

20.11 Condition Numbers of Eigenvalues

The Bauer–Fike theorem of §04 bounds the movement of every eigenvalue of a diagonalizable matrix at once, by $\kappa_2(\mathbf{X})\|\mathbf{E}\|_2$ (Theorem 20.4.2). That is one number for the whole spectrum, and it depends on a choice of eigenvector matrix $\mathbf{X}$. §04 warned that a well-behaved eigenvalue can sit inside a badly conditioned $\mathbf{X}$, so the bound can be far too pessimistic for some eigenvalues. This section asks about **one** eigenvalue at a time. If λ is a simple eigenvalue of $\mathbf{A}$, how fast does it move when $\mathbf{A}$ is pushed in a direction $\mathbf{E}$? The answer is an exact derivative, $\mathbf{y}^*\mathbf{E}\mathbf{x}/(\mathbf{y}^*\mathbf{x})$, built from a right and a left eigenvector. Its worst size over all $\mathbf{E}$ with $\|\mathbf{E}\|_2 = 1$ is the condition number of that eigenvalue.

The field is $\mathbb{C}$. Throughout, $\mathbf{A}, \mathbf{E} \in M_n(\mathbb{C})$, and $\|\cdot\|$ on vectors is the Euclidean norm.

20.11.1 Left and right eigenvectors

Chapter 9 defined a left eigenvector as a nonzero $\mathbf{y}$ with $\mathbf{y}^\top \mathbf{A} = \lambda \mathbf{y}^\top$ (Definition 9.10.1). Over $\mathbb{C}$ it is more convenient to put a conjugate on it. In this section, a **left eigenvector** of $\mathbf{A}$ for λ is a nonzero $\mathbf{y} \in \mathbb{C}^n$ with

$$\mathbf{y}^* \mathbf{A} = \lambda \mathbf{y}^*,$$

which is the same as $\mathbf{A}^* \mathbf{y} = \bar{\lambda} \mathbf{y}$. Since $\mathbf{y}^* = \bar{\mathbf{y}}^\top$, the equation $\mathbf{y}^* \mathbf{A} = \lambda \mathbf{y}^*$ says exactly that $\bar{\mathbf{y}}$ is a left eigenvector in the sense of Definition 9.10.1. So the two conventions differ only by conjugating the vector. The conjugate is worth it for one reason: the number $\mathbf{y}^* \mathbf{x}$ is then the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ of Chapter 11, and Cauchy–Schwarz applies to it.

For a **simple** eigenvalue, that is, one of algebraic multiplicity 1 (Definition 9.3.1), the right and left eigenvectors are as rigid as they can be, and they are never orthogonal. That last fact is what the whole section divides by.

Lemma 20.11.1. Left and Right Eigenvectors of a Simple Eigenvalue

Let $\mathbf{A} \in M_n(\mathbb{C})$ and let λ be a simple eigenvalue of $\mathbf{A}$, with a right eigenvector $\mathbf{x}$ and a left eigenvector $\mathbf{y}$.

- (a) The right eigenvectors of $\mathbf{A}$ for λ are exactly the nonzero multiples of $\mathbf{x}$, and the left eigenvectors are exactly the nonzero multiples of $\mathbf{y}$.
- (b) $\mathbf{y}^* \mathbf{x} \neq 0$.

Idea. Split λ off from the rest of the spectrum. In a basis that begins with $\mathbf{x}$, the matrix becomes block triangular with the 1×1 block (λ) in the corner and a block $\mathbf{B}$ with $\lambda \notin \text{spec}(\mathbf{B})$. Chapter 12 §10 clears the off-diagonal corner whenever the two diagonal blocks have disjoint spectra, so $\mathbf{A}$ is similar to $(\lambda) \oplus \mathbf{B}$. In that picture everything is visible: $\mathbf{e}_1$ is the right eigenvector, $\mathbf{e}_1^*$ the left one, and $\mathbf{e}_1^* \mathbf{e}_1 = 1$.

Proof. If $n = 1$, both claims hold because every nonzero vector of $\mathbb{C}^1$ is a multiple of every other, and $\mathbf{y}^* \mathbf{x} = \bar{y}_1 x_1$ is a product of two nonzero numbers. Let $n \geq 2$.

By Theorem 1.6.5, extend the independent list $(\mathbf{x})$ to a basis of $\mathbb{C}^n$, and let $\mathbf{P}_0$ be the invertible matrix with these columns, so that $\mathbf{P}_0 \mathbf{e}_1 = \mathbf{x}$. Then $\mathbf{P}_0^{-1} \mathbf{A} \mathbf{P}_0 \mathbf{e}_1 = \mathbf{P}_0^{-1} \lambda \mathbf{x} = \lambda \mathbf{e}_1$,

so

$$\mathbf{P}_0^{-1}\mathbf{A}\mathbf{P}_0 = \begin{pmatrix} \lambda & \mathbf{c} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}$$

for some $\mathbf{c} \in M_{1 \times (n-1)}(\mathbb{C})$ and $\mathbf{B} \in M_{n-1}(\mathbb{C})$. By Theorem 7.10.3 and Theorem 7.8.4, $p_{\mathbf{A}}(x) = (x - \lambda)p_{\mathbf{B}}(x)$. If $p_{\mathbf{B}}(\lambda) = 0$, then $x - \lambda$ would divide $p_{\mathbf{B}}$ by the Factor Theorem (Lemma 3.7.1), so $(x - \lambda)^2$ would divide $p_{\mathbf{A}}$, against $a_{\mathbf{A}}(\lambda) = 1$. Hence $\lambda \notin \text{spec}(\mathbf{B})$ (Theorem 9.2.1). The spectra $\{\lambda\}$ and $\text{spec}(\mathbf{B})$ are disjoint, so by Corollary 12.10.3 (c) and (a) there is $\mathbf{Q} = \begin{pmatrix} 1 & \mathbf{z} \\ \mathbf{0} & \mathbf{I}_{n-1} \end{pmatrix}$ with $\mathbf{Q}^{-1}\mathbf{P}_0^{-1}\mathbf{A}\mathbf{P}_0\mathbf{Q} = (\lambda) \oplus \mathbf{B}$. Put $\mathbf{P} = \mathbf{P}_0\mathbf{Q}$. Then

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = (\lambda) \oplus \mathbf{B}, \quad \mathbf{P}\mathbf{e}_1 = \mathbf{P}_0\mathbf{Q}\mathbf{e}_1 = \mathbf{P}_0\mathbf{e}_1 = \mathbf{x}.$$

(a) Let $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ and write $\mathbf{P}^{-1}\mathbf{v} = (u_1, \mathbf{u}')$ with $\mathbf{u}' \in \mathbb{C}^{n-1}$. Then $((\lambda) \oplus \mathbf{B})\mathbf{P}^{-1}\mathbf{v} = \lambda\mathbf{P}^{-1}\mathbf{v}$, whose last $n - 1$ rows say $(\mathbf{B} - \lambda\mathbf{I})\mathbf{u}' = \mathbf{0}$. Since $\lambda \notin \text{spec}(\mathbf{B})$, the matrix $\mathbf{B} - \lambda\mathbf{I}$ is invertible (Theorem 9.2.1), so $\mathbf{u}' = \mathbf{0}$ and $\mathbf{v} = u_1\mathbf{P}\mathbf{e}_1 = u_1\mathbf{x}$. For the left side, let $\mathbf{w}^*\mathbf{A} = \lambda\mathbf{w}^*$ and write $\mathbf{w}^*\mathbf{P} = (w_1, \mathbf{w}')$. Then $(\mathbf{w}^*\mathbf{P})((\lambda) \oplus \mathbf{B}) = \lambda\mathbf{w}^*\mathbf{P}$, whose last $n - 1$ entries say $\mathbf{w}'^*(\mathbf{B} - \lambda\mathbf{I}) = \mathbf{0}$. Multiplying on the right by $(\mathbf{B} - \lambda\mathbf{I})^{-1}$ gives $\mathbf{w}' = \mathbf{0}$, so $\mathbf{w}^* = w_1\mathbf{e}_1^*\mathbf{P}^{-1}$. Every left eigenvector is therefore a multiple of the nonzero row $\mathbf{e}_1^*\mathbf{P}^{-1}$, and in particular every left eigenvector is a multiple of any other. This proves (a).

(b) By the computation just made, $\mathbf{y}^* = w_1\mathbf{e}_1^*\mathbf{P}^{-1}$ with $w_1 \neq 0$, because $\mathbf{y} \neq \mathbf{0}$. Hence

$$\mathbf{y}^*\mathbf{x} = w_1\mathbf{e}_1^*\mathbf{P}^{-1}\mathbf{P}\mathbf{e}_1 = w_1\mathbf{e}_1^*\mathbf{e}_1 = w_1 \neq 0.$$

This proves the lemma. ■

The proof produced a little more than was asked: the left eigenvector $\mathbf{e}_1^*\mathbf{P}^{-1}$ is normalized against $\mathbf{x}$, in the sense that $(\mathbf{e}_1^*\mathbf{P}^{-1})\mathbf{x} = 1$. This is the one-eigenvalue version of the biorthogonal normalization $\mathbf{y}_i^*\mathbf{v}_i = 1$ of Proposition 9.10.7, and it exists for a simple eigenvalue of any matrix, diagonalizable or not.

Warning. Simple is essential: at a Jordan block, the two eigenvectors are orthogonal. For $\mathbf{J}_2(0) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ the only eigenvalue is 0, twice. The right eigenvectors are the multiples of $\mathbf{x} = \mathbf{e}_1$, since $\mathbf{J}_2(0)\mathbf{e}_1 = \mathbf{0}$. The left eigenvectors are the multiples of $\mathbf{y} = \mathbf{e}_2$, since $\mathbf{e}_2^*\mathbf{J}_2(0)$ is the second row, which is zero. So $\mathbf{y}^*\mathbf{x} = 0$. Every formula in this section that divides by $\mathbf{y}^*\mathbf{x}$ is meaningless here, and §03's example shows why it must be: the eigenvalues of $\mathbf{J}_2(0) + t\mathbf{E}_{21}$ are $\pm\sqrt{t}$, which have no derivative at $t = 0$.

Quick check. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$. Find a right and a left eigenvector for the eigenvalue 1, and check that $\mathbf{y}^*\mathbf{x} \neq 0$.

Solution. $\mathbf{A}\mathbf{e}_1 = \mathbf{e}_1$, so $\mathbf{x} = \mathbf{e}_1$. A left eigenvector solves $\mathbf{y}^*(\mathbf{A} - \mathbf{I}) = \mathbf{0}$, and $\mathbf{A} - \mathbf{I} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$; for $\mathbf{y}^* = (a, b)$ this reads $(0, a + b) = (0, 0)$, so $\mathbf{y} = (1, -1)$ will do. Then $\mathbf{y}^*\mathbf{x} = 1 \neq 0$, as Lemma 20.11.1 (b) requires. Here $\mathbf{y}$ is not a multiple of $\mathbf{x}$: the matrix is not normal.

20.11.2 The derivative of a simple eigenvalue

Now move $\mathbf{A}$ along a straight line $\mathbf{A} + t\mathbf{E}$, with t real. Chapter 16 proved that the eigenvalues move continuously (Corollary 16.7.4), but only as an unnumbered list, and Proposition 16.7.2 showed that no single eigenvalue can be followed continuously in general. A simple eigenvalue is the exception: it stays isolated from the rest for small t , so “the eigenvalue near λ ” makes sense. The theorem says that this eigenvalue is differentiable at $t = 0$, and it names the derivative.

Theorem 20.11.2. Derivative of a Simple Eigenvalue

Let $\mathbf{A} \in M_n(\mathbb{C})$, let λ be a simple eigenvalue of $\mathbf{A}$ with a right eigenvector $\mathbf{x}$ and a left eigenvector $\mathbf{y}$, and let $\mathbf{E} \in M_n(\mathbb{C})$. Then there are $r > 0$ and $t_0 > 0$ with the following properties.

- (a) For every **real** t with $|t| < t_0$, the matrix $\mathbf{A} + t\mathbf{E}$ has **exactly one** eigenvalue, counted with algebraic multiplicity, in the open disc $\{z \in \mathbb{C} : |z - \lambda| < r\}$. Call it $\lambda(t)$. Thus $\lambda(0) = \lambda$, and $\lambda(t)$ is a simple eigenvalue of $\mathbf{A} + t\mathbf{E}$.
- (b) $\lambda(t) \rightarrow \lambda$ as $t \rightarrow 0$.
- (c) The function $t \mapsto \lambda(t)$ is differentiable at $t = 0$, and

$$\lambda'(0) = \lim_{t \rightarrow 0} \frac{\lambda(t) - \lambda}{t} = \frac{\mathbf{y}^* \mathbf{E} \mathbf{x}}{\mathbf{y}^* \mathbf{x}}.$$

In other words, $\lambda(t) = \lambda + t \mathbf{y}^* \mathbf{E} \mathbf{x} / (\mathbf{y}^* \mathbf{x}) + \rho(t)$ with $\rho(t)/t \rightarrow 0$. The denominator is nonzero by Lemma 20.11.1 (b), and the quotient does not change when $\mathbf{x}$ or $\mathbf{y}$ is rescaled, which by part (a) of the lemma is the only freedom there is.

Idea. The target identity falls out of one line if we have eigenvectors $\mathbf{x}_t$ of $\mathbf{A} + t\mathbf{E}$ that converge to an eigenvector of $\mathbf{A}$: multiply $(\mathbf{A} + t\mathbf{E})\mathbf{x}_t = \lambda(t)\mathbf{x}_t$ on the left by $\mathbf{y}^*$ and use $\mathbf{y}^* \mathbf{A} = \lambda \mathbf{y}^*$, to get

$$(\lambda(t) - \lambda) \mathbf{y}^* \mathbf{x}_t = t \mathbf{y}^* \mathbf{E} \mathbf{x}_t.$$

Divide by $t \mathbf{y}^* \mathbf{x}_t$ and let $t \rightarrow 0$. So the plan has three steps.

- ① Isolate $\lambda(t)$ from the rest of the spectrum with Corollary 16.7.4.
- ② Produce the eigenvectors $\mathbf{x}_t$ **by a formula**. A column of the adjugate $\text{adj}(\lambda(t)\mathbf{I} - \mathbf{A} - t\mathbf{E})$ is an eigenvector, and its entries are polynomials in $\lambda(t)$ and t , so it converges because $\lambda(t)$ does.
- ③ Check that the limit is a nonzero multiple of $\mathbf{x}$, and pass to the limit in the quotient.

Proof. Let $\lambda_1 = \lambda, \lambda_2, \dots, \lambda_n$ be the eigenvalues of $\mathbf{A}$, listed with algebraic multiplicity. Since λ is simple, $\lambda_j \neq \lambda$ for $j \geq 2$.

Step 1: the isolated eigenvalue. Put $r = \frac{1}{2} \min_{j \geq 2} |\lambda_j - \lambda|$ if $n \geq 2$, and $r = 1$ if $n = 1$; in both cases $r > 0$. By Corollary 16.7.4 with the norm $\|\cdot\|_2$ and with $\varepsilon = r$, there is $\delta > 0$ such that every $\mathbf{B}$ with $\|\mathbf{B} - \mathbf{A}\|_2 < \delta$ has eigenvalues $\mu_1, \dots, \mu_n$ that can be numbered so that $|\mu_i - \lambda_i| < r$ for every i . Put $t_0 = \delta / (\|\mathbf{E}\|_2 + 1)$. If $|t| < t_0$, then $\|(\mathbf{A} + t\mathbf{E}) - \mathbf{A}\|_2 = |t| \|\mathbf{E}\|_2 < \delta$,

so the eigenvalues of $\mathbf{A} + t\mathbf{E}$ can be so numbered. Then $|\mu_1 - \lambda| < r$, while for $j \geq 2$

$$|\mu_j - \lambda| \geq |\lambda_j - \lambda| - |\mu_j - \lambda_j| > 2r - r = r$$

by Theorem 0.6.4. So exactly one entry of the list lies in the open disc of radius r about λ . This proves (a).

Step 2: continuity. Let $0 < \varepsilon < r$. By Corollary 16.7.4 again, now with this ε , there is $\delta' > 0$ such that for $|t|\|\mathbf{E}\|_2 < \delta'$ the eigenvalues of $\mathbf{A} + t\mathbf{E}$ can be numbered as $\mu'_1, \dots, \mu'_n$ with $|\mu'_i - \lambda_i| < \varepsilon$. If moreover $|t| < t_0$, then μ'_1 is an eigenvalue of $\mathbf{A} + t\mathbf{E}$ in the disc of radius $\varepsilon < r$ about λ , so $\mu'_1 = \lambda(t)$ by (a). Hence $|\lambda(t) - \lambda| < \varepsilon$ whenever $|t| < \min(t_0, \delta' / (\|\mathbf{E}\|_2 + 1))$. This proves (b).

Step 3: an eigenvector given by a formula. Put $\mathbf{N}(t) = \lambda(t)\mathbf{I} - \mathbf{A} - t\mathbf{E}$ for $|t| < t_0$, so $\mathbf{N}(0) = \lambda\mathbf{I} - \mathbf{A}$. We first find a column of $\text{adj } \mathbf{N}(0)$ that is not zero. If $n = 1$, then $\text{adj } \mathbf{N}(0) = (1)$ by Definition 7.6.2. If $n \geq 2$, then $1 \leq g_{\mathbf{A}}(\lambda) \leq a_{\mathbf{A}}(\lambda) = 1$ by Theorem 9.3.1, so $\text{rank } \mathbf{N}(0) = n - 1 \geq 1$ by Theorem 3.5.7. By Theorem 7.7.1 (a), $\mathbf{N}(0)$ has a nonzero $(n-1) \times (n-1)$ minor M_{ij} . Then the cofactor $C_{ij} = (-1)^{i+j}M_{ij}$ is nonzero, and it is the (j, i) entry of $\text{adj } \mathbf{N}(0)$ (Definition 7.6.2). In either case some column k of $\text{adj } \mathbf{N}(0)$ is nonzero. Define

$$\mathbf{x}_t = \text{adj}(\mathbf{N}(t)) \mathbf{e}_k \quad (|t| < t_0).$$

Since $\lambda(t)$ is an eigenvalue of $\mathbf{A} + t\mathbf{E}$, we have $\det \mathbf{N}(t) = p_{\mathbf{A}+t\mathbf{E}}(\lambda(t)) = 0$ (Theorem 9.2.1). So Theorem 7.6.2 gives $\mathbf{N}(t) \text{adj } \mathbf{N}(t) = \mathbf{O}$, and in particular

$$(\mathbf{A} + t\mathbf{E}) \mathbf{x}_t = \lambda(t) \mathbf{x}_t \quad (|t| < t_0).$$

At $t = 0$ this says $\mathbf{A}\mathbf{x}_0 = \lambda\mathbf{x}_0$, and $\mathbf{x}_0 \neq \mathbf{0}$ by the choice of k . By Lemma 20.11.1 (a), $\mathbf{x}_0 = c\mathbf{x}$ for some $c \neq 0$.

Next, $\mathbf{x}_t \rightarrow \mathbf{x}_0$ as $t \rightarrow 0$. Each entry of $\text{adj } \mathbf{B}$ is $\pm$ the determinant of a submatrix of $\mathbf{B}$, or the constant 1 when $n = 1$, hence a polynomial in the entries of $\mathbf{B}$ by Theorem 7.4.3. So $\mathbf{B} \mapsto \text{adj } \mathbf{B}$ is continuous by Lemma 16.7.1. And $\|\mathbf{N}(t) - \mathbf{N}(0)\|_2 \leq |\lambda(t) - \lambda| + |t|\|\mathbf{E}\|_2$, which tends to 0 by (b). Hence $\text{adj } \mathbf{N}(t) \rightarrow \text{adj } \mathbf{N}(0)$, and $\|\mathbf{x}_t - \mathbf{x}_0\| \leq \|\text{adj } \mathbf{N}(t) - \text{adj } \mathbf{N}(0)\|_2$ by Theorem 16.3.2 (a), since $\|\mathbf{e}_k\| = 1$. So $\mathbf{x}_t \rightarrow \mathbf{x}_0$.

Step 4: the derivative. Multiply the eigen-equation of Step 3 on the left by $\mathbf{y}^*$, and use $\mathbf{y}^* \mathbf{A} = \lambda \mathbf{y}^*$:

$$\lambda \mathbf{y}^* \mathbf{x}_t + t \mathbf{y}^* \mathbf{E} \mathbf{x}_t = \lambda(t) \mathbf{y}^* \mathbf{x}_t, \quad \text{so} \quad (\lambda(t) - \lambda) \mathbf{y}^* \mathbf{x}_t = t \mathbf{y}^* \mathbf{E} \mathbf{x}_t.$$

By Cauchy–Schwarz (Theorem 11.1.5), $|\mathbf{y}^* \mathbf{x}_t - \mathbf{y}^* \mathbf{x}_0| \leq \|\mathbf{y}^*\| \|\mathbf{x}_t - \mathbf{x}_0\| \rightarrow 0$, and likewise $\mathbf{y}^* \mathbf{E} \mathbf{x}_t \rightarrow \mathbf{y}^* \mathbf{E} \mathbf{x}_0$, because $\|\mathbf{E} \mathbf{x}_t - \mathbf{E} \mathbf{x}_0\| \leq \|\mathbf{E}\|_2 \|\mathbf{x}_t - \mathbf{x}_0\|$. The limit $\mathbf{y}^* \mathbf{x}_0 = c \mathbf{y}^* \mathbf{x}$ is nonzero by Lemma 20.11.1 (b). So there is $t_1 \in (0, t_0]$ with $|\mathbf{y}^* \mathbf{x}_t - \mathbf{y}^* \mathbf{x}_0| < |\mathbf{y}^* \mathbf{x}_0|$, and hence $\mathbf{y}^* \mathbf{x}_t \neq 0$, for $|t| < t_1$. For $0 < |t| < t_1$ we may divide:

$$\begin{aligned} \frac{\lambda(t) - \lambda}{t} &= \frac{\mathbf{y}^* \mathbf{E} \mathbf{x}_t}{\mathbf{y}^* \mathbf{x}_t} \longrightarrow \frac{\mathbf{y}^* \mathbf{E} \mathbf{x}_0}{\mathbf{y}^* \mathbf{x}_0} \quad (t \rightarrow 0), \\ \frac{\mathbf{y}^* \mathbf{E} \mathbf{x}_0}{\mathbf{y}^* \mathbf{x}_0} &= \frac{c \mathbf{y}^* \mathbf{E} \mathbf{x}}{c \mathbf{y}^* \mathbf{x}} = \frac{\mathbf{y}^* \mathbf{E} \mathbf{x}}{\mathbf{y}^* \mathbf{x}}, \end{aligned}$$

by the algebra of limits, the denominator's limit being nonzero. This proves (c). ■

No implicit function theorem was needed. The usual textbook route to this formula differentiates $\det(z\mathbf{I} - \mathbf{A} - t\mathbf{E}) = 0$ implicitly, which presupposes that a differentiable solution $z = \lambda(t)$ exists; that is the implicit function theorem, a piece of analysis this book has not imported. The proof above needed only the continuity of Corollary 16.7.4 and the fact that the adjugate writes down an eigenvector whose entries are polynomials. The analysis consists of limits of polynomials, which is the algebra of limits.

Remark. Since $\lambda(s)$ is a simple eigenvalue of $\mathbf{A} + s\mathbf{E}$ for every $|s| < t_0$, the theorem applies again at $\mathbf{A} + s\mathbf{E}$ in place of $\mathbf{A}$. The eigenvalue it isolates there is $\lambda(s+h)$ for small h , by the uniqueness in (a). So λ is differentiable on the whole interval $(-t_0, t_0)$, with $\lambda'(s) = \mathbf{y}_s^* \mathbf{E} \mathbf{x}_s / (\mathbf{y}_s^* \mathbf{x}_s)$ for right and left eigenvectors $\mathbf{x}_s, \mathbf{y}_s$ of $\mathbf{A} + s\mathbf{E}$ for $\lambda(s)$. We need only $s = 0$.

The formula has a transparent special case. If $\mathbf{A}$ is Hermitian, then $\mathbf{A}^* \mathbf{x} = \mathbf{A} \mathbf{x} = \lambda \mathbf{x}$ with λ real, so $\mathbf{x}$ is also a left eigenvector and we may take $\mathbf{y} = \mathbf{x}$:

$$\lambda'(0) = \frac{\mathbf{x}^* \mathbf{E} \mathbf{x}}{\mathbf{x}^* \mathbf{x}}.$$

When $\mathbf{E}$ is Hermitian too, this is the Rayleigh quotient $R_{\mathbf{E}}(\mathbf{x})$ of Chapter 17 (Definition 17.1.1), which lies in $[\lambda_n(\mathbf{E}), \lambda_1(\mathbf{E})]$ by Proposition 17.1.1. So $|\lambda'(0)| \leq \|\mathbf{E}\|_2$ by Lemma 17.3.5, which is the first-order shadow of Weyl's bound $|\lambda_i(\mathbf{A} + \mathbf{E}) - \lambda_i(\mathbf{A})| \leq \|\mathbf{E}\|_2$ (Corollary 17.3.6).

Quick check. Let $\mathbf{A} = \text{diag}(d_1, \dots, d_n)$ with distinct d_i , and let $\mathbf{E} = (e_{ij})$ be arbitrary. What is the derivative at $t = 0$ of the eigenvalue of $\mathbf{A} + t\mathbf{E}$ near d_i ?

Solution. The eigenvalue d_i is simple, with $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_i$, since $\mathbf{e}_i^* \mathbf{A} = d_i \mathbf{e}_i^*$. So $\lambda'(0) = \mathbf{e}_i^* \mathbf{E} \mathbf{e}_i / (\mathbf{e}_i^* \mathbf{e}_i) = e_{ii}$. To first order, the eigenvalues of a perturbed diagonal matrix with distinct diagonal entries move by the diagonal of the perturbation, and the off-diagonal entries of $\mathbf{E}$ enter only at second order.

20.11.3 The condition number of a simple eigenvalue

The derivative $\mathbf{y}^* \mathbf{E} \mathbf{x} / (\mathbf{y}^* \mathbf{x})$ depends on the direction $\mathbf{E}$. To measure how sensitive λ itself is, bound it over all directions of a given size. By Cauchy–Schwarz and the definition of $\|\mathbf{E}\|_2$,

$$|\lambda'(0)| = \frac{|\mathbf{y}^* \mathbf{E} \mathbf{x}|}{|\mathbf{y}^* \mathbf{x}|} \leq \frac{\|\mathbf{y}\| \|\mathbf{E} \mathbf{x}\|}{|\mathbf{y}^* \mathbf{x}|} \leq \frac{\|\mathbf{x}\| \|\mathbf{y}\|}{|\mathbf{y}^* \mathbf{x}|} \|\mathbf{E}\|_2.$$

The factor in front of $\|\mathbf{E}\|_2$ depends on $\mathbf{A}$ and λ only. It recurs in every estimate below, so we give it a name.

The condition number of an eigenvalue is the largest factor by which a perturbation of the matrix can move that eigenvalue, to first order.

Definition 20.11.1. Condition Number of a Simple Eigenvalue

Let $\mathbf{A} \in M_n(\mathbb{C})$ and let λ be a **simple** eigenvalue of $\mathbf{A}$, with a right eigenvector $\mathbf{x}$ and a left

eigenvector $\mathbf{y}$, so $\mathbf{y}^* \mathbf{A} = \lambda \mathbf{y}^*$. The **condition number** of λ is

$$\kappa_{\mathbf{A}}(\lambda) := \frac{\|\mathbf{x}\| \|\mathbf{y}\|}{|\mathbf{y}^* \mathbf{x}|}.$$

In words: take any right eigenvector and any left eigenvector for λ , multiply their lengths, and divide by the modulus of their inner product. The definition is made only for a **simple** eigenvalue, which is what makes the denominator nonzero (Lemma 20.11.1 (b)). It is also what makes the number well defined: by part (a) of the lemma, any other choice is $\alpha \mathbf{x}$ and $\beta \mathbf{y}$ with $\alpha, \beta \neq 0$, and then

$$\frac{\|\alpha \mathbf{x}\| \|\beta \mathbf{y}\|}{|(\beta \mathbf{y})^* (\alpha \mathbf{x})|} = \frac{|\alpha| |\beta| \|\mathbf{x}\| \|\mathbf{y}\|}{|\bar{\beta} \alpha| |\mathbf{y}^* \mathbf{x}|} = \kappa_{\mathbf{A}}(\lambda).$$

The subscript $\mathbf{A}$ and the argument λ are part of the notation. They keep $\kappa_{\mathbf{A}}(\lambda)$ apart from the condition number $\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$ of a matrix (Definition 16.8.1).

Examples.

- **A 1×1 matrix.** For $\mathbf{A} = (a)$ the eigenvalue a is simple, with $x = y = 1$, so $\kappa_{\mathbf{A}}(a) = 1$. This is the degenerate case, and it is the right answer: the eigenvalue of $(a + te)$ is $a + te$, which moves exactly as fast as the matrix.
- **A diagonal matrix with distinct entries.** For d_i , the Quick check above found $\mathbf{x} = \mathbf{y} = \mathbf{e}_i$, so $\kappa_{\mathbf{A}}(d_i) = 1$. Every eigenvalue is perfectly conditioned.
- **A triangular 2×2 .** For $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ and the eigenvalue 1, the Quick check after the lemma found $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y} = (1, -1)$, with $\mathbf{y}^* \mathbf{x} = 1$. So $\kappa_{\mathbf{A}}(1) = 1 \cdot \sqrt{2}/1 = \sqrt{2}$. The single off-diagonal entry has made the eigenvalue somewhat more sensitive.

Nonexample by minimal change. Change the entry 2 of the last example to $1 + s$, and let $s \rightarrow 0$. For real $s \neq 0$ the matrix $\mathbf{B}_s = \begin{pmatrix} 1 & 1 \\ 0 & 1+s \end{pmatrix}$ has the simple eigenvalue 1, with $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y} = (s, -1)$, so $\kappa_{\mathbf{B}_s}(1) = \sqrt{1+s^2}/|s|$. At $s = 0$ the matrix is $\mathbf{J}_2(1)$, and the eigenvalue 1 is no longer simple. The right and left eigenvectors still exist, $\mathbf{e}_1$ and $\mathbf{e}_2$, but $\mathbf{y}^* \mathbf{x} = 0$, so the clause “simple” fails and the quotient is undefined. The formula for $s \neq 0$ shows that it blows up on the way.

Why this definition. Three choices in it can be checked.

- **Both lengths go in the numerator.** This makes $\kappa_{\mathbf{A}}(\lambda)$ independent of the scaling of $\mathbf{x}$ and $\mathbf{y}$, as just computed. A normalization such as $\mathbf{y}^* \mathbf{x} = 1$ alone would leave one length free.
- **The norm is the Euclidean one,** because then the bound $|\lambda'(0)| \leq \kappa_{\mathbf{A}}(\lambda) \|\mathbf{E}\|_2$ above is attained by some $\mathbf{E}$, as part (b) of the proposition below shows. So $\kappa_{\mathbf{A}}(\lambda)$ is the exact worst case, not merely an upper bound. The same number is also the worst case over $\|\mathbf{E}\|_F = 1$: the bound survives because $\|\mathbf{E}\|_2 \leq \|\mathbf{E}\|_F$ (Proposition 16.3.4), and the extremal $\mathbf{E}$ has rank one, so its two norms agree by the same proposition.
- **The name matches Definition 16.8.1.** There, $\kappa(\mathbf{A})$ is the worst factor by which $\mathbf{A}$ magnifies a relative error in $\mathbf{b}$ into a relative error in the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$. Here $\kappa_{\mathbf{A}}(\lambda)$ is the worst factor by which a change in the matrix is magnified into a change in one eigenvalue, to first order.

Warning. **A condition number is a first-order statement, not a bound.** It governs $\lambda(t)$ only for small t , through $\lambda(t) - \lambda = t\lambda'(0) + \rho(t)$, and the remainder $\rho(t)$ can be as large as the first-order term while $t\|\mathbf{E}\|_2$ still looks small. Take $\mathbf{A} = \begin{pmatrix} 1 & M \\ 0 & 2 \end{pmatrix}$ with $M > 0$, for which $\kappa_{\mathbf{A}}(1) = \sqrt{1+M^2}$, as the worked example below computes, and $\mathbf{E} = \mathbf{E}_{21}$. At $t = -1/(4M)$ the matrix $\mathbf{A} + t\mathbf{E}$ has trace 3 and determinant $2 + \frac{1}{4}$, so both its eigenvalues

equal $\frac{3}{2}$. The eigenvalue 1 has moved by $\frac{1}{2}$, while the first-order prediction is $\kappa_{\mathbf{A}}(1) |t| \|\mathbf{E}\|_2 = \sqrt{1 + M^2}/(4M)$, which is close to $\frac{1}{4}$ for large M . At that point the eigenvalue is no longer simple, and the square root in the exact formula $(3 - \sqrt{1 + 4Mt})/2$ has no derivative there.

The first payoff is a list of properties, which together say what kind of number $\kappa_{\mathbf{A}}(\lambda)$ is.

Proposition 20.11.3. Properties of the Eigenvalue Condition Number

Let $\mathbf{A} \in M_n(\mathbb{C})$ and let λ be a simple eigenvalue of $\mathbf{A}$ with right eigenvector $\mathbf{x}$ and left eigenvector $\mathbf{y}$.

- (a) $\kappa_{\mathbf{A}}(\lambda)$ does not depend on the choice of $\mathbf{x}$ and $\mathbf{y}$, and $\kappa_{\mathbf{A}}(\lambda) \geq 1$, with equality if and only if $\mathbf{y}$ is a multiple of $\mathbf{x}$.
- (b) For every $\mathbf{E} \in M_n(\mathbb{C})$, the eigenvalue $\lambda(t)$ of Theorem 20.11.2 satisfies $|\lambda'(0)| \leq \kappa_{\mathbf{A}}(\lambda) \|\mathbf{E}\|_2$. Equality holds for $\mathbf{E} = \mathbf{y}\mathbf{x}^*/(\|\mathbf{x}\|\|\mathbf{y}\|)$, which has $\|\mathbf{E}\|_2 = 1$. Hence $\kappa_{\mathbf{A}}(\lambda) = \max\{|\lambda'(0)| : \|\mathbf{E}\|_2 = 1\}$.
- (c) If $\mathbf{A}$ is normal, then $\kappa_{\mathbf{A}}(\mu) = 1$ for every simple eigenvalue μ of $\mathbf{A}$.
- (d) If $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$ with $\mathbf{X}$ invertible and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, and λ_i is simple, then $\kappa_{\mathbf{A}}(\lambda_i) \leq \kappa_2(\mathbf{X})$.
- (e) $\kappa_{\mathbf{A}}(\lambda) = 1/\cos \theta$, where $\theta \in [0, \pi/2)$ is the principal angle between the lines $\text{span}(\mathbf{x})$ and $\text{span}(\mathbf{y})$ (Definition 20.8.1).

Proof.

- (a) Independence of the choice was checked after Definition 20.11.1. By Theorem 11.1.5, $|\mathbf{y}^*\mathbf{x}| = |\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\|$, so $\kappa_{\mathbf{A}}(\lambda) \geq 1$, with equality if and only if $(\mathbf{x}, \mathbf{y})$ is linearly dependent, that is, since both are nonzero, if and only if $\mathbf{y}$ is a multiple of $\mathbf{x}$.
- (b) The inequality is the display before Definition 20.11.1, with $\lambda'(0)$ given by Theorem 20.11.2 (c). For the equality, put $\mathbf{E} = \mathbf{y}\mathbf{x}^*/(\|\mathbf{x}\|\|\mathbf{y}\|)$. For every $\mathbf{v}$, $\mathbf{y}\mathbf{x}^*\mathbf{v} = (\mathbf{x}^*\mathbf{v})\mathbf{y}$, so $\|\mathbf{y}\mathbf{x}^*\mathbf{v}\| = |\langle \mathbf{v}, \mathbf{x} \rangle| \|\mathbf{y}\| \leq \|\mathbf{x}\|\|\mathbf{y}\|\|\mathbf{v}\|$ by Cauchy–Schwarz, with equality at $\mathbf{v} = \mathbf{x}$. Hence $\|\mathbf{y}\mathbf{x}^*\|_2 = \|\mathbf{x}\|\|\mathbf{y}\|$ by Definition 16.3.2, and $\|\mathbf{E}\|_2 = 1$. Moreover

$$\mathbf{y}^*\mathbf{E}\mathbf{x} = \frac{(\mathbf{y}^*\mathbf{y})(\mathbf{x}^*\mathbf{x})}{\|\mathbf{x}\|\|\mathbf{y}\|} = \|\mathbf{x}\|\|\mathbf{y}\|, \quad \text{so} \quad |\lambda'(0)| = \frac{\|\mathbf{x}\|\|\mathbf{y}\|}{|\mathbf{y}^*\mathbf{x}|} = \kappa_{\mathbf{A}}(\lambda).$$

The maximum formula restates the inequality and its equality case.

- (c) Let $\mathbf{A}\mathbf{x} = \mu\mathbf{x}$ with $\mathbf{x} \neq \mathbf{0}$. By Theorem 12.3.3, $\mathbf{A}^*\mathbf{x} = \bar{\mu}\mathbf{x}$, and taking conjugate transposes gives $\mathbf{x}^*\mathbf{A} = \mu\mathbf{x}^*$. So $\mathbf{x}$ is also a left eigenvector for μ , and (a) gives $\kappa_{\mathbf{A}}(\mu) = 1$.
- (d) Write $\mathbf{x}_i = \mathbf{X}\mathbf{e}_i$ and $\mathbf{y}_i^* = \mathbf{e}_i^*\mathbf{X}^{-1}$. From $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}$ and $\mathbf{X}^{-1}\mathbf{A} = \mathbf{\Lambda}\mathbf{X}^{-1}$, column i and row i give $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$ and $\mathbf{y}_i^*\mathbf{A} = \lambda_i\mathbf{y}_i^*$, and both vectors are nonzero because $\mathbf{X}$ is invertible. Also $\mathbf{y}_i^*\mathbf{x}_i = \mathbf{e}_i^*\mathbf{X}^{-1}\mathbf{X}\mathbf{e}_i = 1$. Now $\|\mathbf{x}_i\| \leq \|\mathbf{X}\|_2\|\mathbf{e}_i\| = \|\mathbf{X}\|_2$ by Theorem 16.3.2 (a). For $\mathbf{y}_i$, Cauchy–Schwarz gives

$$\|\mathbf{y}_i\|^2 = \mathbf{y}_i^*\mathbf{y}_i = \mathbf{e}_i^*\mathbf{X}^{-1}\mathbf{y}_i \leq \|\mathbf{X}^{-1}\mathbf{y}_i\| \leq \|\mathbf{X}^{-1}\|_2\|\mathbf{y}_i\|,$$

so $\|\mathbf{y}_i\| \leq \|\mathbf{X}^{-1}\|_2$, after dividing by $\|\mathbf{y}_i\| > 0$. Therefore $\kappa_{\mathbf{A}}(\lambda_i) = \|\mathbf{x}_i\|\|\mathbf{y}_i\|/1 \leq \|\mathbf{X}\|_2\|\mathbf{X}^{-1}\|_2 = \kappa_2(\mathbf{X})$, by (a) and Definition 16.8.1.

- (e) The vectors $\mathbf{x}/\|\mathbf{x}\|$ and $\mathbf{y}/\|\mathbf{y}\|$ are orthonormal bases of the two lines, so by Definition 20.8.1 the single principal angle satisfies $\cos \theta = \sigma_1(c)$, where $c = \mathbf{y}^* \mathbf{x} / (\|\mathbf{x}\| \|\mathbf{y}\|)$ is a 1×1 matrix. Its only singular value is $|c|$, the square root of the eigenvalue $\bar{c}c$ of c^*c (Definition 13.9.1). Hence $\cos \theta = |\mathbf{y}^* \mathbf{x}| / (\|\mathbf{x}\| \|\mathbf{y}\|) = 1/\kappa_{\mathbf{A}}(\lambda)$, which is positive by Lemma 20.11.1 (b), so $\theta < \pi/2$. This proves the proposition. ■

Part (e) is the picture to keep. An eigenvalue is well conditioned when its right and left eigenvectors point in nearly the same direction, and ill conditioned when they are nearly perpendicular. For a normal matrix they coincide, by (c). At a Jordan block they are exactly perpendicular, by the warning after the lemma.

Part (d) places the new number beside §04. Bauer–Fike bounds the movement of every eigenvalue by $\kappa_2(\mathbf{X})\|\mathbf{E}\|_2$, for every size of $\mathbf{E}$. Part (d) says that, to first order, each simple eigenvalue individually does at least as well, with its own constant $\kappa_{\mathbf{A}}(\lambda_i)$, and that constant involves no choice of $\mathbf{X}$. This is the refinement §04's warning promised. The first-order statement is sharper. The Bauer–Fike bound is global and needs no smallness.

20.11.4 A worked example

Upper triangular 2×2 matrices show everything, because their eigenvectors can be written down at once.

Example. The condition numbers of a triangular 2×2 matrix

Let $a, b, d \in \mathbb{C}$ with $a \neq d$, and $\mathbf{A} = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$.

- Compute $\kappa_{\mathbf{A}}(a)$ and $\kappa_{\mathbf{A}}(d)$.
- For $a = 1$, $b = M > 0$, $d = 2$, compare with the smallest value of $\kappa_2(\mathbf{X})$ over all diagonalizing $\mathbf{X}$.
- For the same matrix and $\mathbf{E} = \mathbf{E}_{21}$, compute $\lambda'(0)$ for the eigenvalue 1 from Theorem 20.11.2, and compare with the exact eigenvalues of $\mathbf{A} + t\mathbf{E}$ for $M = 100$ and $t = 10^{-4}$.

Solution.

- (a) Both eigenvalues are simple, since $a \neq d$. For a : $\mathbf{A}\mathbf{e}_1 = a\mathbf{e}_1$, so $\mathbf{x} = \mathbf{e}_1$. A left eigenvector solves $\mathbf{y}^*(\mathbf{A} - a\mathbf{I}) = \mathbf{0}$, where $\mathbf{A} - a\mathbf{I} = \begin{pmatrix} 0 & b \\ 0 & d-a \end{pmatrix}$. For a row $\mathbf{y}^* = (u, v)$ the product is $(0, ub + v(d-a))$, so $\mathbf{y}^* = (d-a, -b)$ works. Then $\mathbf{y}^*\mathbf{x} = d-a$, and

$$\kappa_{\mathbf{A}}(a) = \frac{1 \cdot \sqrt{|d-a|^2 + |b|^2}}{|d-a|} = \sqrt{1 + \frac{|b|^2}{|a-d|^2}}.$$

For d : $\mathbf{y} = \mathbf{e}_2$ is a left eigenvector, because $\mathbf{e}_2^* \mathbf{A} = (0, d) = d\mathbf{e}_2^*$. A right eigenvector is $\mathbf{x} = (b, d-a)$, since $(\mathbf{A} - d\mathbf{I})\mathbf{x} = ((a-d)b + b(d-a), 0) = \mathbf{0}$. Then $\mathbf{y}^*\mathbf{x} = d-a$, and $\kappa_{\mathbf{A}}(d)$ is given by the same formula. The two condition numbers are equal. They are large exactly when the off-diagonal entry is large compared with the gap between the eigenvalues.

- (b) Here $\kappa_{\mathbf{A}}(1) = \kappa_{\mathbf{A}}(2) = \sqrt{1 + M^2}$. §04 computed κ_2 for two diagonalizing matrices of

this $\mathbf{A}$ (Example (A large corner, separated eigenvalues)), and the better one, with unit columns, gave $M + \sqrt{1 + M^2}$. We show that no choice does better. The eigenvectors are the multiples of $\mathbf{e}_1$ and of $(M, 1)$, and the columns of a diagonalizing $\mathbf{X}$ are eigenvectors (Theorem 9.4.1). So, up to the order of its columns, which does not change $\kappa_2(\mathbf{X})$, every diagonalizing $\mathbf{X}$ has the form $\mathbf{X} = \begin{pmatrix} \alpha & \beta M \\ 0 & \beta \end{pmatrix}$ with $\alpha, \beta \neq 0$, by Lemma 20.11.1 (a). Multiplying $\mathbf{X}$ by a scalar does not change $\kappa_2(\mathbf{X})$, so we may take $\alpha = 1$. Let $\sigma_1 \geq \sigma_2 > 0$ be the singular values of $\mathbf{X}$. They are the square roots of the eigenvalues of $\mathbf{X}^* \mathbf{X}$ (Definition 13.9.1), so $\sigma_1^2 \sigma_2^2 = \det(\mathbf{X}^* \mathbf{X}) = |\det \mathbf{X}|^2$, giving $\sigma_1 \sigma_2 = |\beta|$, and $\sigma_1^2 + \sigma_2^2 = \text{tr}(\mathbf{X}^* \mathbf{X}) = 1 + |\beta|^2(1 + M^2)$. With $k = \kappa_2(\mathbf{X}) = \sigma_1/\sigma_2$,

$$k + \frac{1}{k} = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1 \sigma_2} = \frac{1}{|\beta|} + |\beta|(1 + M^2) \geq 2\sqrt{1 + M^2},$$

by the inequality $u + v \geq 2\sqrt{uv}$ for $u, v > 0$, with equality exactly when $|\beta| = 1/\sqrt{1 + M^2}$, that is, when both columns of $\mathbf{X}$ are unit vectors. The function $k \mapsto k + 1/k$ is increasing for $k \geq 1$, so the smallest possible $\kappa_2(\mathbf{X})$ is the root $k \geq 1$ of $k + 1/k = 2\sqrt{1 + M^2}$, namely

$$\min_{\mathbf{X}} \kappa_2(\mathbf{X}) = M + \sqrt{1 + M^2}.$$

Indeed $(M + \sqrt{1 + M^2})^{-1} = \sqrt{1 + M^2} - M$, and the two add up to $2\sqrt{1 + M^2}$. So the per-eigenvalue constant $\sqrt{1 + M^2}$ is smaller than the best Bauer–Fike constant, by a factor that tends to 2 as M grows, in agreement with Proposition 20.11.3 (d).

(c) With $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y}^* = (1, -M)$ from (a), $\mathbf{y}^* \mathbf{E}_{21} \mathbf{x} = \mathbf{y}^* \mathbf{e}_2 = -M$ and $\mathbf{y}^* \mathbf{x} = 1$, so

$$\lambda'(0) = -M, \quad |\lambda'(0)| = M \leq \sqrt{1 + M^2} \|\mathbf{E}_{21}\|_2 = \kappa_{\mathbf{A}}(1) \|\mathbf{E}\|_2.$$

Exactly: $\mathbf{A} + t\mathbf{E} = \begin{pmatrix} 1 & M \\ t & 2 \end{pmatrix}$ has trace 3 and determinant $2 - Mt$, so its eigenvalues are

$$\lambda_{\pm}(t) = \frac{3 \pm \sqrt{1 + 4Mt}}{2},$$

and the one near 1 is $\lambda_-(t)$. For $M = 100$ and $t = 10^{-4}$ we get $4Mt = 0.04$, and $\lambda_-(t) = (3 - \sqrt{1.04})/2 \approx 0.990098$. The eigenvalue has moved by about 0.0099, and the first-order prediction $|\lambda'(0)| = M$ is 0.01. Differentiating $\lambda_-(t)$ directly gives $\lambda'_-(0) = -\frac{1}{4} \cdot 4M = -M$, as the theorem says.

§04 set this matrix beside Chapter 12's Example (A perturbation that moves the eigenvalues a thousand times as far), which perturbed the same corner of $\begin{pmatrix} 1 & M \\ 0 & 1 \end{pmatrix}$. There the diagonal entries are equal, and the eigenvalue 1 moved by $\sqrt{Mt}$; with $M = 100$ and $t = 10^{-4}$ that is 0.1, ten times the movement found in (c). The derivative explains the difference. With a gap of 1 between the eigenvalues, the movement is linear in t to first order, with slope M , at most $\kappa_{\mathbf{A}}(1) = \sqrt{1 + M^2}$. With no gap there is no derivative at all. Chapter 12's closing remark, that a matrix with well-separated eigenvalues and a large defect behaves less dramatically, is now a formula.

20.11.5 When eigenvalues coalesce

The two computations just compared are the ends of a single family. Put

$$\mathbf{A}_s = \begin{pmatrix} 1 & M \\ 0 & 1 + s \end{pmatrix} \quad (M > 0, s > 0).$$

By [Example \(The condition numbers of a triangular \$2 \times 2\$ matrix\)](#) (a), with $b = M$ and $a - d = -s$,

$$\kappa_{\mathbf{A}_s}(1) = \kappa_{\mathbf{A}_s}(1 + s) = \sqrt{1 + \frac{M^2}{s^2}} \rightarrow \infty \quad (s \rightarrow 0^+).$$

As the two eigenvalues move together, the left eigenvector $(s, -M)$ of the eigenvalue 1 turns towards $\mathbf{e}_2$, perpendicular to the right eigenvector $\mathbf{e}_1$. At $s = 0$ the matrix is a Jordan block in disguise, as Chapter 12 observed, the product $\mathbf{y}^* \mathbf{x}$ is 0, and the derivative no longer exists. Instead the eigenvalue moves by $\sqrt{Mt}$ ([Example \(A perturbation that moves the eigenvalues a thousand times as far\)](#)), the case $n = 2$ of the n -th root behavior of §03 ([Example \(The Jordan block meets Elsner’s bound\)](#)).

So the first-order theory and the Hölder behavior of §03 are not two separate phenomena. In this family a large condition number is the sign of a nearby coalescence. Near such a coalescence, the linear regime $|\lambda(t) - \lambda| \approx \kappa|t|\|\mathbf{E}\|_2$ holds only for smaller and smaller t , and beyond it the square root takes over. The warning above showed the same thing from the other side: for $\begin{pmatrix} 1 & M \\ 0 & 2 \end{pmatrix}$, a perturbation of size $1/(4M)$, about $1/(4\kappa)$, is enough to make the two eigenvalues collide.

Remark. This completes the thread that Chapter 10 §03 started when it warned that the Jordan form is not stable and said that “Chapter 20 returns to this point”. §03 measured the instability, a movement of $\varepsilon^{1/n}$ for a block of size n , and showed that the exponent is the worst possible ([Example \(The Jordan block meets Elsner’s bound\)](#), Theorem 20.3.3). §04 bounded it for every matrix, with an m -th root for a largest block of size m (Theorem 20.4.5). This section adds the first-order view. A **defective** eigenvalue, one with a Jordan block of size $m \geq 2$, has no finite condition number in any sense: along $\mathbf{J}_m(\lambda) + t\mathbf{E}_{m1}$ every eigenvalue satisfies $|\lambda(t) - \lambda|/|t| = |t|^{1/m-1}$, which is unbounded as $t \rightarrow 0$. Any defective eigenvalue inherits this by similarity: if $\mathbf{A} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$ with $\mathbf{J}$ in Jordan form (Theorem 10.3.1) and $\mathbf{E} = \mathbf{S}\mathbf{E}'\mathbf{S}^{-1}$, where $\mathbf{E}'$ carries a single 1 in the lower-left corner of a block $\mathbf{J}_m(\lambda)$ and zeros elsewhere, then $\mathbf{A} + t\mathbf{E}$ is similar to $\mathbf{J} + t\mathbf{E}'$, so among its eigenvalues are the m roots of $(z - \lambda)^m = t$, at distance $|t|^{1/m}$ from λ . And, as the family $\mathbf{A}_s$ shows, simple eigenvalues that are about to merge into such a block have condition numbers that blow up. A multiple eigenvalue that is **not** defective is different. For $\mathbf{A} = \mathbf{I}_2$, every eigenvalue of $\mathbf{I}_2 + t\mathbf{E}$ lies within $|t|\|\mathbf{E}\|_2$ of 1 by Corollary 20.4.3 (a), so nothing is infinite. What fails there is only the numbering: the two eigenvalues of $\mathbf{I}_2 + t\mathbf{E}$ cannot in general be told apart as “the one that came from 1”, so [Definition 20.11.1](#) does not apply.

20.11.6 Summary and transfer

The chapter set out to answer four questions about a perturbed matrix $\mathbf{A} + \mathbf{E}$: where its eigenvalues are, how far they move, which one goes where, and what happens to eigenvectors, invariant subspaces, singular values, and polar factors.

- **Where.** Gershgorin’s discs (§01) and the other inclusion regions (§02) locate the eigenvalues without computing any of them.
- **How far.** The continuity of roots (§03) is now a theorem of the book, proved by compactness. It comes with Elsner’s Hölder bound, whose exponent $1/n$ cannot be improved, and with a matching version of that bound. Bauer–Fike (§04) replaces the exponent by 1 for diagonalizable matrices, at the price of the factor $\kappa_2(\mathbf{X})$.
- **Which goes where.** Hoffman–Wielandt (§05) matches the eigenvalues of two normal matrices in the Frobenius norm.
- **Hermitian matrices** (§06) got sharper tools: residual bounds, a quadratic bound for the

Rayleigh quotient, and relative bounds under congruence.

- **Eigenvectors and subspaces.** The Sylvester equation (§07) and principal angles (§08) together give Davis–Kahan (§09): an invariant subspace moves by at most the perturbation divided by the gap.
- **Singular values and polar factors** (§10) follow by the dilation and by the Sylvester bounds.
- **To first order.** This section gave each simple eigenvalue its own derivative and condition number.

Nine moves did the work.

- **Dominant coordinate plus the triangle inequality.** Look at the row of the eigen-equation where the eigenvector is largest (§01), or at the two largest coordinates (§02). *Transfer:* to locate an eigenvalue without computing it, read the eigen-equation at the coordinate where the eigenvector peaks. There the eigenvalue cannot hide behind the other terms.
- **A determinant bounded two ways.** The product of the distances from μ to the eigenvalues of $\mathbf{A}$ is $|\det(\mathbf{A} - \mu\mathbf{I})|$, and Hadamard’s inequality bounds that determinant by column lengths (§03, Elsner). *Transfer:* when a product of unknown quantities is a determinant, bound the determinant from above and the product from below, and read off the smallest factor.
- **Compactness plus uniqueness.** Roots are bounded, so a subsequence of root vectors converges. Its limit is a root vector of the limit polynomial, and unique factorization says which one (§03). *Transfer:* to prove that a solution depends continuously on the data, bound the solutions, extract a convergent subsequence, and let uniqueness of the limiting solution do the rest.
- **An integer-valued continuous function is constant.** Deform $\mathbf{A}$ to its diagonal, and count the eigenvalues in a region whose boundary no eigenvalue crosses (§03, the counting theorem). *Transfer:* to count objects that move continuously, move to a case where the count is obvious, along a path that never lets one of them escape.
- **Factor out the resolvent.** Write $\mathbf{A} + \mathbf{E} - \mu\mathbf{I} = (\mathbf{A} - \mu\mathbf{I})(\mathbf{I} + (\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{E})$, so that the Neumann series makes the second factor invertible whenever $\|(\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{E}\| < 1$ (§04). *Transfer:* to show a perturbed matrix is invertible, factor out the unperturbed one and ask whether what remains is close to $\mathbf{I}$.
- **Birkhoff turns a doubly stochastic weight into a permutation.** The weights $|w_{ij}|^2$ of a unitary matrix form a doubly stochastic matrix, and a linear function on those attains its minimum at a permutation matrix (§05). *Transfer:* to optimize over matchings, relax to doubly stochastic matrices. A linear objective has its optimum at a vertex, and the vertices are the matchings.
- **The perturbation of a subspace is a Sylvester equation.** The block that couples the old invariant subspace to the new complement satisfies $\tilde{\mathbf{A}}_2\mathbf{X} - \mathbf{X}\mathbf{A}_1 = (\text{a piece of } \mathbf{E})$, so the separation of the two spectra is the conditioning (§07, §09). *Transfer:* to measure how far an invariant subspace moves, write down the coupling block and recognize the equation it solves.
- **The singular-pair trick.** Test a matrix equation against the top singular pair of the unknown $\mathbf{X}$. This turns $\|\mathbf{X}\|_2$ into a scalar that the hypotheses bound directly (§07, used twice in §10). *Transfer:* to bound the spectral norm of the solution of a linear matrix equation, evaluate the equation at the vectors where that norm is attained.
- **The adjugate as a polynomial eigenvector.** A column of $\text{adj}(\mu\mathbf{I} - \mathbf{B})$ is an eigenvector of $\mathbf{B}$ whenever μ is an eigenvalue, and its entries are polynomials in μ and the entries of $\mathbf{B}$. So it moves continuously when they do (this section). *Transfer:* when an object must be chosen continuously and there is no canonical choice, look for a formula that is polynomial

in the data. The implicit function theorem then becomes unnecessary.

The chapter also paid promises made in earlier chapters.

The promise	Made in	Paid by
“This chapter quotes this theorem, and Chapter 20 proves it” (continuous dependence of roots on coefficients)	Chapter 16 §07	§03 (Continuity of roots)
“Chapter 20 returns to this point” (the Jordan form is not stable)	Chapter 10 §03	§03, §04, and this section
“Chapter 20 develops both and proves it” (Bauer–Fike)	Chapter 12 §11	§04 (The Bauer–Fike theorem)
“Chapter 20 proves the theorem of Bauer and Fike”	Chapter 16 §07	§04
“the theorem of Bauer and Fike, in Chapter 20”	Chapter 17 §03	§04
“the bounds that do exist, Weyl’s and Bauer–Fike’s, are Chapters 17 and 20”	Chapter 16 §08	§04
“Bounds for the eigenvalues of non-Hermitian matrices under perturbation, beginning with Bauer–Fike, are Chapter 20’s”	Chapter 17 §11	§§03–05
“its quantitative form appears in Chapter 20” (the Sylvester equation)	Chapter 8 §04	§07 (The Sylvester equation)
“belongs with perturbation theory in Chapter 20” (conditioning of the Sylvester equation)	Chapter 12 §10	§07
“Chapter 20 leans on it for perturbation bounds” (the Hermitian dilation)	Chapter 17 §09	§10 (Singular values and polar factors)

Some things the chapter did not do. The matching form of Elsner’s bound in §03 carries a factor $2n - 1$, which is not the best possible, and the better constants were not pursued. The spectral-norm form of Davis–Kahan for a cluster lying between two parts of the rest of the spectrum was stated without proof (§09), and nothing depends on it. The bounds of §05, and Mirsky’s theorem in §10, hold in the Frobenius norm; §10’s square-root and polar bounds hold in the spectral norm as well. Extending the eigenvalue and singular value bounds to every unitarily invariant norm belongs to Chapter 21; the square-root and polar bounds are not extended there, and stay in the two norms proved here. Finally, the residual bounds of §04 and §06 say that a computed eigenpair is an exact eigenpair of a nearby matrix. Turning that into a theory of backward error, and into algorithms that compute eigenvalues, is the subject of Chapter 24.

20.11.7 Exercises

20.11.7.1 A. Check your understanding

Exercise. A1

- (a) Define the condition number $\kappa_{\mathbf{A}}(\lambda)$, saying for which λ it is defined.
- (b) State the formula for the derivative of a simple eigenvalue of $\mathbf{A} + t\mathbf{E}$ at $t = 0$.
- (c) True or false: for every eigenvalue of every $\mathbf{A} \in M_n(\mathbb{C})$, a right eigenvector $\mathbf{x}$ and a left eigenvector $\mathbf{y}$ satisfy $\mathbf{y}^*\mathbf{x} \neq 0$. Justify your answer.
- (d) True or false: if $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$ with $\kappa_2(\mathbf{X}) \geq 10^6$, then some eigenvalue of $\mathbf{A}$ has $\kappa_{\mathbf{A}}(\lambda) \geq 10^6$. Justify your answer.
- (e) True or false: $\kappa_{2\mathbf{A}+3\mathbf{I}}(2\lambda+3) = \kappa_{\mathbf{A}}(\lambda)$ for every simple eigenvalue λ of $\mathbf{A}$. Justify your answer.
- (f) The proof of [Theorem 20.11.2](#) needed eigenvectors of $\mathbf{A} + t\mathbf{E}$ that converge as $t \rightarrow 0$. Where did they come from?

Solution.

- (a) For a **simple** eigenvalue λ of $\mathbf{A} \in M_n(\mathbb{C})$ with right eigenvector $\mathbf{x}$ and left eigenvector $\mathbf{y}$ ($\mathbf{y}^*\mathbf{A} = \lambda\mathbf{y}^*$), $\kappa_{\mathbf{A}}(\lambda) = \|\mathbf{x}\|\|\mathbf{y}\|/|\mathbf{y}^*\mathbf{x}|$ ([Definition 20.11.1](#)).
- (b) $\lambda'(0) = \mathbf{y}^*\mathbf{E}\mathbf{x}/(\mathbf{y}^*\mathbf{x})$, where $\lambda(t)$ is the unique eigenvalue of $\mathbf{A} + t\mathbf{E}$ near λ ([Theorem 20.11.2](#)).
- (c) False. For $\mathbf{J}_2(0)$ the eigenvectors are the multiples of $\mathbf{x} = \mathbf{e}_1$ and of $\mathbf{y} = \mathbf{e}_2$, and $\mathbf{y}^*\mathbf{x} = 0$, as in the warning after [Lemma 20.11.1](#). The eigenvalue 0 is not simple, and simplicity is the hypothesis of that lemma.
- (d) False. Let $\mathbf{A} = \text{diag}(1, 2)$ and $\mathbf{X} = \text{diag}(1, 10^6)$. Then $\mathbf{X}\mathbf{A}\mathbf{X}^{-1} = \mathbf{A}$ because diagonal matrices commute, so $\mathbf{A} = \mathbf{X}\mathbf{A}\mathbf{X}^{-1}$, and $\kappa_2(\mathbf{X}) = 10^6$. But $\mathbf{A}$ is normal, so both eigenvalues have condition number 1 by [Proposition 20.11.3](#) (c). A large $\kappa_2(\mathbf{X})$ may just reflect a poor choice of $\mathbf{X}$.
- (e) True. The eigenvalue $2\lambda + 3$ of $2\mathbf{A} + 3\mathbf{I}$ is simple, since $p_{2\mathbf{A}+3\mathbf{I}}(x) = 2^n p_{\mathbf{A}}((x-3)/2)$. From $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{y}^*\mathbf{A} = \lambda\mathbf{y}^*$ we get $(2\mathbf{A} + 3\mathbf{I})\mathbf{x} = (2\lambda + 3)\mathbf{x}$ and $\mathbf{y}^*(2\mathbf{A} + 3\mathbf{I}) = (2\lambda + 3)\mathbf{y}^*$. So the same $\mathbf{x}$ and $\mathbf{y}$ serve, and the formula gives the same number.
- (f) From the adjugate: a nonzero column of $\text{adj}(\lambda\mathbf{I} - \mathbf{A})$ was fixed, and the same column of $\text{adj}(\lambda(t)\mathbf{I} - \mathbf{A} - t\mathbf{E})$ is an eigenvector of $\mathbf{A} + t\mathbf{E}$ for $\lambda(t)$ by [Theorem 7.6.2](#). Its entries are polynomials in $\lambda(t)$, t , and the entries of $\mathbf{A}$ and $\mathbf{E}$, so it converges because $\lambda(t)$ does ([Corollary 16.7.4](#)). No implicit function theorem was used.

20.11.7.2 B. Practice

Exercise. B1: Three condition numbers

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 5 & 4 \\ 0 & 0 & 2 \end{pmatrix}.$$

- (a) For each eigenvalue of $\mathbf{A}$, find a right and a left eigenvector with integer entries, and

compute $\kappa_{\mathbf{A}}(1)$, $\kappa_{\mathbf{A}}(2)$, and $\kappa_{\mathbf{A}}(5)$.

- (b) Hence decide which eigenvalue is the most sensitive to first order, and find $\mathbf{E}$ with $\|\mathbf{E}\|_2 = 1$ for which that eigenvalue of $\mathbf{A} + t\mathbf{E}$ has $|\lambda'(0)|$ as large as possible.

Solution.

- (a) The matrix is upper triangular with diagonal 1, 5, 2, so these are the eigenvalues, each simple.

Eigenvalue 1. $\mathbf{x} = \mathbf{e}_1$. For $\mathbf{y}^* = (u, v, w)$, the equation $\mathbf{y}^*(\mathbf{A} - \mathbf{I}) = \mathbf{0}$ reads $(0, 3u + 4v, 4v + w) = \mathbf{0}$. Take $u = 4, v = -3, w = 12$, so $\mathbf{y} = (4, -3, 12)$, with $\|\mathbf{y}\| = \sqrt{16 + 9 + 144} = 13$ and $\mathbf{y}^*\mathbf{x} = 4$. Hence $\kappa_{\mathbf{A}}(1) = 13/4$.

Eigenvalue 2. $\mathbf{y} = \mathbf{e}_3$, since $\mathbf{e}_3^*\mathbf{A} = (0, 0, 2)$. For $\mathbf{x} = (p, q, r)$, the equation $(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \mathbf{0}$ reads $-p + 3q = 0, 3q + 4r = 0$. Take $r = 3, q = -4, p = -12$, so $\mathbf{x} = (-12, -4, 3)$, with $\|\mathbf{x}\| = \sqrt{144 + 16 + 9} = 13$ and $\mathbf{y}^*\mathbf{x} = 3$. Hence $\kappa_{\mathbf{A}}(2) = 13/3$.

Eigenvalue 5. For $\mathbf{x} = (p, q, 0)$, $(\mathbf{A} - 5\mathbf{I})\mathbf{x} = (-4p + 3q, 0, 0)$, so $\mathbf{x} = (3, 4, 0)$. For $\mathbf{y}^* = (0, v, w)$, $\mathbf{y}^*(\mathbf{A} - 5\mathbf{I}) = (0, 0, 4v - 3w)$, so $\mathbf{y} = (0, 3, 4)$. Both have length 5, and $\mathbf{y}^*\mathbf{x} = 12$. Hence $\kappa_{\mathbf{A}}(5) = 25/12$.

- (b) Since $13/3 > 13/4 > 25/12$, the eigenvalue 2 is the most sensitive, although the eigenvalue 1 is exactly as close to its nearest neighbor and less sensitive: the gaps alone do not decide. By Proposition 20.11.3 (b), the worst direction is

$$\mathbf{E} = \frac{\mathbf{y}\mathbf{x}^*}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{1}{13} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -12 & -4 & 3 \end{pmatrix},$$

with $\|\mathbf{E}\|_2 = 1$ and $\lambda'(0) = \|\mathbf{x}\|\|\mathbf{y}\|/(\mathbf{y}^*\mathbf{x}) = 13/3$. Hence, to first order, a perturbation of size 10^{-3} in this direction moves the eigenvalue 2 by about 4.3×10^{-3} .

Exercise. B2: A derivative checked against the exact answer

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}$ and $\mathbf{E} = \mathbf{E}_{12}$.

- (a) Find the eigenvalues of $\mathbf{A}$ with right and left eigenvectors, and compute both condition numbers.
 (b) Use Theorem 20.11.2 to find $\lambda'(0)$ for each eigenvalue of $\mathbf{A} + t\mathbf{E}$.
 (c) Find the eigenvalues of $\mathbf{A} + t\mathbf{E}$ exactly. Hence compare the first-order estimate with the exact value at $t = 10^{-3}$.

Solution.

- (a) $p_{\mathbf{A}}(x) = x^2 - 3x + 2 = (x - 1)(x - 2)$. For 1: $\mathbf{A} - \mathbf{I} = \begin{pmatrix} -1 & 1 \\ -2 & 2 \end{pmatrix}$, so $\mathbf{x}_1 = (1, 1)$, and $\mathbf{y}^*(\mathbf{A} - \mathbf{I}) = \mathbf{0}$ for $\mathbf{y}^* = (u, v)$ means $-u - 2v = 0$, so $\mathbf{y}_1 = (2, -1)$. Then $\mathbf{y}_1^*\mathbf{x}_1 = 1$ and $\kappa_{\mathbf{A}}(1) = \sqrt{2}\sqrt{5} = \sqrt{10}$. For 2: $\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} -2 & 1 \\ -2 & 1 \end{pmatrix}$, so $\mathbf{x}_2 = (1, 2)$ and $\mathbf{y}_2 = (1, -1)$, with $\mathbf{y}_2^*\mathbf{x}_2 = -1$ and $\kappa_{\mathbf{A}}(2) = \sqrt{5}\sqrt{2} = \sqrt{10}$.

- (b) $\mathbf{y}^*\mathbf{E}_{12}\mathbf{x} = \overline{y_1}x_2$. For 1: $2 \cdot 1/1 = 2$. For 2: $1 \cdot 2/(-1) = -2$.

- (c) $\mathbf{A} + t\mathbf{E} = \begin{pmatrix} 0 & 1+t \\ -2 & 3 \end{pmatrix}$ has trace 3 and determinant $2 + 2t$, so its eigenvalues are $(3 \mp \sqrt{1 - 8t})/2$, for $t < 1/8$. Differentiating at $t = 0$ gives ± 2 , as in (b). At $t = 10^{-3}$, the

first-order estimates are 1.002 and 1.998, and the exact values are $(3 \mp \sqrt{0.992})/2 \approx 1.002004$ and 1.997996 . Hence the error of the first-order estimate is about 4×10^{-6} , which is of the order of t^2 , as expected of a first-order approximation.

Exercise. B3: Invariance of the condition number

Let λ be a simple eigenvalue of $\mathbf{A} \in M_n(\mathbb{C})$. Prove that:

- (a) $\bar{\lambda}$ is a simple eigenvalue of $\mathbf{A}^*$, and $\kappa_{\mathbf{A}^*}(\bar{\lambda}) = \kappa_{\mathbf{A}}(\lambda)$;
- (b) $\kappa_{\mathbf{U}^* \mathbf{A} \mathbf{U}}(\lambda) = \kappa_{\mathbf{A}}(\lambda)$ for every unitary $\mathbf{U}$.
- (c) Explain why (b) fails for a general invertible $\mathbf{S}$ in place of $\mathbf{U}$, using $\mathbf{A} = \text{diag}(1, 2)$.

Solution.

- (a) The characteristic polynomial of $\mathbf{A}^*$ has as roots the conjugates of those of $\mathbf{A}$, with the same multiplicities: $p_{\mathbf{A}^*}(x) = \det(x\mathbf{I} - \mathbf{A}^*) = \overline{\det(\bar{x}\mathbf{I} - \mathbf{A})}$, because the determinant of the conjugate transpose is the conjugate of the determinant. So $\bar{\lambda}$ is a simple eigenvalue of $\mathbf{A}^*$. If $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{y}^*\mathbf{A} = \lambda\mathbf{y}^*$, then taking conjugate transposes gives $\mathbf{x}^*\mathbf{A}^* = \bar{\lambda}\mathbf{x}^*$ and $\mathbf{A}^*\mathbf{y} = \bar{\lambda}\mathbf{y}$. So for $\mathbf{A}^*$ and $\bar{\lambda}$, the vector $\mathbf{y}$ is a right eigenvector and $\mathbf{x}$ a left one. Hence $\kappa_{\mathbf{A}^*}(\bar{\lambda}) = \|\mathbf{y}\|\|\mathbf{x}\|/|\mathbf{x}^*\mathbf{y}| = \kappa_{\mathbf{A}}(\lambda)$, since $|\mathbf{x}^*\mathbf{y}| = |\overline{\mathbf{y}^*\mathbf{x}}| = |\mathbf{y}^*\mathbf{x}|$.
- (b) $\mathbf{U}^*\mathbf{A}\mathbf{U}$ has the same characteristic polynomial as $\mathbf{A}$ (Theorem 7.10.3), so λ is simple for it. From $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{y}^*\mathbf{A} = \lambda\mathbf{y}^*$ we get $(\mathbf{U}^*\mathbf{A}\mathbf{U})(\mathbf{U}^*\mathbf{x}) = \lambda\mathbf{U}^*\mathbf{x}$ and $(\mathbf{U}^*\mathbf{y})^*(\mathbf{U}^*\mathbf{A}\mathbf{U}) = \mathbf{y}^*\mathbf{A}\mathbf{U} = \lambda(\mathbf{U}^*\mathbf{y})^*$. Since $\mathbf{U}^*$ is unitary, $\|\mathbf{U}^*\mathbf{x}\| = \|\mathbf{x}\|$ and $\|\mathbf{U}^*\mathbf{y}\| = \|\mathbf{y}\|$ (Theorem 11.7.3), and $(\mathbf{U}^*\mathbf{y})^*(\mathbf{U}^*\mathbf{x}) = \mathbf{y}^*\mathbf{U}\mathbf{U}^*\mathbf{x} = \mathbf{y}^*\mathbf{x}$. So all three ingredients of Definition 20.11.1 are unchanged.
- (c) For $\mathbf{S} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$. By Example (The condition numbers of a triangular 2×2 matrix) (a) with $b = -1$ and $|a - d| = 1$, its eigenvalue 1 has condition number $\sqrt{2}$, whereas $\kappa_{\mathbf{A}}(1) = 1$ because $\mathbf{A}$ is normal. A nonunitary similarity changes lengths and angles, so it changes the angle between $\mathbf{x}$ and $\mathbf{y}$ in Proposition 20.11.3 (e).

20.11.7.3 C. Going deeper

Exercise. C1: Perfect conditioning characterizes normality

- (a) Let $\mathbf{A} \in M_n(\mathbb{C})$ have n distinct eigenvalues. Prove that $\mathbf{A}$ is normal if and only if $\kappa_{\mathbf{A}}(\lambda) = 1$ for every eigenvalue λ .
- (b) Show that the hypothesis of distinct eigenvalues cannot be replaced by “ $\kappa_{\mathbf{A}}(\lambda) = 1$ for every simple eigenvalue λ ”: give a nonnormal $\mathbf{A} \in M_3(\mathbb{C})$ with a simple eigenvalue, all of whose simple eigenvalues have condition number 1.

Hint: in (a), what does $\kappa = 1$ say about $\mathbf{y}$, and what does the computation in Theorem 9.10.6 say about $\mathbf{y}_j^* \mathbf{x}_i$?

Solution.

- (a) ($\Rightarrow$) This is Proposition 20.11.3 (c).
- ($\Leftarrow$) Let $\lambda_1, \dots, \lambda_n$ be the distinct eigenvalues, with right eigenvectors $\mathbf{x}_1, \dots, \mathbf{x}_n$. Since

$\kappa_{\mathbf{A}}(\lambda_j) = 1$, [Proposition 20.11.3](#) (a) says that a left eigenvector for λ_j is a multiple of $\mathbf{x}_j$, so $\mathbf{x}_j$ itself is one: $\mathbf{x}_j^* \mathbf{A} = \lambda_j \mathbf{x}_j^*$. Let $i \neq j$. Computing $\mathbf{x}_j^* \mathbf{A} \mathbf{x}_i$ in two ways,

$$\lambda_j \mathbf{x}_j^* \mathbf{x}_i = (\mathbf{x}_j^* \mathbf{A}) \mathbf{x}_i = \mathbf{x}_j^* (\mathbf{A} \mathbf{x}_i) = \lambda_i \mathbf{x}_j^* \mathbf{x}_i,$$

so $(\lambda_j - \lambda_i) \mathbf{x}_j^* \mathbf{x}_i = 0$, and $\mathbf{x}_j^* \mathbf{x}_i = 0$ because $\lambda_i \neq \lambda_j$. So the vectors $\mathbf{x}_i / \|\mathbf{x}_i\|$ form an orthonormal list of n vectors in $\mathbb{C}^n$, hence an orthonormal basis of eigenvectors. With $\mathbf{U}$ the unitary matrix having these columns, $\mathbf{A} = \mathbf{U} \operatorname{diag}(\lambda_1, \dots, \lambda_n) \mathbf{U}^*$, so $\mathbf{A}$ is normal by Corollary 12.4.3.

- (b) Let $\mathbf{A} = \mathbf{J}_2(0) \oplus (1) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Its eigenvalues are 0, 0, 1, so 1 is the only simple eigenvalue. Its right and left eigenvectors are both $\mathbf{e}_3$, since $\mathbf{A} \mathbf{e}_3 = \mathbf{e}_3$ and $\mathbf{e}_3^* \mathbf{A} = \mathbf{e}_3^*$, so $\kappa_{\mathbf{A}}(1) = 1$. But $\mathbf{A}$ is not normal: $\mathbf{A}^* \mathbf{A} = \operatorname{diag}(0, 1, 1)$ while $\mathbf{A} \mathbf{A}^* = \operatorname{diag}(1, 0, 1)$.

Exercise. C2: The derivatives add up to the trace

Let $\mathbf{A} \in M_n(\mathbb{C})$ have n distinct eigenvalues $\lambda_1, \dots, \lambda_n$, and let $\mathbf{E} \in M_n(\mathbb{C})$. Write $\lambda_i(t)$ for the eigenvalue of $\mathbf{A} + t\mathbf{E}$ near λ_i , from [Theorem 20.11.2](#).

- (a) Prove that $\lambda_1'(0) + \dots + \lambda_n'(0) = \operatorname{tr} \mathbf{E}$.
 (b) Deduce this again from $\operatorname{tr}(\mathbf{A} + t\mathbf{E}) = \lambda_1(t) + \dots + \lambda_n(t)$, valid for all small t , and explain why the two arguments agree.

Hint: in (a), use a diagonalizing $\mathbf{X}$ and the rows of $\mathbf{X}^{-1}$ as left eigenvectors.

Solution.

- (a) $\mathbf{A}$ is diagonalizable by Corollary 9.4.2, say $\mathbf{A} = \mathbf{X} \mathbf{\Lambda} \mathbf{X}^{-1}$. As in the proof of [Proposition 20.11.3](#) (d), $\mathbf{x}_i = \mathbf{X} \mathbf{e}_i$ and $\mathbf{y}_i^* = \mathbf{e}_i^* \mathbf{X}^{-1}$ are right and left eigenvectors for λ_i , with $\mathbf{y}_i^* \mathbf{x}_i = 1$. So by [Theorem 20.11.2](#),

$$\sum_{i=1}^n \lambda_i'(0) = \sum_{i=1}^n \mathbf{e}_i^* \mathbf{X}^{-1} \mathbf{E} \mathbf{X} \mathbf{e}_i = \operatorname{tr}(\mathbf{X}^{-1} \mathbf{E} \mathbf{X}) = \operatorname{tr} \mathbf{E},$$

where the middle equality sums the diagonal entries of $\mathbf{X}^{-1} \mathbf{E} \mathbf{X}$ and the last uses $\operatorname{tr}(\mathbf{BC}) = \operatorname{tr}(\mathbf{CB})$ (Theorem 0.9.6 (3)).

- (b) Let g be the smallest distance between two of the λ_i . By [Theorem 20.11.2](#) (b), $|\lambda_i(t) - \lambda_i| < g/2$ for every i once $|t|$ is small enough, so $\lambda_1(t), \dots, \lambda_n(t)$ are n distinct eigenvalues of $\mathbf{A} + t\mathbf{E}$. They are therefore all of its eigenvalues, each simple, and Theorem 9.3.3 gives $\operatorname{tr}(\mathbf{A} + t\mathbf{E}) = \sum_i \lambda_i(t)$. The left side is $\operatorname{tr} \mathbf{A} + t \operatorname{tr} \mathbf{E}$ and $\operatorname{tr} \mathbf{A} = \sum_i \lambda_i$, so $\sum_i (\lambda_i(t) - \lambda_i)/t = \operatorname{tr} \mathbf{E}$ for small $t \neq 0$. Letting $t \rightarrow 0$ and using (c) of the theorem for each i gives the claim. The two arguments agree because both express the same fact: the sum of the eigenvalues is a linear function of the matrix, so its derivative in the direction $\mathbf{E}$ is $\operatorname{tr} \mathbf{E}$, whatever the individual derivatives are.

Exercise. C3: The adjugate at a simple eigenvalue

Let $n \geq 2$, let λ be a simple eigenvalue of $\mathbf{A} \in M_n(\mathbb{C})$ with right eigenvector $\mathbf{x}$ and left eigenvector $\mathbf{y}$, and put $\mathbf{N} = \lambda \mathbf{I} - \mathbf{A}$.

- (a) Prove that $\text{adj } \mathbf{N} = c \mathbf{x} \mathbf{y}^*$ for some scalar $c \neq 0$.
 (b) Prove that $\text{tr}(\text{adj } \mathbf{N}) = p'_{\mathbf{A}}(\lambda)$, the derivative of the characteristic polynomial at λ .
 (c) Deduce a second proof that $\mathbf{y}^* \mathbf{x} \neq 0$, and the formula

$$\text{adj}(\lambda \mathbf{I} - \mathbf{A}) = p'_{\mathbf{A}}(\lambda) \frac{\mathbf{x} \mathbf{y}^*}{\mathbf{y}^* \mathbf{x}}.$$

Hint: for (a), use both products in Theorem 7.6.2; for (b), apply Theorem 16.6.6 to $t \mapsto (\lambda + t)\mathbf{I} - \mathbf{A}$.

Solution.

- (a) Since λ is an eigenvalue, $\det \mathbf{N} = 0$, so Theorem 7.6.2 gives $\mathbf{N} \text{adj } \mathbf{N} = \mathbf{O}$ and $\text{adj } \mathbf{N} \mathbf{N} = \mathbf{O}$. The first says every column of $\text{adj } \mathbf{N}$ lies in $\ker(\lambda \mathbf{I} - \mathbf{A})$, so by Lemma 20.11.1 (a) every column is a multiple of $\mathbf{x}$: $\text{adj } \mathbf{N} = \mathbf{x} \mathbf{w}^*$ for some $\mathbf{w} \in \mathbb{C}^n$, whose conjugated entries are those multiples. Step 3 of the proof of Theorem 20.11.2 showed that $\text{adj } \mathbf{N} \neq \mathbf{O}$, so $\mathbf{w} \neq \mathbf{0}$. The second identity gives $\mathbf{x}(\mathbf{w}^* \mathbf{N}) = \mathbf{O}$, and since $\mathbf{x} \neq \mathbf{0}$ this forces $\mathbf{w}^* \mathbf{N} = \mathbf{0}^*$, that is, $\mathbf{w}^* \mathbf{A} = \lambda \mathbf{w}^*$. So $\mathbf{w}$ is a left eigenvector, hence $\mathbf{w} = \bar{c} \mathbf{y}$ for some $c \neq 0$ by Lemma 20.11.1 (a), and $\text{adj } \mathbf{N} = c \mathbf{x} \mathbf{y}^*$.
- (b) Let $\mathbf{B}(t) = (\lambda + t)\mathbf{I} - \mathbf{A}$ for real t , so $\mathbf{B}'(0) = \mathbf{I}$ and $\mathbf{B}(0) = \mathbf{N}$. By Theorem 16.6.6, $t \mapsto \det \mathbf{B}(t)$ is differentiable at 0 with derivative $\text{tr}(\text{adj } \mathbf{N} \mathbf{I}) = \text{tr}(\text{adj } \mathbf{N})$. On the other hand $\det \mathbf{B}(t) = p_{\mathbf{A}}(\lambda + t)$. Expanding each power $(\lambda + t)^k$ by the binomial theorem gives $p_{\mathbf{A}}(\lambda + t) = p_{\mathbf{A}}(\lambda) + p'_{\mathbf{A}}(\lambda)t + t^2 h(t)$ for a polynomial h , because the coefficient of t in $\sum_k c_k(\lambda + t)^k$ is $\sum_k k c_k \lambda^{k-1} = p'_{\mathbf{A}}(\lambda)$. So the difference quotient at 0 is $p'_{\mathbf{A}}(\lambda) + t h(t) \rightarrow p'_{\mathbf{A}}(\lambda)$. The two derivatives agree: $\text{tr}(\text{adj } \mathbf{N}) = p'_{\mathbf{A}}(\lambda)$.
- (c) By (a) and (b), $p'_{\mathbf{A}}(\lambda) = \text{tr}(c \mathbf{x} \mathbf{y}^*) = c \mathbf{y}^* \mathbf{x}$, since the trace of $\mathbf{x} \mathbf{y}^*$ is $\sum_i x_i \bar{y}_i = \mathbf{y}^* \mathbf{x}$. Because λ is a simple root of $p_{\mathbf{A}}$, Theorem 6.4.5 (a) gives $p'_{\mathbf{A}}(\lambda) \neq 0$. Hence $\mathbf{y}^* \mathbf{x} \neq 0$, which reproves Lemma 20.11.1 (b) from (a) of that lemma. Moreover $c = p'_{\mathbf{A}}(\lambda)/(\mathbf{y}^* \mathbf{x})$, which is the displayed formula. For example, for $\mathbf{A} = \begin{pmatrix} 1 & M \\ 0 & 2 \end{pmatrix}$ and $\lambda = 1$, we have $\text{adj}(\mathbf{I} - \mathbf{A}) = \begin{pmatrix} -1 & M \\ 0 & 0 \end{pmatrix}$ and $p'_{\mathbf{A}}(1) = -1$, while $\mathbf{x} \mathbf{y}^*/(\mathbf{y}^* \mathbf{x}) = \begin{pmatrix} 1 & -M \\ 0 & 0 \end{pmatrix}$, in agreement.