Appendix A

Background material

In this chapter, we collect basic results that are useful for this course.

A.1 Inner products and norms

We begin by recalling the definitions of the fundamental concepts of *norm* and *inner product*. For generality, we consider the case of a *complex* vector space, i.e. a vector space for which the scalar field is $\mathbf{C}$.

Definition A.1. A norm on a complex vector space $\mathcal{X}$ is a function $\|\bullet\| : \mathcal{X} \to \mathbf{R}$ satisfying the following axioms:

- **Positivity**: $\forall \boldsymbol{x} \in \mathcal{X} \backslash \{\mathbf{0}\}, \quad \|\boldsymbol{x}\| > 0.$
- **Homogeneity**: $\forall (c, \boldsymbol{x}) \in \mathbf{C} \times \mathcal{X}, \quad \|c\boldsymbol{x}\| = |c| \, \|\boldsymbol{x}\|.$
- **Triangular inequality**: $\forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{X}, \quad \|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|.$

For example, the Euclidean norm on $\mathbf{C}^n$ is given by

$$\|\boldsymbol{x}\| = \sqrt{|x_1|^2 + \ldots + |x_n|^2}.$$

Definition A.2. An inner product on a *complex* vector space $\mathcal{X}$ is a function

$$\langle \bullet, \bullet \rangle : \mathcal{X} \times \mathcal{X} \to \mathbf{C}$$

satisfying the following axioms:

- **Conjugate symmetry**: Here $\overline{\bullet}$ denotes the complex conjugate.

$$\forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{X}, \qquad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}.$$

- **Linearity**: For all $(\alpha, \beta) \in \mathbf{C}^2$ and all $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{X}^3$, it holds that

$$\langle \alpha \boldsymbol{x} + \beta \boldsymbol{y}, \boldsymbol{z} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \beta \langle \boldsymbol{y}, \boldsymbol{z} \rangle.$$

- **Positive-definiteness**:

$$\forall \boldsymbol{x} \in \mathcal{X} \backslash \{0\}, \qquad \langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0.$$

For example, the familiar Euclidean inner product on $\mathbf{C}^n$ is given by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{n} x_i \overline{y}_i.$$

A vector space with an inner product is called an *inner product space*. Any inner product on $\mathcal{X}$ induces a norm via the formula

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}. \tag{A.1}$$

The Cauchy–Schwarz inequality enables to bound inner products using norms. It is also useful for showing that the functional defined in (A.1) satisfies the triangle inequality, which is the goal of Exercise A.2.

Proposition A.1 (Cauchy–Schwarz inequality)**.** *Let $\mathcal{X}$ be an inner product space. Then*

$$\forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{X}, \qquad |\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \|\boldsymbol{y}\|. \tag{A.2}$$

Proof. The statement is obvious if $\boldsymbol{y} = \boldsymbol{0}$, so we assume in the rest of the proof that $\boldsymbol{y} \neq \boldsymbol{0}$. Let us define $p \colon \mathbf{R} \ni t \mapsto \|\boldsymbol{x} + t\boldsymbol{y}\|^2$. Using the bilinearity of the inner product, we have

$$p(t) = \|\boldsymbol{x}\|^2 + 2t \langle \boldsymbol{x}, \boldsymbol{y} \rangle + t^2 \|\boldsymbol{y}\|^2.$$

This shows that p is a convex second-order polynomial with a minimum at $t_* = -\langle \boldsymbol{x}, \boldsymbol{y} \rangle / \|\boldsymbol{y}\|^2$. Substituting this value in the expression of p, we obtain

$$p(t_*) = \|\boldsymbol{x}\|^2 - 2 \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\|\boldsymbol{y}\|^2} + \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\|\boldsymbol{y}\|^2} = \|\boldsymbol{x}\|^2 - \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\|\boldsymbol{y}\|^2}.$$

Since $p(t_*) \geqslant 0$ by definition of p, we obtain (A.2). □

Several norms can be defined on the same vector space $\mathcal{X}$. Two norms $\|\bullet\|_\alpha$ and $\|\bullet\|_\beta$ on $\mathcal{X}$ are said to be equivalent if there exist positive real numbers c_ℓ and c_u such that

$$\forall \boldsymbol{x} \in \mathcal{X}, \qquad c_\ell \|\boldsymbol{x}\|_\alpha \leqslant \|\boldsymbol{x}\|_\beta \leqslant c_u \|\boldsymbol{x}\|_\alpha. \tag{A.3}$$

As the terminology indicates, norm equivalence is an *equivalence relation*. When working with norms on finite-dimensional vector spaces, it is important to keep in mind the following result. The proof is provided for information purposes only.

Proposition A.2. *Assume that $\mathcal{X}$ is a finite-dimensional vector space. Then all the norms defined on $\mathcal{X}$ are pairwise equivalent.*

Proof. Let $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$ be a basis of $\mathcal{X}$. Any $\boldsymbol{x} \in \mathcal{X}$ admits a unique representation in this basis as $\boldsymbol{x} = \lambda_1 \boldsymbol{e}_1 + \dots + \lambda_n \boldsymbol{e}_n$. We will show that any norm $\|\bullet\|$ on $\mathcal{X}$ is equivalent to the norm $\|\bullet\|_*$ given by

$$\|\boldsymbol{x}\|_* = |\lambda_1| + \ldots + |\lambda_n|. \tag{A.4}$$

By the triangle inequality, it holds that

$$\begin{aligned} \|\boldsymbol{x}\| \leqslant |\lambda_1|\|\boldsymbol{e}_1\| + \dots + |\lambda_n|\|\boldsymbol{e}_n\| &\leqslant \Big(|\lambda_1| + \dots + |\lambda_n|\Big) \max\Big\{\|\boldsymbol{e}_1\|, \dots, \|\boldsymbol{e}_n\|\Big\} \\ &= \|\boldsymbol{x}\|_* \max\Big\{\|\boldsymbol{e}_1\|, \dots, \|\boldsymbol{e}_n\|\Big\}. \end{aligned} \tag{A.5}$$

It remains to show that there exists a positive constant ℓ such that

$$\forall \boldsymbol{x} \in \mathcal{X}, \qquad \|\boldsymbol{x}\| \geqslant \ell\Big(|\lambda_1| + \dots + |\lambda_n|\Big). \tag{A.6}$$

To this end, we reason by contradiction. If this inequality were not true, then there would exist a sequence $(\boldsymbol{x}^{(i)})_{i \in \mathbf{N}}$ such that $\|\boldsymbol{x}^{(i)}\| \to 0$ as $i \to \infty$ and $\|\boldsymbol{x}^{(i)}\|_* = 1$ for all $i \in \mathbf{N}$. Since $\lambda_1^{(i)} \in [-1, 1]$ for all $i \in \mathbf{N}$, we can extract a subsequence, still denoted by $(\boldsymbol{x}^{(i)})_{i \in \mathbf{N}}$ for simplicity, such that the corresponding coefficient $\lambda_1^{(i)}$ satisfies $\lambda_1^{(i)} \to \lambda_1^* \in [-1, 1]$, by compactness of the interval $[-1, 1]$. Repeating this procedure for $\lambda_2, \lambda_3, \dots,$ taking a new subsequence every time, we obtain a subsequence $(\boldsymbol{x}^{(i)})_{i \in \mathbf{N}}$ such that $\lambda_j^{(i)} \to \lambda_j^*$ in the limit as $i \to \infty$, for all $j \in \{1, \dots, n\}$. Therefore, it holds that $\boldsymbol{x}^{(i)} \to \boldsymbol{x}^* := \lambda_1^* \boldsymbol{e}_1 + \cdots \lambda_n^* \boldsymbol{e}_n$ in the $\|\bullet\|_*$ norm, and thus also in the $\|\bullet\|$ norm by (A.5). Since $\boldsymbol{x}^{(i)} \to \boldsymbol{0}$ in the latter norm by assumption, we deduce that $\boldsymbol{x}^* = \boldsymbol{0}$. But the vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ are linearly independent, and so this implies that $\lambda_1^* = \dots = \lambda_n^* = 0$, which is a contradiction because we also have that

$$|\lambda_1^*| + \dots + |\lambda_n^*| = \lim_{i \to \infty} \left|\lambda_1^{(i)}\right| + \dots + \left|\lambda_n^{(i)}\right| = 1.$$

This concludes the proof of (A.6). □

Exercise A.1. *Show that $\|\bullet\|_* \colon \mathcal{X} \to \mathbf{R}$ defined in (A.4) is indeed a norm.*

Exercise A.2. *Using Proposition A.1, show that the function $\|\bullet\|$ defined by (A.1) satisfies the triangle inequality.*

A.2 Completeness

Assume that $\mathcal{X}$ is a vector space with a norm $\|\bullet\|$. Together, $(\mathcal{X}, \|\bullet\|)$ form a *normed vector space*. A sequence $(\boldsymbol{x}_n)_{n \geqslant 0}$ in $\mathcal{X}$ is convergent in this space if there exists $\boldsymbol{x}_* \in \mathcal{X}$ such that

$$\|\boldsymbol{x}_n - \boldsymbol{x}_*\| \to 0 \quad \text{ in the limit } n \to \infty.$$

In this case, we write $\lim_{n \to \infty} \boldsymbol{x}_n = \boldsymbol{x}_*$ or $\boldsymbol{x}_n \to \boldsymbol{x}_*$.

Definition A.3 (Cauchy sequence)**.** A sequence $(\boldsymbol{x}_n)_{n\geqslant 0}$ in $\mathcal{X}$ is said to be Cauchy if

$$\lim_{m,n\to+\infty} \|\boldsymbol{x}_n - \boldsymbol{x}_m\| = 0.$$

The normed vector space $(\mathcal{X}, \|\bullet\|)$ is called *complete* if every Cauchy sequence is convergent.

Every convergent sequence is Cauchy, but the converse is not always true.

Example A.1. Consider the case where $\mathcal{X} = C([-1,1])$, the space of continuous functions from $[-1,1]$ to $\mathbf{R}$, endowed with the norm

$$\|f\| = \int_{-1}^{1} |f(x)| \, \mathrm{d}x.$$

The sequence of functions $(f_n)_{n\geqslant 0}$ in $\mathcal{X}$ given by

$$f_n(x) = x^{\frac{1}{2n+1}} \tag{A.7}$$

is Cauchy but not convergent. Indeed, assume for contradiction that there existed $f_* \in \mathcal{X}$ such that

$$\|f_n - f_*\| \xrightarrow[n\to\infty]{} 0. \tag{A.8}$$

Then f_* necessarily coincides with sign function:

$$\operatorname{sgn}(x) := \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

However, this function is discontinuous, which contradicts the statement that $f_* \in \mathcal{X}$.

In this course, all the vector spaces encountered are complete. For example,

- $\mathbf{R}^n$ with any vector norm is complete;
- $\mathbf{C}^{m\times n}$ with any matrix norm is complete.

In order to show that a sequence is convergent in a complete normed vector space, it is sufficient to show that the sequence is Cauchy. This approach is used in Lemma 4.2 and Theorem A.3.

Exercise A.3. *Prove that every convergent sequence is Cauchy.*

A.3 Contraction mappings and the Banach fixed point theorem

Let $(\mathcal{X}, \|\bullet\|)$ denote a normed vector space. A map $\phi\colon \mathcal{X} \to \mathcal{X}$ is called a contraction mapping if there is a constant $L \in (0,1)$ such that

$$\forall (x,y) \in \mathcal{X}\times\mathcal{X}, \qquad \|\phi(x) - \phi(y)\| \leqslant L\|x - y\|.$$

The importance of contraction mappings in this course stems from the following theorem.

Theorem A.3 (Banach fixed point theorem)**.** *Let $(\mathcal{X}, \|\bullet\|)$ be a complete normed space, and let $\phi\colon \mathcal{X} \to \mathcal{X}$ be a contraction mapping. Then ϕ has a unique fixed point in $\mathcal{X}$.*

Proof. We prove first existence and then uniqueness.
Existence. Take $x_0 \in \mathcal{X}$, and define the sequence $(x_k)_{k\in\mathbf{N}}$ inductively by

$$x_{k+1} = \phi(x_k). \tag{A.9}$$

It holds that

$$\|x_{k+1} - x_k\| = \|\phi(x_k) - \phi(x_{k-1})\| \leqslant L\|x_k - x_{k-1}\| \leqslant \ldots \leqslant L^k\|x_1 - x_0\|.$$

Therefore, for any $n \geqslant m$, we have by the triangle inequality

$$\begin{aligned}\|x_n - x_m\| &\leqslant \|x_n - x_{n-1}\| + \cdots + \|x_{m+1} - x_m\| \\ &\leqslant (L^{n-1} + \cdots + L^m)\|x_1 - x_0\| \\ &\leqslant L^m(1 + L + L^2 + \cdots)\|x_1 - x_0\| = \frac{L^m}{1-L}\|x_1 - x_0\|.\end{aligned}$$

It follows that the sequence $(x_k)_{k\in\mathbf{N}}$ is Cauchy in $\mathcal{X}$, implying by completeness that $x_k \to x_*$ in the limit as $k \to \infty$, for some limit $x_* \in \mathcal{X}$. Being a contraction, the mapping ϕ is continuous, and so taking the limit $k \to \infty$ in (A.9), we obtain that

$$x_* = \lim_{k\to\infty} x_{k+1} = \lim_{k\to\infty} \phi(x_k) = \phi\left(\lim_{k\to\infty} x_k\right) = \phi(x_*).$$

In other words, x_* is a fixed point of ϕ.

Uniqueness. Assume that $y_* \in \mathcal{X}$ is a fixed point. Then,

$$\|y_* - x_*\| = \|\phi(y_*) - \phi(x_*)\| \leqslant L\|y_* - x_*\|,$$

which implies that $y_* = x_*$ since $L < 1$. □

Remark A.1. The Banach fixed point theorem holds also in complete metric spaces.

A.4 Vector norms

In the vector space $\mathbf{C}^n$, the most commonly used norms are particular cases of the p-norm, also called Hölder norm.

Definition A.4. Given $p \in [1, \infty]$, the p-norm of a vector $\boldsymbol{x} \in \mathbf{C}^n$ is defined as follows:

$$\|\boldsymbol{x}\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} & \text{if } p < \infty, \\ \max\Big\{|x_1|, \ldots, |x_n|\Big\} & \text{if } p = \infty. \end{cases}$$

The values of p most commonly encountered in applications are 1, 2 and ∞. The 1-norm is sometimes called the *taxicab* or *Manhattan* norm, and the 2-norm is usually called the *Euclidean norm*. The explicit expressions of these norms are

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Notice that the infinity norm $\|\bullet\|_\infty$ may be defined as the limit of the p-norm as $p \to \infty$:

$$\|\boldsymbol{x}\|_\infty := \lim_{p \to \infty} \|\boldsymbol{x}\|_p.$$

In the rest of this chapter, the notations $\langle \bullet, \bullet \rangle$ and $\|\bullet\|$ without subscript always refer to the Euclidean inner product (A.1) and induced norm, unless specified otherwise.

A.5 Matrix norms

Given two norms $\|\bullet\|_\alpha$ and $\|\bullet\|_\beta$ on $\mathbf{C}^m$ and $\mathbf{C}^n$, respectively, we define the *operator norm* induced by $\|\bullet\|_\alpha$ and $\|\bullet\|_\beta$ of the matrix A as

$$\|\mathsf{A}\|_{\alpha,\beta} = \sup\left\{\|\mathsf{A}\boldsymbol{x}\|_\alpha : \boldsymbol{x} \in \mathbf{C}^n, \|\boldsymbol{x}\|_\beta \leqslant 1\right\}. \tag{A.10}$$

The term *operator norm* is motivated by the fact that, to any matrix $\mathsf{A} \in \mathbf{C}^{m \times n}$, there naturally corresponds the linear operator from $\mathbf{C}^n$ to $\mathbf{C}^m$ with action $\boldsymbol{x} \mapsto \mathsf{A}\boldsymbol{x}$. Matrix norms of the type (A.10) are also called *subordinate* matrix norms. An immediate corollary of the definition (A.10) is that, for all $\boldsymbol{x} \in \mathbf{C}^n$,

$$\|\mathsf{A}\boldsymbol{x}\|_\alpha = \|\mathsf{A}\widehat{\boldsymbol{x}}\|_\alpha \|\boldsymbol{x}\|_\beta \leqslant \sup\left\{\|\mathsf{A}\boldsymbol{y}\|_\alpha : \|\boldsymbol{y}\|_\beta \leqslant 1\right\} \|\boldsymbol{x}\|_\beta = \|\mathsf{A}\|_{\alpha,\beta} \|\boldsymbol{x}\|_\beta, \qquad \widehat{\boldsymbol{x}} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_\beta}. \tag{A.11}$$

Exercise A.4. *Show that equation* (A.10) *defines a norm on* $\mathbf{C}^{m \times n}$.

The matrix p-norm is defined as the operator norm (A.10) in the particular case where $\|\bullet\|_\alpha$ and $\|\bullet\|_\beta$ are both Hölder norms with the same value of p.

Definition A.5. Given $p \in [1, \infty]$, the p-norm of a matrix $\mathsf{A} \in \mathbf{C}^{m \times n}$ is given by

$$\|\mathsf{A}\|_p := \sup\left\{\|\mathsf{A}\boldsymbol{x}\|_p : \boldsymbol{x} \in \mathbf{C}^n, \|\boldsymbol{x}\|_p \leqslant 1\right\}. \tag{A.12}$$

Not all matrix norms are induced by vector norms. For example, the Frobenius norm, which is widely used in applications, is not induced by a vector norm. It is, however, induced by an inner product on $\mathbf{C}^{m \times n}$.

Definition A.6. The Frobenius norm of $\mathsf{A} \in \mathbf{C}^{m\times n}$ is given by

$$\|\mathsf{A}\|_{\mathrm{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}. \tag{A.13}$$

A matrix norm $\|\bullet\|$ is said to be submultiplicative if, for any two matrices $\mathsf{A} \in \mathbf{C}^{m\times n}$ and $\mathsf{B} \in \mathbf{C}^{n\times \ell}$, it holds that

$$\|\mathsf{AB}\| \leqslant \|\mathsf{A}\|\|\mathsf{B}\|.$$

All subordinate matrix norms, for example the p-norms, are submultiplicative, and so is the Frobenius norm.

Exercise A.5. *Write down the inner product on $\mathbf{C}^{m\times n}$ corresponding to (A.13).*

Exercise A.6. *Show that the matrix p-norm is submultiplicative.*

A.6 Diagonalization and spectral theorem

Definition A.7. A square matrix $\mathsf{A} \in \mathbf{C}^{n\times n}$ is said to be diagonalizable if there exists an invertible matrix $\mathsf{P} \in \mathbf{C}^{n\times n}$ and a diagonal matrix $\mathsf{D} \in \mathbf{C}^{n\times n}$ such that

$$\mathsf{AP} = \mathsf{PD}. \tag{A.14}$$

In this case, the diagonal elements of D are called the eigenvalues of A, and the columns of P are called the eigenvectors of A.

Denoting by $\boldsymbol{e}_i$ the i-th column of P and by λ_i the i-th diagonal element of D, we have by (A.14) that $\mathsf{A}\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i$ or, equivalently, $(\mathsf{A}-\lambda_i \mathsf{I}_n)\boldsymbol{e}_i = \mathbf{0}$. Here I_n is the $\mathbf{C}^{n\times n}$ identity matrix. Therefore, a complex number λ is an eigenvalue of A if and only if $\det(\mathsf{A} - \lambda \mathsf{I}_n) = 0$. In other words, the eigenvalues of A are the roots of $\det(\mathsf{A} - \lambda \mathsf{I}_n)$, which is called the *characteristic polynomial*.

Symmetric matrices and spectral theorem

The transpose of a matrix $\mathsf{A} \in \mathbf{C}^{m\times n}$ is denoted by $\mathsf{A}^T \in \mathbf{C}^{n\times m}$ and defined as the matrix with entries $a^T_{ij} = a_{ji}$. The conjugate transpose of A, denoted by A^*, is the matrix obtained by taking the transpose and taking the complex conjugate of all the entries. A real matrix that is equal to its transpose is necessarily square and called *symmetric*, and a complex matrix that is equal to its conjugate transpose is called *Hermitian*. Hermitian matrices, of which real symmetric matrices are a subset, enjoy many nice properties, the main one being that they are diagonalizable with a matrix Q that is unitary, i.e. such that $\mathsf{Q}^{-1} = \mathsf{Q}^*$. This is the content of the *spectral theorem*, a pillar of linear algebra with important generalizations to infinite-dimensional operators.

Theorem A.4 (Spectral theorem for Hermitian matrices)**.** *If $\mathsf{A} \in \mathbf{C}^{n\times n}$ is Hermitian, then there exists an unitary matrix $\mathsf{Q} \in \mathbf{C}^{n\times n}$ and a diagonal matrix $\mathsf{D} \in \mathbf{R}^{n\times n}$ such that*

$$\mathsf{AQ} = \mathsf{QD}.$$

Sketch of the proof. The result is trivial for $n = 1$. Reasoning by induction, we assume that the result is true for Hermitian matrices in $\mathbf{C}^{n-1\times n-1}$ and prove that it then also holds for $\mathsf{A} \in \mathbf{C}^{n\times n}$.

Step 1. Existence of a real eigenvalue. By the fundamental theorem of algebra, there exists at least one solution $\lambda_1 \in \mathbf{C}$ to the equation $\det(\mathsf{A} - \lambda \mathsf{I}_n) = 0$, to which there corresponds at least one solution $\boldsymbol{q}_1 \in \mathbf{C}^n$ of norm 1 to the equation $(\mathsf{A} - \lambda_1 \mathsf{I}_n)\boldsymbol{q}_1 = \mathbf{0}$. The eigenvalue λ_1 is necessarily real because

$$\lambda_1 \langle \boldsymbol{q}_1, \boldsymbol{q}_1 \rangle = \langle \lambda_1 \boldsymbol{q}_1, \boldsymbol{q}_1 \rangle = \langle \mathsf{A}\boldsymbol{q}_1, \boldsymbol{q}_1 \rangle = \langle \boldsymbol{q}_1, \mathsf{A}\boldsymbol{q}_1 \rangle = \langle \boldsymbol{q}_1, \lambda_1 \boldsymbol{q}_1 \rangle = \overline{\lambda}_1 \langle \boldsymbol{q}_1, \boldsymbol{q}_1 \rangle.$$

Step 2. Using the induction hypothesis. Next, take an orthonormal basis $(\boldsymbol{e}_2, \dots, \boldsymbol{e}_n)$ of the orthogonal complement $\mathrm{Span}\{\boldsymbol{q}_1\}^\perp$ and construct the unitary matrix

$$V = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{e}_2 & \dots & \boldsymbol{e}_n \end{pmatrix},$$

i.e. the matrix with columns $\boldsymbol{q}_1$, $\boldsymbol{e}_2$, etc. A calculation gives,

$$\mathsf{V}^*\mathsf{AV} = \begin{pmatrix} \langle \boldsymbol{q}_1, \mathsf{A}\boldsymbol{q}_1\rangle & \langle \boldsymbol{q}_1, \mathsf{A}\boldsymbol{e}_2\rangle & \dots & \langle \boldsymbol{q}_1, \mathsf{A}\boldsymbol{e}_n\rangle \\ \langle \boldsymbol{e}_2, \mathsf{A}\boldsymbol{q}_1\rangle & \langle \boldsymbol{e}_2, \mathsf{A}\boldsymbol{e}_2\rangle & \dots & \langle \boldsymbol{e}_2, \mathsf{A}\boldsymbol{e}_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{e}_n, \mathsf{A}\boldsymbol{q}_1\rangle & \langle \boldsymbol{e}_n, \mathsf{A}\boldsymbol{e}_2\rangle & \dots & \langle \boldsymbol{e}_n, \mathsf{A}\boldsymbol{e}_n\rangle \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \langle \boldsymbol{e}_2, \mathsf{A}\boldsymbol{e}_2\rangle & \dots & \langle \boldsymbol{e}_2, \mathsf{A}\boldsymbol{e}_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \langle \boldsymbol{e}_n, \mathsf{A}\boldsymbol{e}_2\rangle & \dots & \langle \boldsymbol{e}_n, \mathsf{A}\boldsymbol{e}_n\rangle \end{pmatrix}.$$

Let us denote the $n-1 \times n-1$ lower right block of this matrix by V_{n-1}. This is a Hermitian matrix of size $n-1$ so, using the induction hypothesis, we deduce that $\mathsf{V}_{n-1} = \mathsf{Q}_{n-1}\mathsf{D}_{n-1}\mathsf{Q}_{n-1}^*$ for appropriate matrices $\mathsf{Q}_{n-1} \in \mathbf{C}^{n-1\times n-1}$ and $\mathsf{D}_{n-1} \in \mathbf{R}^{n-1\times n-1}$ which are unitary and diagonal, respectively.

Step 3. Constructing Q and D. Define now

$$\mathsf{Q} = \mathsf{V}\begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{Q}_{n-1} \end{pmatrix}.$$

It is not difficult to verify that Q is a unitary matrix, and we have

$$\mathsf{Q}^*\mathsf{AQ} = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{Q}_{n-1}^* \end{pmatrix} \mathsf{V}^*\mathsf{AV} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{Q}_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{Q}_{n-1}^* \end{pmatrix} \begin{pmatrix} \lambda_1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{V}_{n-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{Q}_{n-1} \end{pmatrix}.$$

Developing the last expression, we obtain

$$\mathsf{Q}^*\mathsf{AQ} = \begin{pmatrix} \lambda_1 & \mathbf{0}^T \\ \mathbf{0} & \mathsf{D}_{n-1} \end{pmatrix},$$

which concludes the proof. □

We deduce, as a corollary of the spectral theorem, that if $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are eigenvectors of a Hermitian matrix associated with different eigenvalues, then they are necessarily orthogonal for the Euclidean inner product. Indeed, since $\mathsf{A} = \mathsf{A}^*$ and the eigenvalues are real, it holds that

$$\begin{aligned} (\lambda_1 - \lambda_2)\langle \boldsymbol{e}_1, \boldsymbol{e}_2\rangle &= \langle \lambda_1 \boldsymbol{e}_1, \boldsymbol{e}_2\rangle - \langle \boldsymbol{e}_1, \overline{\lambda}_2 \boldsymbol{e}_2\rangle \\ &= \langle \mathsf{A}\boldsymbol{e}_1, \boldsymbol{e}_2\rangle - \langle \boldsymbol{e}_1, \mathsf{A}\boldsymbol{e}_2\rangle = \langle \mathsf{A}\boldsymbol{e}_1, \boldsymbol{e}_2\rangle - \langle \mathsf{A}^*\boldsymbol{e}_1, \boldsymbol{e}_2\rangle = 0. \end{aligned}$$

The largest eigenvalue of a matrix, in modulus, is called the *spectral radius* and denoted by ρ. The following result relates the 2-norm of a matrix to the spectral radius of AA^*.

Proposition A.5. *It holds that* $\|\mathsf{A}\|_2 = \sqrt{\rho(\mathsf{A}^*\mathsf{A})}$.

Proof. Since $\mathsf{A}^*\mathsf{A}$ is Hermitian, it holds by the spectral theorem that $\mathsf{A}^*\mathsf{A} = \mathsf{QDQ}^*$ for some unitary matrix Q and real diagonal matrix D. Therefore, denoting by $(\mu_i)_{1\leqslant i\leqslant n}$ the (positive) diagonal elements of D and introducing $\boldsymbol{y} := \mathsf{Q}^*\boldsymbol{x}$, we have

$$
\begin{aligned}
\|\mathsf{A}\boldsymbol{x}\| &= \sqrt{\boldsymbol{x}^*\mathsf{A}^*\mathsf{A}\boldsymbol{x}} = \sqrt{\boldsymbol{x}^*\mathsf{QDQ}^*\boldsymbol{x}} \\
&= \sqrt{\sum_{i=1}^{n} \mu_i y_i^2} \leqslant \sqrt{\rho(\mathsf{A}^*\mathsf{A})}\sqrt{\sum_{i=1}^{n} y_i^2} = \sqrt{\rho(\mathsf{A}^*\mathsf{A})}\|\boldsymbol{x}\|,
\end{aligned}
\tag{A.15}
$$

where we used in the last equality the fact that $\boldsymbol{y}$ has the same norm as $\boldsymbol{x}$, because Q is unitary. It follows from (A.15) that $\|\mathsf{A}\| \leqslant \sqrt{\rho(\mathsf{A}^*\mathsf{A})}$, and the converse inequality also holds true since $\|\mathsf{A}\boldsymbol{x}\| = \sqrt{\rho(\mathsf{A}^*\mathsf{A})}\|\boldsymbol{x}\|$ if $\boldsymbol{x}$ is the eigenvector of $\mathsf{A}^*\mathsf{A}$ corresponding to an eigenvalue of modulus $\rho(\mathsf{A}^*\mathsf{A})$. □

To conclude this section, we recall and prove the Courant–Fisher theorem.

Theorem A.6 (Courant–Fisher Min-Max theorem)**.** *The eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ of a Hermitian matrix are characterized by the relation*

$$
\lambda_k = \max_{\mathcal{S},\dim(\mathcal{S})=k} \left(\min_{\boldsymbol{x}\in\mathcal{S}\setminus\{0\}} \frac{\boldsymbol{x}^*\mathsf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} \right). \tag{A.16}
$$

Proof. Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ be normalized and pairwise orthogonal eigenvectors associated with the eigenvalues $\lambda_1, \dots, \lambda_n$, and let $\mathcal{S}_k = \operatorname{Span}\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$. Any $\boldsymbol{x} \in \mathcal{S}_k$ may be expressed as a linear combination $\boldsymbol{x} = \alpha_1\boldsymbol{v}_1 + \cdots + \alpha_k\boldsymbol{v}_k$, and so

$$
\forall \boldsymbol{x} \in \mathcal{S}_k, \qquad \frac{\boldsymbol{x}^*\mathsf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} = \frac{\sum_{i=1}^{k} \lambda_i|\alpha_i|^2}{\sum_{i=1}^{k}|\alpha_i|^2} \geqslant \lambda_k.
$$

Therefore, it holds that

$$
\min_{\boldsymbol{x}\in\mathcal{S}_k\setminus\{0\}} \frac{\boldsymbol{x}^*\mathsf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} \geqslant \lambda_k,
$$

which proves the $\geqslant$ direction of (A.16). For the $\leqslant$ direction, let $\mathcal{U}_k = \operatorname{Span}\{\boldsymbol{v}_k, \dots \boldsymbol{v}_n\}$. Using a well-known result from linear algebra, we calculate that, for any subspace $\mathcal{S} \subset \mathbf{C}^n$ of dimension k,

$$
\begin{aligned}
\dim(\mathcal{S} \cap \mathcal{U}_k) &= \dim(\mathcal{S}) + \dim(\mathcal{U}_k) - \dim(\mathcal{S} + \mathcal{U}_k) \\
&\geqslant k + (n - k + 1) - n = 1.
\end{aligned}
$$

Therefore, any $\mathcal{S} \subset \mathbf{C}^n$ of dimension k has a nonzero intersection with $\mathcal{U}_k$. But since any vector in $\mathcal{U}_k$ can be expanded as $\beta_1\boldsymbol{v}_k + \cdots + \beta_n\boldsymbol{v}_n$, we have

$$
\forall \boldsymbol{x} \in \mathcal{U}_k, \qquad \frac{\boldsymbol{x}^*\mathsf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} = \frac{\sum_{i=k}^{n} \lambda_i|\alpha_i|^2}{\sum_{i=k}^{n}|\alpha_i|^2} \leqslant \lambda_k.
$$

This shows that

$$
\forall \mathcal{S} \subset \mathbf{C}^n \text{ with } \dim(\mathcal{S}) = k, \qquad \min_{\boldsymbol{x}\in\mathcal{S}\setminus\{0\}} \frac{\boldsymbol{x}^*\mathsf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} \leqslant \lambda_k,
$$

which enables to conclude the proof. □

Exercise A.7. *Prove that if $\mathsf{A} \in \mathbf{R}^{n\times n}$ is diagonalizable as in (A.14), then $\mathsf{A}^n = \mathsf{PD}^n\mathsf{P}^{-1}$.*

A.7 Similarity transformation and Jordan normal form

In this section, we work with matrices in $\mathbf{C}^{n\times n}$. A *similarity transformation* is a mapping of the type $\mathbf{C}^{n\times n} \ni \mathsf{A} \mapsto \mathsf{P}^{-1}\mathsf{A}\mathsf{P} \in \mathbf{C}^{n\times n}$, where $\mathsf{P} \in \mathbf{C}^{n\times n}$ is a nonsingular matrix. If two matrices are related by a similarity transformation, then they are called *similar*, because they may be viewed as two representations of the same linear mapping in different bases.

Definition A.8 (Jordan block)**.** A Jordan block with dimension n is a matrix of the form

$$\mathsf{J}_n(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$$

The parameter $\lambda \in \mathbf{C}$ is called the eigenvalue of the Jordan block.

A Jordan block is diagonalizable if and only if it is of dimension 1. The only eigenvector of a Jordan block is $\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^T$. The power of a Jordan block admits an explicit expression.

Lemma A.7. *It holds that*

$$\mathsf{J}_n(\lambda)^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots & \dots & \binom{k}{n-1}\lambda^{k-n+1} \\ & \lambda^k & \binom{k}{1}\lambda^{k-1} & \dots & \dots & \binom{k}{n-2}\lambda^{k-n+2} \\ & & \ddots & \ddots & & \vdots \\ & & & \ddots & \ddots & \vdots \\ & & & & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ & & & & & \lambda^k \end{pmatrix}. \tag{A.17}$$

Proof. The explicit expression of the Jordan block can be obtained by decomposing the block as $\mathsf{J}_n(\lambda) = \lambda\mathsf{I} + \mathsf{N}$ and using the binomial formula:

$$(\lambda\mathsf{I} + \mathsf{N})^k = \sum_{i=1}^{k} \binom{k}{i} (\lambda\mathsf{I})^{k-i}\mathsf{N}^i.$$

To conclude the proof, we use the fact that N^i is a matrix with zeros everywhere except for i-th super-diagonal, which contains only ones. Moreover $\mathsf{N}^i = 0_{n\times n}$ if $i \geqslant n$. ☐

A matrix is said to be of *Jordan normal form* if it is block-diagonal with Jordan blocks on the diagonal. In other words, a matrix $\mathsf{J} \in \mathbf{C}^{n\times n}$ is of Jordan normal form if

$$\mathsf{J} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & & \\ & J_{n_2}(\lambda_2) & & & \\ & & \ddots & & \\ & & & J_{n_{k-1}}(\lambda_{k-1}) & \\ & & & & J_{n_k}(\lambda_k) \end{pmatrix}$$

with $n_1 + \cdots + n_k = n$. Note that $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of A. We state without proof the following important result.

Proposition A.8 (Jordan normal form). *Any matrix $\mathsf{A} \in \mathbf{C}^{n\times n}$ is* similar *to a matrix in Jordan normal form. In other words, there exists an invertible matrix $\mathsf{P} \in \mathbf{C}^{n\times n}$ and a matrix in normal Jordan form $\mathsf{J} \in \mathbf{C}^{n\times n}$ such that*

$$\mathsf{A} = \mathsf{P}\mathsf{J}\mathsf{P}^{-1}$$

A.8 Oldenburger's theorem and Gelfand's formula

The following result establishes a necessary and sufficient condition for the convergence of $\|\mathsf{A}^k\|$ to 0 in terms of the spectral radius of A, and for any matrix norm $\|\bullet\|$.

Proposition A.9 (Oldenburger). *Let $\rho(\mathsf{A})$ denote the spectral radius of $\mathsf{A} \in \mathbf{C}^{n\times n}$ and $\|\bullet\|$ be a matrix norm. Then*

- $\|\mathsf{A}^k\| \to 0$ *in the limit as $k \to \infty$ if and only if $\rho(\mathsf{A}) < 1$.*
- $\|\mathsf{A}^k\| \to \infty$ *in the limit as $k \to \infty$ if and only if $\rho(\mathsf{A}) > 1$.*

Proof. Since all matrix norms are equivalent, we can assume without loss of generality that $\|\bullet\|$ is the 2-norm. We prove only the equivalence $\|\mathsf{A}^k\| \to 0 \Leftrightarrow \rho(\mathsf{A}) < 1$. The other statement can be proved similarly. By Proposition A.8, there exists a nonsingular matrix P such that $\mathsf{A} = \mathsf{P}\mathsf{J}\mathsf{P}^{-1}$, for a matrix $\mathsf{J} \in \mathbf{C}^{n\times n}$ which is in normal Jordan form. Since $\rho(\mathsf{A}) = \rho(\mathsf{J})$ and $\|\mathsf{A}^k\| \to 0$ if and only if $\|\mathsf{J}^k\| \to 0$, it is sufficient to show that $\|\mathsf{J}^k\| \to 0 \Leftrightarrow \rho(\mathsf{J}) < 1$. The latter statement follows from the expression of the power of a Jordan block given in Lemma A.7. □

With this result, we can prove Gelfand's formula, which relates the spectral radius to the asymptotic growth of $\|\mathsf{A}^k\|$, and is used in Chapter 4.

Proposition A.10 (Gelfand's formula). *Let $\mathsf{A} \in \mathbf{C}^{n\times n}$. It holds for any norm that*

$$\lim_{k\to\infty} \|\mathsf{A}^k\|^{\frac{1}{k}} = \rho(\mathsf{A})$$

Proof. Let $0 < \varepsilon < \rho(\mathsf{A})$ and define $\mathsf{A}^+ = \frac{\mathsf{A}}{\rho(\mathsf{A})+\varepsilon}$ and $\mathsf{A}^- = \frac{\mathsf{A}}{\rho(\mathsf{A})-\varepsilon}$. It holds by construction that $\rho(\mathsf{A}^+) < 1$ and $\rho(\mathsf{A}^-) > 1$. Using Proposition A.9, we deduce that

$$\lim_{k\to\infty} \|(\mathsf{A}^+)^k\| = 0, \qquad \lim_{k\to\infty} \|(\mathsf{A}^-)^k\| = \infty.$$

Therefore, it holds that

$$\limsup_{k\to\infty} \left\|(\mathsf{A}^+)^k\right\|^{\frac{1}{k}} \leqslant 1, \qquad \liminf_{k\to\infty} \left\|(\mathsf{A}^-)^k\right\|^{\frac{1}{k}} \geqslant 1.$$

Substituting the expressions of A^+ and A^-, we deduce that

$$\limsup_{k\to\infty} \left\|\mathsf{A}^k\right\|^{\frac{1}{k}} \leqslant \rho(\mathsf{A}) + \varepsilon, \qquad \liminf_{k\to\infty} \left\|\mathsf{A}^k\right\|^{\frac{1}{k}} \geqslant \rho(\mathsf{A}) - \varepsilon.$$

Since ε was arbitrary, we obtain that

$$\rho(\mathsf{A}) \leqslant \liminf_{k\to\infty} \left\|\mathsf{A}^k\right\|^{\frac{1}{k}} \leqslant \limsup_{k\to\infty} \left\|\mathsf{A}^k\right\|^{\frac{1}{k}} \leqslant \rho(\mathsf{A}),$$

which implies the statement. □