\mathbf{A}(t)$ for every $t \in D$. Consequently there is no continuous $\Lambda: M_2(\mathbb{C}) \rightarrow \mathbb{C}$ with $\Lambda(\mathbf{B}) \in \text{spec}(\mathbf{B})$ for every $\mathbf{B}$.

Idea. Walk t once around the unit circle. The two roots $\pm\sqrt{t}$ trade places on the way round, so a choice that is continuous along the walk comes home as the *other* root. The bookkeeping is done by dividing the chosen root by one fixed root, which turns the walk into a continuous function with only the two values 1 and -1 ; on an interval such a function cannot jump, and it must.

Proof. The characteristic polynomial is $\det\begin{pmatrix} x & -1 \\ -t & x \end{pmatrix} = x^2 - t$, so μ is an eigenvalue of $\mathbf{A}(t)$ if and only if $\mu^2 = t$ (Theorem 9.2.1).

Suppose such a λ exists, so $\lambda(t)^2 = t$ for all $|t| \leq 1$. Define $g: [0, 2\pi] \rightarrow \mathbb{C}$ by

$$g(\theta) = \lambda(e^{i\theta}) e^{-i\theta/2}.$$

It is continuous, being a composite and a product of continuous functions — the elementary algebra of limits recorded in the chapter introduction, applied to $\theta \mapsto e^{i\theta}$, to λ , and to $\theta \mapsto e^{-i\theta/2}$. Squaring,

$$g(\theta)^2 = \lambda(e^{i\theta})^2 e^{-i\theta} = e^{i\theta} e^{-i\theta} = 1,$$

so $g(\theta) \in \{1, -1\}$ for every θ . In particular g is real-valued, and a continuous real-valued function on an interval taking only the values 1 and -1 is constant: if it took both, the intermediate value theorem would produce a θ with $g(\theta) = 0$. Hence $g(2\pi) = g(0)$.

But $e^{i \cdot 0} = e^{i \cdot 2\pi} = 1$, so $g(0) = \lambda(1)$ while $g(2\pi) = \lambda(1)e^{-i\pi} = -\lambda(1)$. Therefore $\lambda(1) = -\lambda(1)$, that is $\lambda(1) = 0$, contradicting $\lambda(1)^2 = 1$. So no such λ exists.

For the last sentence, if Λ were such a function then $t \mapsto \Lambda(\mathbf{A}(t))$ would be continuous, because $t \mapsto \mathbf{A}(t)$ is continuous by [Lemma 16.7.1](#), and it would select an eigenvalue of $\mathbf{A}(t)$ for each t . This proves the proposition. ■

The intermediate value theorem is quoted here, and it is the only place in this chapter that uses it; it is fact (A5) of the chapter introduction, borrowed and not proved.

Two things are worth reading off the argument. First, the obstruction is not exotic: $\mathbf{A}(0) = \mathbf{J}_2(0)$ is the standard nilpotent matrix, and the trouble happens exactly at the parameter where the two eigenvalues collide. Second, nothing is wrong with the pair $\{\sqrt{t}, -\sqrt{t}\}$, which shrinks to $\{0, 0\}$ as $t \rightarrow 0$ in any reasonable sense. It is only the attempt to call one of them “the first” that fails. This is why the correct statement speaks of the roots as a **multiset**: a list with multiplicities and no order.

16.7.3 The imported fact

Here is what is true. It is the one statement in this chapter that the chapter does not prove.

Theorem 16.7.3. Continuous Dependence of the Roots on the Coefficients

Let $n \geq 1$ and let

$$p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0 \in \mathbb{C}[x]$$

be monic, with roots $\mu_1, \dots, \mu_n \in \mathbb{C}$ listed with multiplicity (Corollary 6.5.5). Let $\varepsilon > 0$. Then there is $\delta > 0$ with the following property. If

$$q(x) = x^n + b_{n-1}x^{n-1} + \cdots + b_1x + b_0$$

is monic of the **same degree** n and $|b_j - c_j| < \delta$ for every j , then the roots $\nu_1, \dots, \nu_n$ of q , listed with multiplicity, can be **numbered** so that

$$|\nu_i - \mu_i| < \varepsilon \quad (i = 1, \dots, n).$$

This chapter quotes this theorem, and Chapter 20 proves it. The classical proof uses complex analysis: one encircles each root of p by a small circle on which p does not vanish, counts the roots inside by a contour integral, and observes that an integer-valued quantity depending

continuously on the coefficients cannot change. The book never develops that machinery, and it does not need to: compactness, fact (A3), together with unique factorization in $\mathbb{C}[x]$, is enough, and Chapter 20 §03 gives that proof. It is postponed there because it belongs with the quantitative bounds on how far roots move. Until then the statement is recorded in the manner this book has used before — Chapter 14 §06 with Taylor’s theorem, Chapter 11 with the theorem of Abel and Ruffini: named at the point of use, never buried inside a proof that presents itself as self-contained, and never strengthened in passing. What *can* be proved with the tools at hand is that the numbering in the statement has to be chosen afresh for each q , and that was [Proposition 16.7.2](#).

Two clauses of the statement carry weight. **Monic of the same degree** is essential: the roots of $\delta x^2 + x$ are 0 and $-1/\delta$, which runs off to infinity as $\delta \rightarrow 0$, while the limit polynomial x has the single root 0. Allowing the leading coefficient to vanish destroys everything, and monic polynomials of a fixed degree are exactly the family where it cannot. And **numbered so that** is the multiset formulation: the permutation matching up the two lists depends on q , and [Proposition 16.7.2](#) says it must.

16.7.4 Eigenvalues, and the spectral radius

Now the two halves are assembled.

Corollary 16.7.4. Eigenvalues Depend Continuously on the Matrix

Let $\mathbf{A} \in M_n(\mathbb{C})$ with eigenvalues $\lambda_1, \dots, \lambda_n$ listed with algebraic multiplicity, let $\|\cdot\|$ be any norm on $M_n(\mathbb{C})$, and let $\varepsilon > 0$. Then there is $\delta > 0$ such that every $\mathbf{B} \in M_n(\mathbb{C})$ with $\|\mathbf{B} - \mathbf{A}\| < \delta$ has eigenvalues $\mu_1, \dots, \mu_n$, listed with algebraic multiplicity, which can be numbered so that $|\mu_i - \lambda_i| < \varepsilon$ for every i .

Idea. The characteristic polynomial’s coefficients are polynomials in the entries, so they move continuously with $\mathbf{A}$; the quoted root theorem then moves the roots. The only work is composing the two estimates, choosing the tolerance for the coefficients small enough that the roots come out within ε .

Proof. By Theorem 7.10.2 the polynomials $p_{\mathbf{A}}$ and $p_{\mathbf{B}}$ are monic of degree n , and by Theorem 9.2.1 together with Definition 9.3.1 their root lists with multiplicity are the eigenvalue lists with algebraic multiplicity. Let $\delta' > 0$ be the number supplied by [Theorem 16.7.3](#) for $p = p_{\mathbf{A}}$ and this ε . The map $\mathbf{B} \mapsto (c_0(\mathbf{B}), \dots, c_{n-1}(\mathbf{B}))$ is continuous by [Lemma 16.7.1](#), as noted after that lemma, so there is $\delta > 0$ such that $\|\mathbf{B} - \mathbf{A}\| < \delta$ forces $|c_j(\mathbf{B}) - c_j(\mathbf{A})| < \delta'$ for every j . For such a $\mathbf{B}$, [Theorem 16.7.3](#) applies to $q = p_{\mathbf{B}}$. This proves the corollary. ■

The corollary says exactly as much as [Theorem 16.7.3](#) and no more: the numbering is not canonical and depends on $\mathbf{B}$. Any quantity that ignores the numbering, however, inherits honest continuity, and the spectral radius of Section 4 is the first example.

Quick check. Does [Corollary 16.7.4](#) make $\rho: M_n(\mathbb{C}) \rightarrow \mathbb{R}$ continuous, given that no individual eigenvalue is a continuous function of $\mathbf{A}$?

Solution. Yes. Let $\varepsilon > 0$ and take δ from [Corollary 16.7.4](#). If $\|\mathbf{B} - \mathbf{A}\| < \delta$, fix a numbering with $|\mu_i - \lambda_i| < \varepsilon$. Then $|\mu_i| \leq |\lambda_i| + \varepsilon \leq \rho(\mathbf{A}) + \varepsilon$ for each i by Definition 16.4.1, so $\rho(\mathbf{B}) \leq \rho(\mathbf{A}) + \varepsilon$; the same numbering read the other way gives $|\lambda_i| \leq |\mu_i| + \varepsilon \leq \rho(\mathbf{B}) + \varepsilon$ for each i , hence $\rho(\mathbf{A}) \leq \rho(\mathbf{B}) + \varepsilon$, so $|\rho(\mathbf{B}) - \rho(\mathbf{A})| \leq \varepsilon$. The argument never names an individual eigenvalue, which is why it survives: ρ is a function of the multiset alone.

16.7.5 The limit argument, made routine

Chapter 12 §01 called Corollary 12.1.6 “the tool behind a standard strategy: *prove it for diagonalizable matrices, then take a limit*”, and said that this chapter, “where matrices acquire norms and limits can be spoken of properly, runs arguments of exactly this shape”. Here is the shape. It costs two corollaries, and afterwards the strategy is three lines every time.

Corollary 16.7.5. Diagonalizable Matrices Are Dense

Let $\|\cdot\|$ be any norm on $M_n(\mathbb{C})$. For every $\mathbf{A} \in M_n(\mathbb{C})$ and every $\varepsilon > 0$ there is a **diagonalizable** $\mathbf{A}' \in M_n(\mathbb{C})$ with $\|\mathbf{A} - \mathbf{A}'\| < \varepsilon$. Equivalently, every complex square matrix is the limit of a sequence of diagonalizable matrices, in every norm.

Idea. Chapter 12 produced a diagonalizable matrix near $\mathbf{A}$ with the *entries* close. “Close in every norm” is now the same thing, by Corollary 16.2.4, so the only task is to convert one tolerance into the other.

Proof. Let $N_\infty(\mathbf{M}) = \max_{i,j} |m_{ij}|$, which is a norm on $M_n(\mathbb{C})$: it is the ∞ -norm of Example (The 1-, 2-, and ∞ -norms on F^n) read on the list of n^2 entries. By Theorem 16.2.2 there is $C > 0$ with $\|\mathbf{M}\| \leq C N_\infty(\mathbf{M})$ for every $\mathbf{M}$. Apply Corollary 12.1.6 with the tolerance ε/C : it supplies a diagonalizable $\mathbf{A}'$ with $|a_{ij} - a'_{ij}| < \varepsilon/C$ for all i, j , hence with $N_\infty(\mathbf{A} - \mathbf{A}') < \varepsilon/C$, a maximum of finitely many numbers each below that bound. Therefore $\|\mathbf{A} - \mathbf{A}'\| \leq C N_\infty(\mathbf{A} - \mathbf{A}') < \varepsilon$. Taking $\varepsilon = 1/k$ gives a sequence $\mathbf{A}_k$ of diagonalizable matrices with $\|\mathbf{A} - \mathbf{A}_k\| \rightarrow 0$. This proves the corollary. ■

Corollary 16.7.6. Proving an Identity by Density

Let $\Phi: M_n(\mathbb{C}) \rightarrow M_{r \times s}(\mathbb{C})$ be a map each of whose output entries is a polynomial in the entries of the input. If $\Phi(\mathbf{A}) = \mathbf{O}$ for every **diagonalizable** $\mathbf{A} \in M_n(\mathbb{C})$, then $\Phi(\mathbf{A}) = \mathbf{O}$ for every $\mathbf{A} \in M_n(\mathbb{C})$.

Idea. Both sides of the identity are continuous in $\mathbf{A}$, and they agree on the diagonalizable matrices, which are dense. A continuous function is determined by its values on a dense set, so they agree everywhere.

Proof. Fix $\mathbf{A}$ and a norm on each side. By [Corollary 16.7.5](#) choose diagonalizable $\mathbf{A}_k$ with $\|\mathbf{A} - \mathbf{A}_k\| \rightarrow 0$. By [Lemma 16.7.1](#), $\|\Phi(\mathbf{A}_k) - \Phi(\mathbf{A})\| \rightarrow 0$. Since $\Phi(\mathbf{A}_k) = \mathbf{O}$ by hypothesis, that says $\|\Phi(\mathbf{A})\| \rightarrow 0$; but $\|\Phi(\mathbf{A})\|$ does not depend on k , so $\|\Phi(\mathbf{A})\| = 0$ and $\Phi(\mathbf{A}) = \mathbf{O}$ by the positivity clause of Definition 16.1.1. This proves the corollary. ■

The hypothesis “each output entry is a polynomial in the input entries” is the one to check, and it is usually free: any expression built from $\mathbf{A}$ by addition, multiplication, transposition, conjugate transposition of a real matrix, determinants, adjugates, traces, and the coefficients of $p_{\mathbf{A}}$ qualifies. What does *not* qualify is anything built from an inverse, an eigenvector, or a choice of ordering. Here is the strategy in action, on the theorem that Chapter 9 proved the long way.

Example. Cayley–Hamilton, by a limit

Prove that $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$ for every $\mathbf{A} \in M_n(\mathbb{C})$, using density rather than the adjugate identity of Theorem 7.6.2.

Solution. Put $\Phi(\mathbf{A}) = p_{\mathbf{A}}(\mathbf{A})$. Each entry of $\Phi(\mathbf{A})$ is a polynomial in the entries of $\mathbf{A}$: the coefficients $c_j(\mathbf{A})$ are, as noted after [Lemma 16.7.1](#), and $\Phi(\mathbf{A}) = \mathbf{A}^n + c_{n-1}(\mathbf{A})\mathbf{A}^{n-1} + \cdots + c_0(\mathbf{A})\mathbf{I}$ is then built from them by products and sums.

Let $\mathbf{A}$ be diagonalizable, say $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ with $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$. By Theorem 9.4.3, $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{P} \text{diag}(p_{\mathbf{A}}(d_1), \dots, p_{\mathbf{A}}(d_n)) \mathbf{P}^{-1}$. Each d_i is an eigenvalue of $\mathbf{A}$, hence a root of $p_{\mathbf{A}}$ (Theorem 9.2.1), so every diagonal entry is 0 and $\Phi(\mathbf{A}) = \mathbf{O}$.

By [Corollary 16.7.6](#), $\Phi(\mathbf{A}) = \mathbf{O}$ for every $\mathbf{A} \in M_n(\mathbb{C})$, which is Theorem 9.6.1 over $\mathbb{C}$.

That is the whole strategy: check the identity where the matrix is diagonal in some basis, where everything is a scalar computation, and let density carry it to the rest. It proves nothing new here — Theorem 9.6.1 holds over every field, and the density argument is confined to $\mathbb{C}$ — but the three lines at the end are the part to transfer. [Exercise \(C2: Density of the invertible matrices, and \$\mathbf{AB}\$ versus \$\mathbf{BA}\$ \)](#) runs the same argument with a different dense set.

16.7.6 The warning: continuous is not Lipschitz

Everything so far has been qualitative. The quantitative question is whether there is a constant L with

$$(\text{eigenvalue movement}) \leq L \|\mathbf{B} - \mathbf{A}\|,$$

which is what “continuous” is usually imagined to mean and what a numerical analyst actually needs. There is not, and the failure is severe.

Example. A Jordan block’s eigenvalues move like $\varepsilon^{1/k}$

Let $k \geq 2$, let $\varepsilon > 0$, and perturb the nilpotent Jordan block in its bottom left corner:

$$\mathbf{A} = \mathbf{J}_k(0), \quad \mathbf{A}_{\varepsilon} = \mathbf{J}_k(0) + \varepsilon \mathbf{E}_{k1}.$$

Find the eigenvalues of $\mathbf{A}_\varepsilon$ and compare their distance from $\text{spec}(\mathbf{A}) = \{0\}$ with $\|\mathbf{A}_\varepsilon - \mathbf{A}\|$.

Solution. The characteristic polynomial is $x^k - \varepsilon$. Write $\mathbf{M} = x\mathbf{I} - \mathbf{A}_\varepsilon$, so $\mathbf{M}$ has x on the diagonal, -1 on the superdiagonal, $-\varepsilon$ in position $(k, 1)$, and zeros elsewhere. Expand along the first column (Theorem 7.6.1 (a)), whose only nonzero entries are $m_{11} = x$ and $m_{k1} = -\varepsilon$.

Deleting row 1 and column 1 leaves the $(k-1) \times (k-1)$ upper triangular matrix with x on the diagonal, of determinant x^{k-1} (Theorem 7.4.6). Deleting row k and column 1 leaves the matrix with rows $1, \dots, k-1$ and columns $2, \dots, k$ of $\mathbf{M}$; its (i, j) entry is $m_{i,j+1}$, which is -1 when $j = i$ and x when $i = j+1$, so it is lower triangular with -1 all along its diagonal and determinant $(-1)^{k-1}$. The cofactor signs are $(-1)^{1+1} = 1$ and $(-1)^{k+1}$, so

$$\det \mathbf{M} = x \cdot x^{k-1} + (-\varepsilon)(-1)^{k+1}(-1)^{k-1} = x^k - \varepsilon,$$

since $(-1)^{k+1}(-1)^{k-1} = (-1)^{2k} = 1$.

The eigenvalues. They are the k -th roots of ε , namely $\varepsilon^{1/k} e^{2\pi i j/k}$ for $j = 0, \dots, k-1$, each of modulus $\varepsilon^{1/k}$. For $k = 2$ this is the pair $\pm\sqrt{\varepsilon}$, and $\mathbf{A}_\varepsilon$ is the matrix $\mathbf{A}(\varepsilon)$ of Proposition 16.7.2.

The comparison. The perturbation $\mathbf{E} = \varepsilon \mathbf{E}_{k1}$ has a single nonzero entry, so its largest absolute row sum and largest absolute column sum are both ε , and $\mathbf{E}^* \mathbf{E} = \varepsilon^2 \mathbf{E}_{11}$ has largest eigenvalue ε^2 ; hence $\|\mathbf{E}\|_1 = \|\mathbf{E}\|_\infty = \|\mathbf{E}\|_2 = \varepsilon$ by Theorem 16.3.3. Every eigenvalue of $\mathbf{A}$ is 0, so the eigenvalues have moved a distance $\varepsilon^{1/k}$ while the matrix moved a distance ε . The ratio is

$$\frac{\varepsilon^{1/k}}{\varepsilon} = \varepsilon^{-(k-1)/k} \xrightarrow{\varepsilon \rightarrow 0^+} \infty.$$

A by-product worth keeping. The k roots $\varepsilon^{1/k} e^{2\pi i j/k}$ are **distinct** for every $\varepsilon > 0$, so $\mathbf{A}_\varepsilon$ has k distinct eigenvalues and is diagonalizable by Corollary 9.4.2. A single nilpotent Jordan block, which is as far from diagonalizable as a matrix gets, therefore has diagonalizable matrices arbitrarily close to it, and here they are written down. That is Corollary 16.7.5 made concrete for one family.

Put numbers on it. For $k = 2$ and $\varepsilon = 10^{-6}$ the matrix moves by one part in a million and the eigenvalues move to $\pm 10^{-3}$: an amplification of a thousand. For $k = 10$ and $\varepsilon = 10^{-10}$ the matrix moves by 10^{-10} and the eigenvalues land on the circle of radius 10^{-1} , an amplification of a billion. No constant L survives this, so the continuity of Corollary 16.7.4 cannot be upgraded to a Lipschitz bound, and the δ it produces genuinely depends on $\mathbf{A}$ and not only on n .

Warning. Continuous does not mean stable. Corollary 16.7.4 guarantees that a small enough perturbation moves the eigenvalues a little, and says nothing whatever about how small “small enough” is. For $\mathbf{J}_{10}(0)$ a perturbation of size 10^{-10} — the kind of change one would call negligible in data recorded to ten digits — scatters the eigenvalues over a circle of radius 0.1. Reading a printed eigenvalue of a matrix with a large Jordan block as a property of the matrix, rather than of the matrix plus whatever noise reached it, is the mistake this example exists to prevent.

The opposite extreme is worth seeing beside it. Take the Hermitian $\mathbf{A} = \text{diag}(1, 2)$ and the Hermitian perturbation $\mathbf{E} = \varepsilon(\mathbf{E}_{12} + \mathbf{E}_{21})$, of norm $\|\mathbf{E}\|_2 = \varepsilon$. The matrix $\mathbf{A} + \mathbf{E} = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 2 \end{pmatrix}$ has

trace 3 and determinant $2 - \varepsilon^2$, so its eigenvalues are

$$\frac{3 \pm \sqrt{1 + 4\varepsilon^2}}{2},$$

and since $\sqrt{1+u} \leq 1+u/2$ for $u \geq 0$, each of them lies within ε^2 of the corresponding eigenvalue of $\mathbf{A}$. Here the eigenvalues move *less* than the matrix does, by a whole order in ε . That is not an accident of the example: Chapter 17 proves that the eigenvalues of a Hermitian matrix move by at most $\|\mathbf{E}\|_2$ under a Hermitian perturbation, and Chapter 20 proves the theorem of Bauer and Fike, which bounds the movement for a diagonalizable matrix by $\|\mathbf{E}\|_2$ times a factor measuring how far its eigenvector basis is from orthonormal.

So the exponent $1/k$ in [Example \(A Jordan block's eigenvalues move like \$\varepsilon^{1/k}\$ \)](#) is not about size; it is about shape. It is decided by the largest Jordan block at the eigenvalue (Theorem 10.3.1), and block size is precisely what Chapter 12 §11 was measuring when [Example \(The defect of a Jordan block\)](#) computed

$$\Delta(\mathbf{J}_k(\lambda)) = \sqrt{k-1}$$

for the departure from normality of Definition 12.11.1. The two statements match at both ends: $\Delta = 0$ exactly for the normal matrices, whose eigenvalues are the stable ones, and Δ grows with the block that makes them fragile. They do not match in between, and it would be wrong to read Δ as a formula for the exponent — Δ is one number for the whole matrix, so it cannot report which eigenvalue carries which block, and [Example \(The defect of a Jordan block\)](#) records that it does not depend on λ at all. What $\Delta(\mathbf{A})$ does is tell you whether the question needs asking at all.

16.7.7 Exercises

16.7.7.1 A. Check your understanding

Exercise. A1

- State [Theorem 16.7.3](#), paying attention to the two hypotheses on q .
- Explain in one sentence why the theorem is stated for the multiset of roots rather than for a list of n individual roots.
- Decide whether the following is correct, and justify your answer: “the function sending $\mathbf{A} \in M_2(\mathbb{C})$ to its eigenvalue of largest modulus is continuous”.
- For the $k \times k$ Jordan block, how far do the eigenvalues move under a perturbation of norm ε , and what does that say about a Lipschitz bound?

Solution.

- See [Theorem 16.7.3](#). The two hypotheses are that q is **monic** and of the **same degree** n as p ; without them the roots can escape to infinity, as $\delta x^2 + x$ shows.
- Because no continuous choice of an individual root exists: [Proposition 16.7.2](#) shows that following one square root of t once around the unit circle returns the other one, so the matching numbering must be allowed to depend on q .
- Incorrect as stated, because a matrix can have two eigenvalues of equal largest modulus and then the phrase does not name a number; and incorrect even where it does name a number, since $\mathbf{A}(t)$ of [Proposition 16.7.2](#) has the two eigenvalues $\pm\sqrt{t}$ of equal modulus for every $t \neq 0$. The modulus itself, that is ρ , is continuous — see the Quick check.

- (d) They move a distance $\varepsilon^{1/k}$ ([Example \(A Jordan block's eigenvalues move like \$\varepsilon^{1/k}\$ \)](#)), so the ratio of eigenvalue movement to matrix movement is $\varepsilon^{-(k-1)/k}$, which is unbounded as $\varepsilon \rightarrow 0^+$. No Lipschitz constant exists.

16.7.7.2 B. Practice

Exercise. B1: Two perturbations of the same size

Let $\varepsilon = 10^{-8}$. Compute the eigenvalues of

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ \varepsilon & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \varepsilon & 1 \\ 0 & 0 \end{pmatrix},$$

each of which differs from $\mathbf{J}_2(0)$ by a matrix of 2-norm ε . Hence explain why the constant in [Corollary 16.7.4](#) cannot depend on $\|\mathbf{B} - \mathbf{A}\|$ alone.

Solution. For $\mathbf{B}$: $p_{\mathbf{B}}(x) = x^2 - \varepsilon$ by [Example \(A Jordan block's eigenvalues move like \$\varepsilon^{1/k}\$ \)](#) with $k = 2$, so the eigenvalues are $\pm 10^{-4}$.

For $\mathbf{C}$: the matrix is upper triangular, so its eigenvalues are its diagonal entries $\varepsilon = 10^{-8}$ and 0 (Theorem 9.7.4 (b)).

Both perturbations have 2-norm $\varepsilon = 10^{-8}$, since each is ε times a single matrix unit. The eigenvalues of $\mathbf{B}$ have moved 10^{-4} from $\text{spec}(\mathbf{J}_2(0)) = \{0\}$, those of $\mathbf{C}$ only 10^{-8} : a factor of 10^4 between two perturbations of exactly the same size. So the movement is not a function of $\|\mathbf{E}\|$, and the δ of [Corollary 16.7.4](#) cannot be computed from ε and n alone.

Exercise. B2: A diagonalizable neighbor, explicitly

Let $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ and let $\varepsilon > 0$.

- Show that $\mathbf{A}$ is not diagonalizable.
- Exhibit a diagonalizable $\mathbf{A}'$ with $\|\mathbf{A} - \mathbf{A}'\|_2 < \varepsilon$, and say which theorem certifies that your $\mathbf{A}'$ is diagonalizable.
- What are the eigenvalues of your $\mathbf{A}'$, and how far have they moved?

Solution.

- The only eigenvalue is 2, with algebraic multiplicity 2 (Theorem 9.7.4 (b)). But $\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ has rank 1, so $g_{\mathbf{A}}(2) = 1 < 2 = a_{\mathbf{A}}(2)$ and $\mathbf{A}$ is not diagonalizable by Theorem 9.4.1 (e).
- Take δ with $0 < \delta < \varepsilon$ and $\mathbf{A}' = \begin{pmatrix} 2 & 1 \\ 0 & 2+\delta \end{pmatrix}$. It is upper triangular with distinct diagonal entries 2 and $2 + \delta$, hence has two distinct eigenvalues (Theorem 9.7.4 (b)) and is diagonalizable by Corollary 9.4.2. And $\mathbf{A} - \mathbf{A}' = -\delta\mathbf{E}_{22}$, so $\|\mathbf{A} - \mathbf{A}'\|_2 = \delta < \varepsilon$ by Theorem 16.3.3 (c).
- They are 2 and $2 + \delta$. One eigenvalue has not moved and the other has moved δ , which

is exactly $\|\mathbf{A} - \mathbf{A}'\|_2$: here the movement is Lipschitz with constant 1. The failure in [Example \(A Jordan block's eigenvalues move like \$\varepsilon^{1/k}\$ \)](#) needed a perturbation *below* the diagonal, which is the direction a triangular form cannot absorb.

Exercise. B3: The determinant and the trace, by density

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$. Use [Corollary 16.7.6](#) to prove that $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ — after first checking that the map in question satisfies the hypothesis of that corollary — and then explain why this is a bad way to prove it.

Solution. Fix $\mathbf{B}$ and put $\Phi(\mathbf{A}) = \text{tr}(\mathbf{AB}) - \text{tr}(\mathbf{BA})$, a 1×1 matrix whose single entry is $\sum_{i,j} (a_{ij}b_{ji} - b_{ij}a_{ji})$, a polynomial in the entries of $\mathbf{A}$; so the hypothesis of [Corollary 16.7.6](#) holds.

Let $\mathbf{A} = \mathbf{PDP}^{-1}$ be diagonalizable, $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$, and put $\mathbf{X} = \mathbf{P}^{-1}\mathbf{BP}$. By Theorem 4.2.5, similar matrices have equal traces, and $\mathbf{AB} = \mathbf{P}(\mathbf{DX})\mathbf{P}^{-1}$ while $\mathbf{BA} = \mathbf{P}(\mathbf{XD})\mathbf{P}^{-1}$, so $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{DX})$ and $\text{tr}(\mathbf{BA}) = \text{tr}(\mathbf{XD})$. For a diagonal $\mathbf{D}$ both of those are $\sum_i d_i x_{ii}$, by inspection of the diagonal entries of the two products. Hence $\Phi(\mathbf{A}) = 0$ for every diagonalizable $\mathbf{A}$, and [Corollary 16.7.6](#) extends it to every $\mathbf{A}$.

It is a bad way to prove it because the identity is a two-line computation from the definition, $\text{tr}(\mathbf{AB}) = \sum_{i,j} a_{ij}b_{ji} = \text{tr}(\mathbf{BA})$, valid over every field; the density argument needs $\mathbb{C}$, an imported theorem, and two corollaries, and delivers less. A density argument earns its keep only when the diagonalizable case is genuinely easier, as in [Example \(Cayley–Hamilton, by a limit\)](#).

16.7.7.3 C. Going deeper

Exercise. C1: How close to singular a nearby eigenvalue forces you to be

Let $\mathbf{A} \in M_n(\mathbb{C})$, let $\|\cdot\|$ be an operator norm on $M_n(\mathbb{C})$ (Definition 16.3.2), and let $\mathbf{E} \in M_n(\mathbb{C})$ with $\mathbf{E} \neq \mathbf{0}$.

- (a) Suppose μ is an eigenvalue of $\mathbf{A} + \mathbf{E}$ and $\mu \notin \text{spec}(\mathbf{A})$. Prove that

$$\|(\mathbf{A} - \mu\mathbf{I})^{-1}\| \geq \frac{1}{\|\mathbf{E}\|}.$$

- (b) Deduce that if $\|(\mathbf{A} - z\mathbf{I})^{-1}\| \leq M$ for every z with $|z - \lambda| = r$, then no eigenvalue of $\mathbf{A} + \mathbf{E}$ lies on that circle as soon as $\|\mathbf{E}\| < 1/M$.

Hint: for (a), factor $\mathbf{A} + \mathbf{E} - \mu\mathbf{I}$ and use Theorem 16.5.1.

Solution.

- (a) Since $\mu \notin \text{spec}(\mathbf{A})$, the matrix $\mathbf{A} - \mu\mathbf{I}$ is invertible (Theorem 9.2.1). Write

$$\mathbf{A} + \mathbf{E} - \mu\mathbf{I} = (\mathbf{A} - \mu\mathbf{I})(\mathbf{I} + (\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{E}).$$

The left side is singular, because μ is an eigenvalue of $\mathbf{A} + \mathbf{E}$. The first factor on the

right is invertible, so the second factor is singular: a product with an invertible left factor is invertible as soon as the right factor is. Put $\mathbf{X} = -(\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{E}$, so $\mathbf{I} - \mathbf{X}$ is singular. By Theorem 16.5.1, $\|\mathbf{X}\| < 1$ would make $\mathbf{I} - \mathbf{X}$ invertible; hence $\|\mathbf{X}\| \geq 1$. Submultiplicativity (Theorem 16.3.2) gives

$$1 \leq \|\mathbf{X}\| \leq \|(\mathbf{A} - \mu\mathbf{I})^{-1}\| \|\mathbf{E}\|,$$

and dividing by $\|\mathbf{E}\| > 0$ gives the claim.

- (b) Suppose μ is an eigenvalue of $\mathbf{A} + \mathbf{E}$ with $|\mu - \lambda| = r$. Then $\|(\mathbf{A} - \mu\mathbf{I})^{-1}\| \leq M$ by hypothesis, and in particular $\mathbf{A} - \mu\mathbf{I}$ is invertible, so $\mu \notin \text{spec}(\mathbf{A})$ and part (a) applies: $M \geq 1/\|\mathbf{E}\|$, that is $\|\mathbf{E}\| \geq 1/M$. Contrapositively, $\|\mathbf{E}\| < 1/M$ leaves no eigenvalue of $\mathbf{A} + \mathbf{E}$ on the circle.

Exercise. C2: Density of the invertible matrices, and $\mathbf{AB}$ versus $\mathbf{BA}$

- (a) Prove that for every $\mathbf{A} \in M_n(\mathbb{C})$ and every $\varepsilon > 0$ there is an **invertible** $\mathbf{A}'$ with $\|\mathbf{A} - \mathbf{A}'\| < \varepsilon$, in any norm. *Hint: consider $\mathbf{A} - t\mathbf{I}$.*
- (b) Prove a version of Corollary 16.7.6 with “invertible” in place of “diagonalizable”.
- (c) Hence give a second proof that $p_{\mathbf{AB}} = p_{\mathbf{BA}}$ for all $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$, a statement first proved in Theorem 9.10.4 (b).

Solution.

- (a) The matrix $\mathbf{A} - t\mathbf{I}$ is singular exactly when $t \in \text{spec}(\mathbf{A})$ (Theorem 9.2.1), and $\text{spec}(\mathbf{A})$ has at most n elements. Let N_∞ and C be as in the proof of Corollary 16.7.5, so $\|\mathbf{M}\| \leq C N_\infty(\mathbf{M})$. Choose $t \neq 0$ with $|t| < \varepsilon/C$ and $t \notin \text{spec}(\mathbf{A})$, possible because the disc of radius ε/C is infinite and only finitely many of its points are forbidden. Then $\mathbf{A}' = \mathbf{A} - t\mathbf{I}$ is invertible and $N_\infty(\mathbf{A} - \mathbf{A}') = |t|$, so $\|\mathbf{A} - \mathbf{A}'\| < \varepsilon$.
- (b) Verbatim the proof of Corollary 16.7.6, with the sequence of diagonalizable matrices replaced by the sequence of invertible matrices supplied by (a) with $\varepsilon = 1/k$.
- (c) Fix $\mathbf{B}$ and set $\Phi(\mathbf{A}) = (c_0(\mathbf{AB}) - c_0(\mathbf{BA}), \dots, c_{n-1}(\mathbf{AB}) - c_{n-1}(\mathbf{BA}))$, a row of n entries. Each entry of $\mathbf{AB}$ and of $\mathbf{BA}$ is a polynomial in the entries of $\mathbf{A}$, and each c_j is a polynomial in the entries of its argument, so Φ satisfies the hypothesis of (b). For invertible $\mathbf{A}$ we have $\mathbf{BA} = \mathbf{A}^{-1}(\mathbf{AB})\mathbf{A}$, so $\mathbf{AB}$ and $\mathbf{BA}$ are similar and $p_{\mathbf{AB}} = p_{\mathbf{BA}}$ by Theorem 7.10.3; that is exactly $\Phi(\mathbf{A}) = \mathbf{O}$, and it is Exercise (B3: $\mathbf{AB}$ and $\mathbf{BA}$) (a). By (b), $\Phi(\mathbf{A}) = \mathbf{O}$ for every $\mathbf{A}$, and two monic polynomials of degree n with the same lower coefficients are equal.

Exercise. C3: Distinct eigenvalues is an open condition

Let $\mathbf{A} \in M_n(\mathbb{C})$ have n **distinct** eigenvalues $\lambda_1, \dots, \lambda_n$, and let $\|\cdot\|$ be a norm on $M_n(\mathbb{C})$.

- (a) Prove that there is $\delta > 0$ such that every $\mathbf{B}$ with $\|\mathbf{B} - \mathbf{A}\| < \delta$ also has n distinct eigenvalues, and is therefore diagonalizable.
- (b) Deduce that the set of matrices in $M_n(\mathbb{C})$ with a repeated eigenvalue contains no ball,

that is, that its complement is dense.

- (c) Is the set of diagonalizable matrices open? Justify your answer.

Solution.

- (a) Put $2\varepsilon = \min_{i \neq j} |\lambda_i - \lambda_j|$, which is positive because the λ_i are distinct and there are finitely many pairs. Let δ be the number that [Corollary 16.7.4](#) supplies for this ε . If $\|\mathbf{B} - \mathbf{A}\| < \delta$, number the eigenvalues of $\mathbf{B}$ so that $|\mu_i - \lambda_i| < \varepsilon$ for every i . For $i \neq j$, Theorem 0.6.4 gives

$$|\mu_i - \mu_j| \geq |\lambda_i - \lambda_j| - |\mu_i - \lambda_i| - |\mu_j - \lambda_j| > 2\varepsilon - \varepsilon - \varepsilon = 0,$$

so the μ_i are pairwise distinct. There are n of them, so $\mathbf{B}$ is diagonalizable by [Corollary 9.4.2](#).

- (b) Let $\mathbf{A}$ be arbitrary and $\varepsilon > 0$. By Exercise (C3: Distinct eigenvalues are dense, but not over the reals) (a) there is $\mathbf{A}'$ with n distinct eigenvalues and $|a_{ij} - a'_{ij}| < \varepsilon/C$ for all i, j , with C as in the proof of [Corollary 16.7.5](#); hence $\|\mathbf{A} - \mathbf{A}'\| < \varepsilon$. So every ball meets the set of matrices with n distinct eigenvalues, and no ball is contained in its complement.
- (c) No. The zero matrix is diagonalizable, and in $M_2(\mathbb{C})$ every ball around it contains $\varepsilon \mathbf{J}_2(0)$, whose only eigenvalue is 0, with eigenspace of dimension 1 and algebraic multiplicity 2; so it is not diagonalizable by [Theorem 9.4.1 \(e\)](#). Dense ([Corollary 16.7.5](#)) and open are independent properties, and the diagonalizable matrices are the first without being the second.