5
Eigentheory

To each linear operator, we introduce a special collection of scalars that interpolate between the determinant and all of the entries of its associated matrix.

5.0 Eigenvalues

Which scalars best encode a linear operator? A surprising amount of information about a linear operator can be recovered by focusing on the simplest type of behaviour: when the linear operator acts as multiplication by a scalar.

5.0.0 Definition. Let V be a $\mathbb{K}$-vector space. A scalar λ in $\mathbb{K}$ is an ***eigenvalue*** of a linear operator $T \colon V \to V$ if there exists a nonzero vector $\boldsymbol{v}$ in V such that $T[\boldsymbol{v}] = \lambda\, \boldsymbol{v}$. Geometrically, the vectors $T[\boldsymbol{v}]$ and $\boldsymbol{v}$ are parallel. Any nonzero vector $\boldsymbol{v}$ in V is an ***eigenvector*** of linear operator T with eigenvalue λ if $T[\boldsymbol{v}] = \lambda\, \boldsymbol{v}$. The set of all eigenvalues for T is called the ***spectrum*** of the linear operator T.

The lowercase "lambda" λ is the eleventh letter in the Greek alphabet.

Since every linear operator sends the zero vector to the zero vector, we exclude this degenerate situation from the definition of an eigenvector.

5.0.1 Problem. Consider the linear operator $T \colon \mathbb{Q}^2 \to \mathbb{Q}^2$ defined by $T[\boldsymbol{x}] \coloneqq \mathbf{A}\, \boldsymbol{x}$ where the matrix $\mathbf{A}$ is $\left[\begin{smallmatrix} 1 & 6 \\ 5 & 2 \end{smallmatrix}\right]$. Show that the vector $\left[\begin{smallmatrix} 6 \\ -5 \end{smallmatrix}\right]$ is an eigenvector of T, but the vector $\left[\begin{smallmatrix} 3 \\ -2 \end{smallmatrix}\right]$ is not.

Solution. For any scalar λ in $\mathbb{Q}$, we have

$$\begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = (-4) \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

so the vector $\left[\begin{smallmatrix} 6 \\ -5 \end{smallmatrix}\right]$ is an eigenvector with eigenvalue -4 of the linear map T whereas the vector $\left[\begin{smallmatrix} 3 \\ -2 \end{smallmatrix}\right]$ is not an eigenvector of T. $\square$

5.0.2 Problem. Let $D \colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the linear operator defined, for any smooth function f, by $D[f] \coloneqq \frac{df}{dt}$. For any scalar a in $\mathbb{R}$, show that e^{at} an eigenvector of D.

Solution. Since $D[e^{at}] = a\, e^{at}$, the function e^{at} is an eigenvector for D with eigenvalue a, so every real number is an eigenvalue of D. $\square$

5.0.3 Problem. Let $J \colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the linear operator defined by $(J[f])(x) \coloneqq \int_0^x f(t)\; dt$. Prove that J has no eigenvalues.

Solution by contradiction. Suppose that the smooth function f is an eigenvector of the linear operator J. By definition, there would exist a scalar λ in $\mathbb{R}$ such that $(J[f])(x) = \int_0^x f(t)\, dt = \lambda\big(f(x)\big)$ and the Fundamental Theorem of Calculus would give

$$f(x) = \frac{d}{dx}\left[\int_0^x f(t)\, dt\right] = \frac{d}{dx}\left[\lambda\big(f(x)\big)\right] = \lambda\big(f'(x)\big)\,.$$

Since eigenvectors are nonzero, we would have $\lambda \neq 0$. The general solution to the differential equation $f'(x) = \frac{1}{\lambda} f(x)$ has the form $f(x) = Ce^{x/\lambda}$ for some real constant C. However, the initial condition $(J[f])(0) = 0 = f(0)$ would imply that $C = 0$, so f would the zero function which is a contradiction. □

5.0.4 Proposition (Characterization of eigenvalues)**.** *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{K}$-vector space V. For any scalar λ in $\mathbb{K}$, the following are equivalent.*

a. *The scalar λ is an eigenvalue of T.*
b. *The linear operator $\lambda \operatorname{id}_V - T$ is not invertible.*
c. *The linear operator $\lambda \operatorname{id}_V - T$ is not injective.*
d. *The linear operator $\lambda \operatorname{id}_V - T$ is not surjective.*

Proof. Suppose that the vector $\boldsymbol{v}$ in V is an eigenvector of T with eigenvalue λ in $\mathbb{K}$. The defining equation $T[\boldsymbol{v}] = \lambda\, \boldsymbol{v}$ is equivalent to $(\lambda \operatorname{id}_V - T)[\boldsymbol{v}] = \mathbf{0}$. Hence, the vector $\boldsymbol{v}$ is an eigenvector of the linear operator T with eigenvalue λ if and only if $\boldsymbol{v}$ lies in $\operatorname{Ker}(\lambda \operatorname{id}_V - T)$. An eigenvector is nonzero, so we have $\operatorname{Ker}(\lambda \operatorname{id}_V - T) \neq \{\mathbf{0}\}$. On a finite-dimensional vector space, the characterization of invertible operators [3.3.5] shows the equivalence of conditions b–d. □

5.0.5 Problem. Show that 7 is an eigenvalue of $\mathbf{A} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$.

Solution. It suffices to show that there exists a nonzero $\boldsymbol{v}$ in $\mathbb{Q}^2$ such that $(7\,\mathbf{I} - \mathbf{A})[\boldsymbol{v}] = \mathbf{0}$. Row operations give

$$7\,\mathbf{I} - \mathbf{A} = 7\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 6 & -6 \\ -5 & 5 \end{bmatrix} \xrightarrow[\sim]{r_1 \mapsto \frac{1}{6} r_1} \begin{bmatrix} 1 & -1 \\ -5 & 5 \end{bmatrix} \xrightarrow[\sim]{r_2 \mapsto r_2 + 5\, r_1} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

which implies that $\operatorname{Ker}(7\,\mathbf{I} - \mathbf{A}) = \operatorname{Span}\left(\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\right) \neq \{\mathbf{0}\}$ and 7 is an eigenvalue of $\mathbf{A}$. □

5.0.6 Theorem (Distinct eigenvalues have independent eigenvectors)**.** *The eigenvectors corresponding to distinct eigenvalues of a linear operator are linearly independent.*

Proof by contradiction. Let V be a $\mathbb{K}$-vector space and let m be a positive integer. Consider a linear operator $T\colon V \to V$ having m distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$. Suppose that the eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$ in V, where $T[\boldsymbol{v}_j] = \lambda_j\, \boldsymbol{v}_j$ for all $1 \leqslant j \leqslant m$, are linearly dependent. Hence, there would exist a smallest positive integer k such that the vector $\boldsymbol{v}_k$ lies in $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{k-1})$, whence there would exist scalars $c_1, c_2, \dots, c_{k-1}$ in $\mathbb{K}$, not all zero, such that $\boldsymbol{v}_k = c_1\,\boldsymbol{v}_1 + c_2\,\boldsymbol{v}_2 + \cdots + c_{k-1}\,\boldsymbol{v}_{k-1}$. It follows that

$$\begin{aligned} \lambda_k\,\boldsymbol{v}_k &= c_1\lambda_k\,\boldsymbol{v}_1 + c_2\lambda_k\,\boldsymbol{v}_2 + \cdots + c_{k-1}\lambda_k\,\boldsymbol{v}_{k-1}\,, \\ T[\boldsymbol{v}_k] &= T[c_1\,\boldsymbol{v}_1 + c_2\,\boldsymbol{v}_2 + \cdots + c_{k-1}\,\boldsymbol{v}_{k-1}] \\ &= c_1\lambda_1\,\boldsymbol{v}_1 + c_2\lambda_2\,\boldsymbol{v}_2 + \cdots + c_{k-1}\lambda_{k-1}\,\boldsymbol{v}_{k-1}\,, \end{aligned}$$

which would imply that

$$\begin{aligned} \mathbf{0} &= T[\boldsymbol{v}_k] - \lambda_k\,\boldsymbol{v}_k \\ &= c_1(\lambda_1 - \lambda_k)\,\boldsymbol{v}_1 + c_2(\lambda_2 - \lambda_k)\,\boldsymbol{v}_2 + \cdots + c_{k-1}(\lambda_{k-1} - \lambda_k)\,\boldsymbol{v}_{k-1}\,. \end{aligned}$$

Our minimal choice of k would ensure that the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{k-1}$ are linearly independent. Since $\lambda_j \neq \lambda_k$ for all $1 \leqslant j \leqslant k-1$, we would have $c_1 = c_2 = \cdots = c_{k-1} = 0$. However, this would mean that $\boldsymbol{v}_k = \mathbf{0}$ which contradicts the hypothesis that $\boldsymbol{v}_k$ is an eigenvector. Thus, we conclude that the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$ are linearly independent. □

5.0.7 Corollary. *Each linear operator on vector space V has at most* $\dim V$ *distinct eigenvalues.*

Proof. The dimension of a vector space give an upper bound on the number of linearly independent vectors [2.2.1]. The assertion follows because distinct eigenvalues have linearly independent eigenvectors [5.0.6]. □

Exercises

5.0.8 Problem. Determine which of the following statements are true. If a statement is false, then provide a counterexample.

i. The scalar 1 is always an eigenvalue for the identity operator.
ii. The only eigenvalue of the identity operator is 1.
iii. On a fixed $\mathbb{K}$-vector space V, every scalar in $\mathbb{K}$ is an eigenvalue for some linear operator on V.
iv. A linear operator on a finite-dimensional vector space may have infinitely many eigenvectors.
v. The scalar 0 can never be an eigenvalue.

5.1 Characteristic Polynomials

HOW DO WE DETERMINE JUST THE EIGENVALUES OF A LINEAR OPERATOR? Geometrically, applying a linear operator to an eigenvector just rescales the eigenvector by the eigenvalue. However, it is useful to have another algebraic mechanism for computing the eigenvalues which does not require knowing the eigenvectors.

5.1.0 Problem. Find the eigenvalues of the linear operator $R\colon \mathbb{C}^2 \to \mathbb{C}^2$ defined, for all scalars x and y in $\mathbb{C}$, by $R[x\,\boldsymbol{e}_1 + y\,\boldsymbol{e}_2] = -y\,\boldsymbol{e}_1 + x\,\boldsymbol{e}_2$.

Solution. Suppose that the scalar λ in $\mathbb{C}$ is an eigenvalue of the linear operator R. By definition [5.0.0], there are scalars x and y in $\mathbb{C}$, not both zero, such that $-y\boldsymbol{e}_1 + x\boldsymbol{e}_2 = R[x\boldsymbol{e}_1 + y\boldsymbol{e}_2] = (\lambda x)\boldsymbol{e}_1 + (\lambda y)\boldsymbol{e}_2$. Comparing coefficients gives $-y = \lambda x$ and $x = \lambda y$. We deduce that both x and y are nonzero. Moreover, we obtain $-y = \lambda^2 y$ and $0 = (1 - \lambda^2) y$ which implies that $\lambda = \pm \mathrm{i}$. □

5.1.1 Remark. Over the real numbers, the linear operator $R\colon \mathbb{R}^2 \to \mathbb{R}^2$, defined by $R[x\boldsymbol{e}_1 + y\boldsymbol{e}_2] = -y\,\boldsymbol{e}_1 + x\,\boldsymbol{e}_2$, does not have any eigenvalues. This linear operator is just a counterclockwise rotation by the angle $\pi/2$ about the origin. This rotation of a nonzero vector in $\mathbb{R}^2$ is never equal a scalar multiple of itself, so R has no real eigenvalues.

5.1.2 Definition. Let $\mathcal{B} := (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ be an ordered basis for a $\mathbb{K}$-vector space V and consider the linear operator $T\colon V \to V$. The ***characteristic polynomial*** of T, lying in $\mathbb{K}[t]_{\leqslant n}$, is equal to

$$\mathrm{p}_T(t) := \det\big((t\,\mathrm{id}_V - T)^{\mathcal{B}}_{\mathcal{B}}\big)\,.$$

The definition requires the vector space to be finite-dimensional because determinants are only defined for square matrices.

5.1.3 Remark. Similar matrices have the same determinant [4.2.4] and represent the same linear operator relative to different bases [4.2.7], so the characteristic polynomial of the linear operator T does not depend on the choice of the basis $\mathcal{B}$.

Choosing the standard basis for $\mathbb{K}^n$, we see that, for any $(n \times n)$-matrix $\mathbf{A}$, the characteristic polynomial is

$$\mathrm{p}_{\mathbf{A}}(t) = \det(t\,\mathbf{I} - \mathbf{A})\,.$$

5.1.4 Problem. Find the characteristic polynomial of the matrix

$$\mathbf{A} := \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Solution. Since the determinant of a triangular matrix is the product of its diagonal entries, we have

$$\det(t\,\mathbf{I} - \mathbf{A}) = \det\left(\begin{bmatrix} t-5 & 2 & -6 & 1 \\ 0 & t-3 & 8 & 0 \\ 0 & 0 & t-5 & -4 \\ 0 & 0 & 0 & t-1 \end{bmatrix}\right) = (t-5)^2(t-3)(t-1)\,,$$

so $\mathrm{p}_{\mathbf{A}}(t) = (t-5)^2(t-3)(t-1) = t^4 - 14\,t^3 + 68\,t^2 - 130\,t + 75$. □

5.1.5 Theorem. *For a linear operator on a finite-dimensional vector space, the roots of its characteristic polynomial coincide with its spectrum.*

Proof. Let V be a vector space of dimension n, let $\mathcal{B}$ be an order basis for V, and let $T\colon V \to V$ be a linear operator. The characterization of the determinant implies that the linear operator $t\,\mathrm{id}_V - T$ is invertible if and only if $\det\big((t\,\mathrm{id}_V - T)^{\mathcal{B}}_{\mathcal{B}}\big) \neq 0$, so the assertion follows from the characterization of eigenvalues [5.0.4]. □

5.1.6 Corollary. *Every linear operator on a nonzero finite-dimensional $\mathbb{C}$-vector space has an eigenvalue.*

Proof. The fundamental theorem of algebra states that any non-constant univariate polynomial with complex coefficients has a complex root. Because the characteristic polynomial of the linear operator has a root, Theorem 5.1.5 demonstrates that the linear operator has an eigenvalue. □

5.1.7 Corollary. *A linear operator on a finite-dimensional vector space is invertible if and only if* 0 *is not an eigenvalue.*

Proof. Since there exists a multiplicative isomorphism between the vector space of linear operators and the vector space of square matrices [4.1.4, 4.2.6], we need only consider an $(n \times n)$-matrix $\mathbf{A}$ for some positive integer n. The characterization of the determinant establishes that the matrix $\mathbf{A}$ is invertible if and only if we have $0 \neq \det(\mathbf{A}) = (-1)^n \det(0\,\mathbf{I} - \mathbf{A})$ which means the number 0 is not a root of the characteristic polynomial $\mathrm{p}_{\mathbf{A}}(t) = \det(t\,\mathbf{I} - \mathbf{A})$. □

5.1.8 Problem. For any (2×2)-matrix $\mathbf{A}$, prove that

$$\mathrm{p}_{\mathbf{A}}(t) = t^2 - \operatorname{tr}(\mathbf{A})\, t + \det(\mathbf{A})\,.$$

Solution. Let $\mathbf{A} := \left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right]$ for some scalars a, b, c, and d. It follows that $\operatorname{tr}(\mathbf{A}) = a + d$, $\det(\mathbf{A}) = ad - bc$, and

$$\begin{aligned} \det\left(t \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} a & c \\ b & d \end{bmatrix} \right) &= \det\left(\begin{bmatrix} t-a & -c \\ -b & t-d \end{bmatrix} \right) \\ &= (t-a)(t-d) - bc = t^2 - (a+d)\,t + (ad-bc). \end{aligned}$$

We conclude that $\mathrm{p}_{\mathbf{A}}(t) = t^2 - \operatorname{tr}(\mathbf{A})\, t + \det(\mathbf{A})$. □

5.1.9 Definition. The ***algebraic multiplicity*** of an eigenvalue is its multiplicity as a root of the characteristic polynomial.

5.1.10 Problem. The characteristic polynomial of some linear operator is $t^6 - 4t^5 - 12t^4$. Find its eigenvalues and their algebraic multiplicities.

Solution. Factoring the characteristic polynomial gives

$$t^6 - 4t^5 - 12t^4 = t^4(t^2 - 4t - 12) = t^4(t-6)(t+2)\,.$$

Hence, the eigenvalues are 0 with algebraic multiplicity 4, 6 with algebraic multiplicity 1, and -2 with algebraic multiplicity 1. □

Exercises

5.1.11 Problem. Determine which of the following statements are true. If a statement is false, then provide a counterexample.

i. Every linear operator over a $\mathbb{K}$-vector space has an eigenvalue.
ii. The characteristic polynomial of a linear operator depends on choice of a basis for its underlying vector space.
iii. Similar matrices have the same characteristic polynomial.
iv. Matrices with the same characteristic polynomial are similar.
v. The algebraic multiplicity of an eigenvalue counts the number of associated eigenvectors.

5.2 Triangularization

WHEN IS THE MATRIX OF A LINEAR OPERATOR TRIANGULAR? As an initial attempt to find tractable matrices associated to a linear operator, we concentrated on ordered bases that produce upper triangular matrices.

5.2.0 Lemma. *Let T be a linear operator on $\mathbb{K}$-vector space V and let $\mathcal{B} := (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ be an ordered basis for V. The associated matrix $(T)_{\mathcal{B}}^{\mathcal{B}}$ is upper-triangular if and only if, for all $1 \leqslant k \leqslant n$, the image vector $T[\boldsymbol{v}_k]$ lies in the linear subspace* $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$.

Proof.
$\Rightarrow$: Suppose that the matrix $\mathbf{A} := (T)_{\mathcal{B}}^{\mathcal{B}}$ of the linear operator T relative to the ordered basis $\mathcal{B}$ is upper-triangular. For all $1 \leqslant j \leqslant n$ and all $1 \leqslant k \leqslant n$, let the scalar $a_{j,k}$ in $\mathbb{K}$ be the (j,k)-entry in the matrix $\mathbf{A}$. Since $a_{j,k} = 0$ when $j > k$, the definition of matrix multiplication implies that, for all $1 \leqslant k \leqslant n$, we have

$$\begin{aligned}(T[\boldsymbol{v}_k])_{\mathcal{B}} &= (T)_{\mathcal{B}}^{\mathcal{B}}\,(\boldsymbol{v}_k)_{\mathcal{B}} = \mathbf{A}\,\boldsymbol{e}_k \\ &= a_{1,k}\,\boldsymbol{e}_1 + a_{2,k}\,\boldsymbol{e}_2 + \cdots + a_{k,k}\,\boldsymbol{e}_k + 0\,\boldsymbol{v}_{k+1} + 0\,\boldsymbol{v}_{k+2} + \cdots + 0\,\boldsymbol{v}_n \\ &= a_{1,k}\,(\boldsymbol{v}_1)_{\mathcal{B}} + a_{2,k}\,(\boldsymbol{v}_2)_{\mathcal{B}} + \cdots + a_{k,k}\,(\boldsymbol{v}_k)_{\mathcal{B}}\,,\end{aligned}$$

For all $1 \leqslant k \leqslant n$, we see that $T[\boldsymbol{v}_k]$ lies in $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$.
$\Leftarrow$: Suppose that $T[\boldsymbol{v}_k] \in \operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ for all $1 \leqslant k \leqslant n$. Hence, there exists scalars $a_{1,k}, a_{2,k}, \dots, a_{k,k}$ in $\mathbb{K}$ such that

$$\begin{aligned}T[\boldsymbol{v}_k] &= a_{1,k}\,\boldsymbol{v}_1 + a_{2,k}\,\boldsymbol{v}_2 + \cdots + a_{k,k}\,\boldsymbol{v}_k \\ &= a_{1,k}\,\boldsymbol{v}_1 + a_{2,k}\,\boldsymbol{v}_2 + \cdots + a_{k,k}\,\boldsymbol{v}_k + 0\,\boldsymbol{v}_{k+1} + 0\,\boldsymbol{v}_{k+2} + \cdots + 0\,\boldsymbol{v}_n\,.\end{aligned}$$

It follows that $(T[\boldsymbol{v}_k])_{\mathcal{B}} = \begin{bmatrix} a_{1,k} & a_{2,k} & \cdots & a_{k,k} & 0 & 0 & \cdots & 0\end{bmatrix}^{\mathsf{T}}$ and the matrix $(T)_{\mathcal{B}}^{\mathcal{B}}$ is upper-triangular. $\square$

5.2.1 Theorem (Triangularization)**.** *Let V be a finite-dimensional $\mathbb{C}$-vector space. For any linear operator $T \colon V \to V$, there exists an ordered basis of V such that the matrix of T relative to this basis is upper-triangular.*

Inductive proof. We proceed by induction on $n := \dim V$. The base case $n = 1$ is vacuous because every (1×1)-matrix is upper-triangular.

Suppose that $n > 1$. The linear operator T has an eigenvalue λ in $\mathbb{C}$; see [5.1.6]. Consider the linear subspace $W := \operatorname{Im}(\lambda \operatorname{id}_V - T)$; see [3.1.2]. The linear operator $\lambda \operatorname{id}_V - T \colon V \to V$ is not surjective [5.0.4], so $\dim W < \dim V$. Given any vector $\boldsymbol{w}$ in W, we have $T[\boldsymbol{w}] = \lambda\, \boldsymbol{w} - (\lambda \operatorname{id}_V - T)[\boldsymbol{w}]$. Since the vector $\boldsymbol{w}$ is in W, the second term $(\lambda \operatorname{id}_V - T)[\boldsymbol{w}]$ is in W, and W is a linear subspace, it follows that image vector $T[\boldsymbol{w}]$ is also in W. Thus, the restriction $T|_W$ is a linear operator on W. By the induction hypothesis, there exists a ordered basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$ of W such that the matrix of the linear operator $T|_W$ relative to this basis is upper-triangular. Hence, the characterization of triangular operators [5.2.0] establishes that, for all $1 \leqslant k \leqslant m$, we have $T[\boldsymbol{v}_k] \in \operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$. Extending the basis for W to a basis for V, there exist vectors $\boldsymbol{v}_{m+1}, \boldsymbol{v}_{m+2}, \dots, \boldsymbol{v}_n$ in V such that $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ is a basis for V. For each $m+1 \leqslant j \leqslant n$, observe that $T[\boldsymbol{v}_j] = \lambda\, \boldsymbol{v}_j - (\lambda \operatorname{id}_V - T)[\boldsymbol{v}_j]$. Because $(\lambda \operatorname{id}_V - T)[\boldsymbol{v}_j]$ is in W, it follows that $T[\boldsymbol{v}_j] \in \operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k, \boldsymbol{v}_j) \subset \operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_j)$. Therefore, the characterization of triangular operators [5.2.0] proves that the matrix of T relative to $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ is upper-triangular. □

The next result highlights some of the benefits of finding a basis for which the matrix of a linear operator is upper-triangular.

5.2.2 Proposition. *Let V be a finite-dimensional $\mathbb{K}$-vector space, let T be a linear operator on V, and let $\mathcal{B}$ an order basis for V such that the matrix $\mathbf{A} := (T)_{\mathcal{B}}^{\mathcal{B}}$ is upper-triangular. The eigenvalues of T are precisely the entries on the diagonal of the matrix $\mathbf{A}$. Furthermore, the linear operator T is invertible if and only if all the entries on the diagonal of $\mathbf{A}$ are nonzero.*

Proof. For all $1 \leqslant j \leqslant n$ and all $1 \leqslant k \leqslant n$, let the scalar $a_{j,k}$ in $\mathbb{K}$ denote the (j,k)-entry in the matrix $\mathbf{A}$. By hypothesis, the matrix $\mathbf{A}$ is upper-triangular, so the matrix $t\,\mathbf{I} - \mathbf{A}$ is also upper-triangular. Since the map sending a linear operator to its corresponding matrix is linear [4.1.4] and the determinant of a triangular matrix is the product of its diagonal entries, the characteristic polynomial of the linear operator T is

$$\mathrm{p}_T(t) = \det\big((t \operatorname{id}_V - T)_{\mathcal{B}}^{\mathcal{B}}\big) = \det(t\,\mathbf{I} - \mathbf{A}) = \prod_{k=1}^{n} (t - a_{k,k})\,.$$

The roots of the characteristic polynomial [5.1.5] coincide with the spectrum, so the diagonal entries in the matrix $\mathbf{A}$ are the eigenvalues of the linear operator T. Lastly, a linear operator is invertible if and only if 0 is an eigenvalue [5.1.7]. □

This proposition fails without the upper-triangular hypothesis.

5.2.3 Problem. Exhibit an invertible linear operator whose matrix has only zeros on the diagonal and a non-invertible linear operator whose matrix as only nonzero entries on the diagonal.

Solution. Let $\mathcal{B} := (v_1, v_2, \dots, v_n)$ be an ordered basis for a $\mathbb{K}$-vector space V. Consider the linear maps $S \colon V \to V$ and $T \colon V \to V$ defined, respectively, by $S[v_n] := v_1$, $S[v_k] := v_{k+1}$ for all $1 \leqslant k < n$, and $T[v_j] = v_1 + v_2 + \cdots + v_n$ for all $1 \leqslant j \leqslant n$. It follows that $S^n = \mathrm{id}_V$,

$$(S)_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \text{and} \quad (T)_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Since $\det\big((T)_{\mathcal{B}}^{\mathcal{B}}\big) = 0$, the linear operator T is not invertible. □

Exercises

5.2.4 Problem. Determine which of the following statements are true. If a statement is false, then provide a counterexample.

i. Given a linear operator on any finite-dimensional $\mathbb{K}$-vector space, there exists an ordered basis such that the associated matrix is upper-triangular.

ii. Given a linear operator on any finite-dimensional $\mathbb{C}$-vector space, there exists an ordered basis such that the associated matrix is lower-triangular.

iii. On any finite-dimensional $\mathbb{C}$-vector space, there exists an ordered basis such that, for every linear operator, the associated matrix is upper-triangular.

5.3 Estimating Eigenvalues

HOW CAN WE FIND BOUNDS ON THE EIGENVALUES OF A MATRIX? There are simple geometric ways to estimate the eigenvalues of a complex square matrix. To describe the heuristic for generating these bounds, let the scalar $a_{j,k}$ in $\mathbb{C}$, for all $1 \leqslant j \leqslant n$ and all $1 \leqslant k \leqslant n$, denote the (j,k)-entry in the matrix $\mathbf{A}$. There is a unique expression $\mathbf{A} = \mathbf{D} + \mathbf{B}$ where

$$\mathbf{D} := \begin{bmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix} \quad \text{and} \quad \mathbf{B} := \begin{bmatrix} 0 & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & 0 & a_{2,3} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & 0 \end{bmatrix}.$$

Consider the perturbations $\mathbf{A}_\varepsilon := \mathbf{D} + \varepsilon\,\mathbf{B}$ for any scalar ε in $\mathbb{C}$. In particular, we have $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_0 = \mathbf{D}$. The eigenvalues of $\mathbf{A}_0 = \mathbf{D}$ are clearly $a_{1,1}, a_{2,2}, \dots, a_{n,n}$ because $\mathbf{D}$ is a triangular matrix. It seems plausible that, when $|\varepsilon|$ is small enough, the eigenvalues of $\mathbf{A}_\varepsilon$ will be near $a_{1,1}, a_{2,2}, \dots, a_{n,n}$. The circle theorem makes this idea precise.

5.3.0 Definition. For all $1 \leqslant j \leqslant n$, let r_j be the sum of the absolute values of the off-diagonal entries in the j-th row of the matrix $\mathbf{A}$;

$$\begin{aligned} r_j &:= \sum_{\substack{k=1\\k\neq j}}^{n} |a_{j,k}| \\ &= |a_{j,1}| + |a_{j,2}| + \cdots + |a_{j,j-1}| + |a_{j,j+1}| + |a_{j,j+2}| + \cdots + |a_{j,n}| \,. \end{aligned}$$

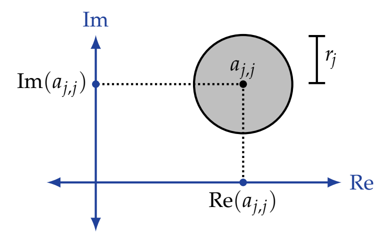

Figure 5.0: A Gershgorin disk D_j

The j-th ***Gershgorin disk*** is the circular disk D_j lying in the complex plane with center $a_{j,j}$ in $\mathbb{C}$ and radius r_j in $\mathbb{R}$:

$$D_j := \left\{ z \in \mathbb{C} \;\middle|\; |z - a_{j,j}| \leqslant r_j \right\}.$$

5.3.1 Problem. Find the Gershgorin disks for the matrix

$$\begin{bmatrix} 1+2\,\mathrm{i} & 1 \\ 4+3\,\mathrm{i} & -3 \end{bmatrix}.$$

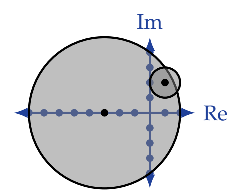

Figure 5.1: The Gershgorin disks for Problem 5.3.1

Solution. We have $r_1 = |1| = 1$ and $r_2 = |4+3\,\mathrm{i}| = \sqrt{4^2+3^2} = 5$, so $D_1 = \left\{ z \in \mathbb{C} \;\middle|\; |z - (1+2\,\mathrm{i})| \leqslant 1 \right\}$ and $D_2 = \left\{ z \in \mathbb{C} \;\middle|\; |z+3| \leqslant 5 \right\}$. □

5.3.2 Theorem (Gershgorin circle)**.** *For any complex $(n \times n)$-matrix $\mathbf{A}$, every eigenvalue of $\mathbf{A}$ is contained in a Gershgorin disk.*

The circle theorem was first published by Semyon Gershgorin in 1931.

Proof. Given an eigenvalue λ in $\mathbb{C}$ of the matrix $\mathbf{A}$ with eigenvector $\boldsymbol{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^\mathsf{T}$, we have $\mathbf{A}\,\boldsymbol{v} = \lambda\,\boldsymbol{v}$. By comparing entries, we obtain $\sum_{k=1}^{n} a_{j,k}\, v_k = \lambda v_j$ for all $1 \leqslant j \leqslant n$. Let v_m be a coordinate of the vector $\boldsymbol{v}$ having the largest absolute value. When $j = m$, the triangle inequality gives

$$\begin{aligned} |\lambda\, v_m - a_{m,m}\, v_m| = \left| \sum_{k=1}^{n} a_{m,k}\, v_k - a_{m,m}\, v_m \right| &= \left| \sum_{\substack{k=1\\k\neq m}}^{n} a_{m,k}\, v_k \right| \\ &\leqslant \sum_{\substack{k=1\\k\neq m}}^{n} |a_{m,k}|\,|v_k| \leqslant \sum_{\substack{k=1\\k\neq m}}^{n} |a_{m,k}|\,|v_m| = |v_m| \sum_{\substack{k=1\\k\neq m}}^{n} |a_{m,k}| = |v_m|\, r_m \,. \end{aligned}$$

We conclude that $|v_m|\,|\lambda - a_{m,m}| \leqslant |v_m|\, r_m$ and $|\lambda - a_{m,m}| \leqslant r_m$. □

5.3.3 Problem. Sketch disks that contain the eigenvalues of

$$\mathbf{C} := \begin{bmatrix} 7 & 2 & -1 \\ -1 & 10 & 1 \\ -1 & 1 & 6 \end{bmatrix}.$$

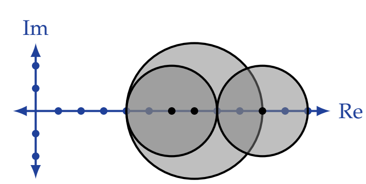

Figure 5.2: The Gershgorin disks for Problem 5.3.3

Solution. The disks are $\left\{ z \in \mathbb{C} \;\middle|\; |z-7| \leqslant 3 \right\}$, $\left\{ z \in \mathbb{C} \;\middle|\; |z-10| \leqslant 2 \right\}$, and $\left\{ z \in \mathbb{C} \;\middle|\; |z-6| \leqslant 2 \right\}$. This matrix is invertible because the Gershgorin disks do not contain the origin. □

5.3.4 Lemma. *For any square matrix* $\mathbf{A}$, *both* $\mathbf{A}$ *and* $\mathbf{A}^{\mathsf{T}}$ *have the same eigenvalues.*

Proof. Since $\det(t\,\mathbf{I} - \mathbf{A}) = \det\big((t\,\mathbf{I} - \mathbf{A})^{\mathsf{T}}\big) = \det(t\,\mathbf{I} - \mathbf{A}^{\mathsf{T}})$ and the roots of the characteristic polynomial equal the spectrum [5.1.5], the claim follows. $\square$

Since $\mathbf{A}^{\mathsf{T}}$ has the same eigenvalues as $\mathbf{A}$, the circle theorem can sometimes be applied to $\mathbf{A}^{\mathsf{T}}$ to get better estimates. In particular, we may intersect the Gershgorin regions for $\mathbf{A}$ and $\mathbf{A}^{\mathsf{T}}$. For instance, the eigenvalues of the matrix $\mathbf{C}$ lie in the region

$$\big\{z \in \mathbb{C} \bigm| |z-7| \leqslant 2\big\} \cup \big\{z \in \mathbb{C} \bigm| |z-10| \leqslant 2\big\} \cup \big\{z \in \mathbb{C} \bigm| |z-6| \leqslant 2\big\}\,.$$

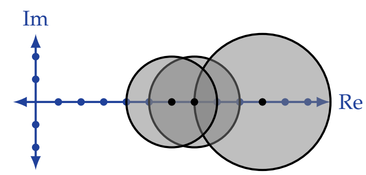

Figure 5.3: The Gershgorin disks for $\mathbf{C}^{\mathsf{T}}$

Let $\mathbf{P}$ be an invertible matrix. Since the matrix $\mathbf{P}^{-1}\,\mathbf{A}\,\mathbf{P}$ has the same eigenvalues as $\mathbf{A}$, we can also apply the circle theorem to $\mathbf{P}^{-1}\,\mathbf{A}\,\mathbf{P}$. For some choice of $\mathbf{P}$, the bounds obtained may be sharper. A convenient choice is $\mathbf{D} = \operatorname{diag}(d_1, d_2, \dots, d_n)$ with $d_j > 0$. In this case, we have $\mathbf{D}^{-1}\,\mathbf{A}\,\mathbf{D} = [d_k a_{j,k}/d_j]$.

5.3.5 Problem. Sketch disks that contain the eigenvalues of

$$\begin{bmatrix} 0 & d_2/d_1 & 0 & 0 \\ d_1/d_2 & 5 & d_3/d_2 & 0 \\ 0 & d_2/d_3 & 20 & d_4/d_3 \\ 0 & 0 & d_3/d_4 & 20 \end{bmatrix}$$

for a few values of $(d_1, d_2, d_3, d_4) \in \mathbb{N}^4$.

Solution. Applying the Gershgorin Circle Theorem yields Table 5.1.

(d_1, d_2, d_3, d_4)	Gershgorin disks
$(1,1,1,1)$	$\big\{z \in \mathbb{C} \bigm\| \|z\| \leqslant 1\big\} \cup \big\{z \in \mathbb{C} \bigm\| \|z-5\| \leqslant 2\big\} \cup \big\{z \in \mathbb{C} \bigm\| \|z-20\| \leqslant 2\big\}$
$(2,1,1,1)$	$\big\{z \in \mathbb{C} \bigm\| \|z\| \leqslant \frac{1}{2}\big\} \cup \big\{z \in \mathbb{C} \bigm\| \|z-5\| \leqslant 3\big\} \cup \big\{z \in \mathbb{C} \bigm\| \|z-20\| \leqslant 3\big\}$
$(1,2,1,1)$	$\big\{z \in \mathbb{C} \bigm\| \|z\| \leqslant 2\big\} \cup \big\{z \in \mathbb{C} \bigm\| \|z-5\| \leqslant 1\big\} \cup \big\{z \in \mathbb{C} \bigm\| \|z-20\| \leqslant 2\big\}$

Table 5.1: Gershgorin disks for various similar matrices

Intersecting these regions shows that the eigenvalues are contained in

$$\big\{z \in \mathbb{C} \bigm| |z| \leqslant \tfrac{1}{2}\big\} \cup \big\{z \in \mathbb{C} \bigm| |z-5| \leqslant 1\big\} \cup \big\{z \in \mathbb{C} \bigm| |z-20| \leqslant 2\big\}\,.$$

Using numerical methods, one shows that the eigenvalues are *approximately* -0.194735438, 5.127483946, 19.03511963, 21.03176997. $\square$

Exercises

5.3.6 Problem. Determine which of the following statements are true. If a statement is false, then provide a counterexample.

i. The Gershgorin Theorem allows one to determine exactly the eigenvalues of any matrix.

ii. The Gerschgorin Theorem never allows one to determine exactly the eigenvalues of a matrix.
iii. The Gerschgorin disks of a triangular matrix have radii of length zero.

5.3.7 Problem. A square matrix in ***strictly diagonally dominant*** if the absolute value of each diagonal entry is greater than the sum of the absolute values of the remaining entries in that row. Prove that a strictly diagonally dominant matrix must be invertible.

5.3.8 Problem. Sketch the Gerschgorin disks in the complex plane that contain the eigenvalues of the matrix

$$\mathbf{A} := \begin{bmatrix} 10 & 0 & -1 & 1 \\ -1 & 12\,\mathrm{i} & -1 & 2 \\ -1 & 3 & 20 & 2 \\ 1 & 2\,\mathrm{i} & 3 & -45 \end{bmatrix} .$$