Chapter 1
Mathematical Preliminaries

In this short chapter we will review some important notions and results from calculus, linear algebra, and topology that will be frequently used throughout the book. This chapter is not intended to be, by any means, a comprehensive treatment of these subjects, and the interested reader can find more material in advanced calculus and linear algebra books.

1.1 ▪ The Space $\mathbb{R}^n$

The vector space $\mathbb{R}^n$ is the set of n-dimensional column vectors with real components endowed with the component-wise addition operator

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

and the scalar-vector product

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix},$$

where in the above $x_1, x_2, \dots, x_n, \lambda$ are real numbers. Throughout the book we will be mainly interested in problems over $\mathbb{R}^n$, although other vector spaces will be considered in a few cases. We will denote the standard basis of $\mathbb{R}^n$ by $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$, where $\mathbf{e}_i$ is the n-length column vector whose ith component is one while all the others are zeros. The column vectors of all ones and all zeros will be denoted by $\mathbf{e}$ and $\mathbf{0}$, respectively, where the length of the vectors will be clear from the context.

Important Subsets of $\mathbb{R}^n$

The *nonnegative orthant* is the subset of $\mathbb{R}^n$ consisting of all vectors in $\mathbb{R}^n$ with nonnegative components and is denoted by $\mathbb{R}^n_+$:

$$\mathbb{R}^n_+ = \left\{ (x_1, x_2, \dots, x_n)^T : x_1, x_2, \dots, x_n \geq 0 \right\}.$$

Similarly, the *positive orthant* consists of all the vectors in $\mathbb{R}^n$ with positive components and is denoted by $\mathbb{R}^n_{++}$:

$$\mathbb{R}^n_{++} = \left\{(x_1, x_2, \ldots, x_n)^T : x_1, x_2, \ldots, x_n > 0\right\}.$$

For given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the *closed line segment* between $\mathbf{x}$ and $\mathbf{y}$ is a subset of $\mathbb{R}^n$ denoted by $[\mathbf{x}, \mathbf{y}]$ and defined as

$$[\mathbf{x}, \mathbf{y}] = \{\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}) : \alpha \in [0, 1]\}.$$

The *open line segment* $(\mathbf{x}, \mathbf{y})$ is similarly defined as

$$(\mathbf{x}, \mathbf{y}) = \{\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}) : \alpha \in (0, 1)\}$$

when $\mathbf{x} \neq \mathbf{y}$ and is the empty set $\emptyset$ when $\mathbf{x} = \mathbf{y}$. The *unit-simplex*, denoted by Δ_n, is the subset of $\mathbb{R}^n$ comprising all nonnegative vectors whose sum is 1:

$$\Delta_n = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}, \mathbf{e}^T \mathbf{x} = 1\right\}.$$

1.2 ▪ The Space $\mathbb{R}^{m \times n}$

The set of all real-valued matrices of order $m \times n$ is denoted by $\mathbb{R}^{m \times n}$. Some special matrices that will be frequently used are the $n \times n$ identity matrix denoted by $\mathbf{I}_n$ and the $m \times n$ zeros matrix denoted by $\mathbf{0}_{m \times n}$. We will frequently omit the subscripts of these matrices when the dimensions will be clear from the context.

1.3 ▪ Inner Products and Norms

Inner Products

We begin with the formal definition of an *inner product*.

Definition 1.1 (inner product). *An* **inner product** *on $\mathbb{R}^n$ is a map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with the following properties:*

1. **(symmetry)** $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ *for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
2. **(additivity)** $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$ *for any* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.
3. **(homogeneity)** $\langle \lambda \mathbf{x}, \mathbf{y} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle$ *for any* $\lambda \in \mathbb{R}$ *and* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
4. **(positive definiteness)** $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ *for any* $\mathbf{x} \in \mathbb{R}^n$ *and* $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

Example 1.2. Perhaps the most widely used inner product is the so-called *dot product* defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i \text{ for any } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Since this is in a sense the "standard" inner product, we will by default assume—unless explicitly stated otherwise—that the underlying inner product is the dot product. ∎

Example 1.3. The dot product is not the only possible inner product on $\mathbb{R}^n$. For example, let $\mathbf{w} \in \mathbb{R}^n_{++}$. Then it is easy to show that the following weighted dot product is also an inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} = \sum_{i=1}^n w_i x_i y_i. \quad ∎$$

Vector Norms

Definition 1.4 (norm). *A* **norm** $||\cdot||$ *on* $\mathbb{R}^n$ *is a function* $||\cdot||:\mathbb{R}^n \to \mathbb{R}$ *satisfying the following:*

1. **(nonnegativity)** $||\mathbf{x}|| \geq 0$ *for any* $\mathbf{x} \in \mathbb{R}^n$ *and* $||\mathbf{x}|| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.
2. **(positive homogeneity)** $||\lambda \mathbf{x}|| = |\lambda| ||\mathbf{x}||$ *for any* $\mathbf{x} \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$.
3. **(triangle inequality)** $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ *for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

One natural way to generate a norm on $\mathbb{R}^n$ is to take any inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ and define the associated norm

$$||\mathbf{x}|| \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \text{ for all } \mathbf{x} \in \mathbb{R}^n,$$

which can be easily seen to be a norm. If the inner product is the dot product, then the associated norm is the so-called *Euclidean norm* or l_2*-norm*:

$$||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^n x_i^2} \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

By default, the underlying norm on $\mathbb{R}^n$ is $||\cdot||_2$, and the subscript 2 will be frequently omitted. The Euclidean norm belongs to the class of l_p norms (for $p \geq 1$) defined by

$$||\mathbf{x}||_p \equiv \sqrt[p]{\sum_{i=1}^n |x_i|^p}.$$

The restriction $p \geq 1$ is necessary since for $0 < p < 1$, the function $||\cdot||_p$ is actually not a norm (see Exercise 1.1). Another important norm is the l_∞ norm given by

$$||\mathbf{x}||_\infty \equiv \max_{i=1,2,\ldots,n} |x_i|.$$

Unsurprisingly, it can be shown (see Exercise 1.2) that

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

An important inequality connecting the dot product of two vectors and their norms is the Cauchy–Schwarz inequality, which will be used frequently throughout the book.

Lemma 1.5 (Cauchy–Schwarz inequality). *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$|\mathbf{x}^T \mathbf{y}| \leq ||\mathbf{x}||_2 \cdot ||\mathbf{y}||_2.$$

Equality is satisfied if and only if $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly dependent.*

Matrix Norms

Similarly to vector norms, we can define the concept of a *matrix norm*.

Definition 1.6. *A* **norm** $\|\cdot\|$ *on* $\mathbb{R}^{m\times n}$ *is a function* $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ *satisfying the following:*

1. **(nonnegativity)** $\|\mathbf{A}\| \geq 0$ *for any* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\|\mathbf{A}\| = 0$ *if and only if* $\mathbf{A} = \mathbf{0}$.
2. **(positive homogeneity)** $\|\lambda\mathbf{A}\| = |\lambda| \cdot \|\mathbf{A}\|$ *for any* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\lambda \in \mathbb{R}$.
3. **(triangle inequality)** $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ *for any* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$.

Many examples of matrix norms are generated by using the concept of induced norms, which we now describe. Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, the *induced matrix norm* $\|\mathbf{A}\|_{a,b}$ is defined by

$$\|\mathbf{A}\|_{a,b} = \max_{\mathbf{x}} \{\|\mathbf{A}\mathbf{x}\|_b : \|\mathbf{x}\|_a \leq 1\}.$$

It can be shown that the above definition implies that for any $\mathbf{x} \in \mathbb{R}^n$ the inequality

$$\|\mathbf{A}\mathbf{x}\|_b \leq \|\mathbf{A}\|_{a,b}\|\mathbf{x}\|_a$$

holds. An induced matrix norm is indeed a norm in the sense that it satisfies the three properties required from a matrix norm (see Definition 1.6): nonnegativity, positive homogeneity, and the triangle inequality. We refer to the matrix norm $\|\cdot\|_{a,b}$ as the (a,b)-norm. When $a = b$ (for example, when the two vector norms are l_a norms), we will simply refer to it as an a-norm and omit one of the subscripts in its notation; that is, the notation is $\|\cdot\|_a$ instead of $\|\cdot\|_{a,a}$.

Example 1.7 (spectral norm). If $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_2$, then the induced norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the maximum singular value of $\mathbf{A}$:

$$\|\mathbf{A}\|_2 = \|\mathbf{A}\|_{2,2} = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})} \equiv \sigma_{\max}(\mathbf{A}).$$

Since the Euclidean norm is the "standard" vector norm, the induced norm, namely the spectral norm, will be the standard matrix norm, and thus the subscripts of this norm will usually be omitted. ∎

Example 1.8 (1-norm). When $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$, the induced matrix norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is given by

$$\|\mathbf{A}\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |A_{i,j}|.$$

This norm is also called the *maximum absolute column sum* norm. ∎

Example 1.9 (∞-norm). When $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$, then the induced matrix norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is given by

$$\|\mathbf{A}\|_\infty = \max_{i=1,2,\ldots,m} \sum_{j=1}^{n} |A_{i,j}|.$$

This norm is also called the *maximum absolute row sum norm*. ∎

An example of a matrix norm that is not an *induced* norm is the *Frobenius norm* defined by

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}, \quad \mathbf{A} \in \mathbb{R}^{m\times n}.$$

1.4 ▪ Eigenvalues and Eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then a nonzero vector $\mathbf{v} \in \mathbb{C}^n$ is called an *eigenvector* of $\mathbf{A}$ if there exists a $\lambda \in \mathbb{C}$ (the complex field) for which

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}.$$

The scalar λ is the *eigenvalue* corresponding to the eigenvector $\mathbf{v}$. In general, real-valued matrices can have complex eigenvalues, but it is well known that all the eigenvalues of symmetric matrices are real. The eigenvalues of a symmetric $n \times n$ matrix $\mathbf{A}$ are denoted by

$$\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A}).$$

The maximum eigenvalue is also denoted by $\lambda_{\max}(\mathbf{A})(= \lambda_1(\mathbf{A}))$ and the minimum eigenvalue is also denoted by $\lambda_{\min}(\mathbf{A})(= \lambda_n(\mathbf{A}))$. One of the most useful results related to eigenvalues is the spectral decomposition theorem, which states that any symmetric matrix $\mathbf{A}$ has an orthonormal basis of eigenvectors.

Theorem 1.10 (spectral decomposition theorem). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an $n \times n$ symmetric matrix. Then there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ ($\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$) and a diagonal matrix $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ for which*

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{D}. \tag{1.1}$$

The columns of the matrix $\mathbf{U}$ in the factorization (1.1) constitute an orthonormal basis comprised of eigenvectors of $\mathbf{A}$ and the diagonal elements of $\mathbf{D}$ are the corresponding eigenvalues. A direct result of the spectral decomposition theorem is that the trace and the determinant of a matrix can be expressed via its eigenvalues:

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i(\mathbf{A}), \tag{1.2}$$

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i(\mathbf{A}). \tag{1.3}$$

Another important consequence of the spectral decomposition theorem is the bounding of the so-called *Rayleigh quotient*. For a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, the Rayleigh quotient is defined by

$$R_{\mathbf{A}}(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} \text{ for any } \mathbf{x} \neq \mathbf{0}.$$

We can now use the spectral decomposition theorem to prove the following lemma providing lower and upper bounds on the Rayleigh quotient.

Lemma 1.11. *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric. Then*

$$\lambda_{\min}(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{x}) \leq \lambda_{\max}(\mathbf{A}) \textit{ for any } \mathbf{x} \neq \mathbf{0}.$$

Proof. By the spectral decomposition theorem there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ such that $\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{D}$. We can assume without the loss of generality that the diagonal elements of $\mathbf{D}$, which are the eigenvalues of $\mathbf{A}$, are ordered nonincreasingly: $d_1 \geq d_2 \geq \cdots \geq d_n$, where $d_1 = \lambda_{\max}(\mathbf{A})$ and

$d_n = \lambda_{\min}(\mathbf{A})$. Making the change of variables $\mathbf{x} = \mathbf{U}\mathbf{y}$ and noting that $\mathbf{U}$ is a nonsingular matrix, we obtain that

$$\max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} = \max_{\mathbf{y}\neq 0} \frac{\mathbf{y}^T\mathbf{U}^T\mathbf{A}\mathbf{U}\mathbf{y}}{\|\mathbf{U}\mathbf{y}\|^2} = \max_{\mathbf{y}\neq 0} \frac{\mathbf{y}^T\mathbf{D}\mathbf{y}}{\|\mathbf{y}\|^2} = \max_{\mathbf{y}\neq 0} \frac{\sum_{i=1}^n d_i y_i^2}{\sum_{i=1}^n y_i^2}.$$

Since $d_i \leq d_1$ for all $i = 1, 2, \ldots, n$, it follows that $\sum_{i=1}^n d_i y_i^2 \leq d_1(\sum_{i=1}^n y_i^2)$, and hence

$$R_{\mathbf{A}}(\mathbf{x}) = \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} \leq \max_{\mathbf{y}\neq 0} \frac{d_1(\sum_{i=1}^n y_i^2)}{\sum_{i=1}^n y_i^2} = d_1 = \lambda_{\max}(\mathbf{A}).$$

The inequality $R_{\mathbf{A}}(\mathbf{x}) \geq \lambda_{\min}(\mathbf{A})$ follows by a similar argument. $\square$

The lower and upper bounds on the Rayleigh quotient given in the last lemma are attained at eigenvectors corresponding to the minimal and maximal eigenvalues respectively. Indeed, if $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors corresponding to the minimal and maximal eigenvalues respectively, then

$$R_{\mathbf{A}}(\mathbf{v}) = \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\|\mathbf{v}\|^2} = \frac{\lambda_{\min}(\mathbf{A})\|\mathbf{v}\|^2}{\|\mathbf{v}\|^2} = \lambda_{\min}(\mathbf{A}),$$
$$R_{\mathbf{A}}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{A}\mathbf{w}}{\|\mathbf{w}\|^2} = \frac{\lambda_{\max}(\mathbf{A})\|\mathbf{w}\|^2}{\|\mathbf{w}\|^2} = \lambda_{\max}(\mathbf{A}).$$

The above facts are summarized in the following lemma.

Lemma 1.12. *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric. Then*

$$\min_{\mathbf{x}\neq 0} R_{\mathbf{A}}(\mathbf{x}) = \lambda_{\min}(\mathbf{A}), \tag{1.4}$$

and the eigenvectors of $\mathbf{A}$ corresponding to the minimal eigenvalue are minimizers of problem (1.4). *In addition,*

$$\max_{\mathbf{x}\neq 0} R_{\mathbf{A}}(\mathbf{x}) = \lambda_{\max}(\mathbf{A}), \tag{1.5}$$

and the eigenvectors of $\mathbf{A}$ corresponding to the maximal eigenvalue are maximizers of problem (1.5).

1.5 ▪ Basic Topological Concepts

We begin with the definition of a ball.

Definition 1.13 (open ball, closed ball). *The* **open ball** *with center $\mathbf{c} \in \mathbb{R}^n$ and radius r is denoted by $B(\mathbf{c}, r)$ and defined by*

$$B(\mathbf{c}, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| < r\}.$$

The **closed ball** *with center $\mathbf{c}$ and radius r is denoted by $B[\mathbf{c}, r]$ and defined by*

$$B[\mathbf{c}, r] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| \leq r\}.$$

Note that the norm used in the definition of the ball is not necessarily the Euclidean norm. As usual, if the norm is not specified, we will assume by default that it is the Euclidean norm. The ball $B(\mathbf{c}, r)$ for some arbitrary $r > 0$ is also referred to as a *neighborhood* of $\mathbf{c}$. The first topological notion we define is that of an *interior point* of a set. This is a point which has a neighborhood contained in the set.

Definition 1.14 (interior points). *Given a set $U \subseteq \mathbb{R}^n$, a point $\mathbf{c} \in U$ is an* **interior point** *of U if there exists $r > 0$ for which $B(\mathbf{c}, r) \subseteq U$.*

The set of all interior points of a given set U is called the *interior* of the set and is denoted by $\text{int}(U)$:

$$\text{int}(U) = \{\mathbf{x} \in U : B(\mathbf{x}, r) \subseteq U \text{ for some } r > 0\}.$$

Example 1.15. Following are some examples of interiors of sets which were previously discussed:

$$\begin{aligned} \text{int}(\mathbb{R}^n_+) &= \mathbb{R}^n_{++}, \\ \text{int}(B[\mathbf{c}, r]) &= B(\mathbf{c}, r) \quad (\mathbf{c} \in \mathbb{R}^n, r \in \mathbb{R}_{++}). \end{aligned} \qquad \blacksquare$$

Definition 1.16 (open sets). *An* **open set** *is a set that contains only interior points. In other words, $U \subseteq \mathbb{R}^n$ is an open set if*

$$\textit{for every } \mathbf{x} \in U \textit{ there exists } r > 0 \textit{ such that } B(\mathbf{x}, r) \subseteq U.$$

Examples of open sets are $\mathbb{R}^n$, open balls (hence the name), and the positive orthant $\mathbb{R}^n_{++}$. A known result is that a union of any number of open sets is an open set and the intersection of a finite number of open sets is open.

Definition 1.17 (closed sets). *A set $U \subseteq \mathbb{R}^n$ is said to be* **closed** *if it contains all the limits of convergent sequences of points in U; that is, U is closed if for every sequence of points $\{\mathbf{x}_i\}_{i \geq 1} \subseteq U$ satisfying $\mathbf{x}_i \to \mathbf{x}^*$ as $i \to \infty$, it holds that $\mathbf{x}^* \in U$.*

A known property is that a set U is closed if and only if its complement U^c is open (see Exercise 1.15). Examples of closed sets are the closed ball $B[\mathbf{c}, r]$, closed lines segments, the nonnegative orthant $\mathbb{R}^n_+$, and the unit simplex Δ_n. The space $\mathbb{R}^n$ is both closed and open. An important and useful result states that level sets, as well as contour sets, of continuous functions are closed. This is stated in the following proposition.

Proposition 1.18 (closedness of level and contour sets of continuous functions). *Let f be a continuous function defined over a closed set $S \subseteq \mathbb{R}^n$. Then for any $\alpha \in \mathbb{R}$ the sets*

$$\begin{aligned} Lev(f, \alpha) &= \{\mathbf{x} \in S : f(\mathbf{x}) \leq \alpha\}, \\ Con(f, \alpha) &= \{\mathbf{x} \in S : f(\mathbf{x}) = \alpha\} \end{aligned}$$

are closed.

Definition 1.19 (boundary points). *Given a set $U \subseteq \mathbb{R}^n$, a* **boundary point** *of U is a point $\mathbf{x} \in \mathbb{R}^n$ satisfying the following: any neighborhood of $\mathbf{x}$ contains at least one point in U and at least one point in its complement U^c.*

The set of all boundary points of a set U is denoted by $\text{bd}(U)$ and is called *the boundary of U*.

Example 1.20. Some examples of boundary sets are

$$\begin{aligned}
\mathrm{bd}(B(\mathbf{c},r)) = \mathrm{bd}(B[\mathbf{c},r]) &= \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}-\mathbf{c}\| = r\} \quad (\mathbf{c}\in\mathbb{R}^n, r\in\mathbb{R}_{++}),\\
\mathrm{bd}(\mathbb{R}^n_{++}) = \mathrm{bd}(\mathbb{R}^n_{+}) &= \left\{\mathbf{x}\in\mathbb{R}^n_+ : \exists i : x_i = 0\right\},\\
\mathrm{bd}(\mathbb{R}^n) &= \emptyset. \quad \blacksquare
\end{aligned}$$

The *closure* of a set $U\subseteq\mathbb{R}^n$ is denoted by $\mathrm{cl}(U)$ and is defined to be the smallest closed set containing U:

$$\mathrm{cl}(U) = \bigcap\{T : U\subseteq T, T \text{ is closed}\}.$$

The closure set is indeed a closed set as an intersection of closed sets. Another equivalent definition of $\mathrm{cl}(U)$ is given by

$$\mathrm{cl}(U) = U\cup\mathrm{bd}(U).$$

Example 1.21.

$$\begin{aligned}
\mathrm{cl}(\mathbb{R}^n_{++}) &= \mathbb{R}^n_+,\\
\mathrm{cl}(B(\mathbf{c},r)) &= B[\mathbf{c},r] \quad (\mathbf{c}\in\mathbb{R}^n, r\in\mathbb{R}_{++}),\\
\mathrm{cl}((\mathbf{x},\mathbf{y})) &= [\mathbf{x},\mathbf{y}] \quad (\mathbf{x},\mathbf{y}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{y}). \quad \blacksquare
\end{aligned}$$

Definition 1.22 (boundedness and compactness).

1. *A set $U\subseteq\mathbb{R}^n$ is called* **bounded** *if there exists $M>0$ for which $U\subseteq B(\mathbf{0},M)$.*

2. *A set $U\subseteq\mathbb{R}^n$ is called* **compact** *if it is closed and bounded.*

Examples of compact sets are closed balls and line segments. The positive orthant is not compact since it is unbounded, and open balls are not compact since they are not closed.

1.5.1 ▪ Differentiability

Let f be a function defined on a set $S\subseteq\mathbb{R}^n$. Let $\mathbf{x}\in\mathrm{int}(S)$ and let $\mathbf{0}\neq\mathbf{d}\in\mathbb{R}^n$. If the limit

$$\lim_{t\to 0^+}\frac{f(\mathbf{x}+t\mathbf{d})-f(\mathbf{x})}{t}$$

exists, then it is called *the directional derivative* of f at $\mathbf{x}$ along the direction $\mathbf{d}$ and is denoted by $f'(\mathbf{x};\mathbf{d})$. For any $i=1,2,\ldots,n$, if the limit

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{t\to 0}\frac{f(\mathbf{x}+t\mathbf{e}_i)-f(\mathbf{x})}{t}$$

exists, then it is called *the ith partial derivative* of f at $\mathbf{x}$.

If all the partial derivatives of a function f exist at a point $\mathbf{x} \in \mathbb{R}^n$, then the *gradient* of f at $\mathbf{x}$ is defined to be the column vector consisting of all the partial derivatives:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called *continuously differentiable over U* if all the partial derivatives exist and are continuous on U. The definition of continuous differentiability can also be extended to nonopen sets by using the convention that a function f is said to be continuously differentiable over a set C if there exists an open set U containing C on which the function is also defined and continuously differentiable. In the setting of continuous differentiability, we have the following important formula for the directional derivative:

$$f'(\mathbf{x}; \mathbf{d}) = \nabla f(\mathbf{x})^T \mathbf{d}$$

for all $\mathbf{x} \in U$ and $\mathbf{d} \in \mathbb{R}^n$. It can also be shown in this setting of continuous differentiability that the following approximation result holds.

Proposition 1.23. *Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is continuously differentiable over U. Then*

$$\lim_{\mathbf{d}\to 0} \frac{f(\mathbf{x}+\mathbf{d}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^T \mathbf{d}}{\|\mathbf{d}\|} = 0 \textit{ for all } \mathbf{x} \in U.$$

Another way to write the above result is as follows:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + o(\|\mathbf{y} - \mathbf{x}\|),$$

where $o(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ is a one-dimensional function satisfying $\frac{o(t)}{t} \to 0$ as $t \to 0^+$.

The partial derivatives $\frac{\partial f}{\partial x_i}$ are themselves real-valued functions that can be partially differentiated. The (i,j)th partial derivatives of f at $\mathbf{x} \in U$ (if it exists) is defined by

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial \left(\frac{\partial f}{\partial x_j} \right)}{\partial x_i}(\mathbf{x}).$$

A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called *twice continuously differentiable over U* if all the second order partial derivatives exist and are continuous over U. Under the assumption of twice continuous differentiability, the second order partial derivatives are symmetric, meaning that for any $i \neq j$ and any $\mathbf{x} \in U$

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}).$$

The *Hessian* of f at a point $\mathbf{x} \in U$ is the $n \times n$ matrix

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix},$$

where all the second order partial derivatives are evaluated at $\mathbf{x}$. Since f is twice continuously differentiable over U, the Hessian matrix is symmetric. There are two main approximation results (linear and quadratic) which are direct consequences of Taylor's approximation theorem that will be used frequently in the book and are thus recalled here.

Theorem 1.24 (linear approximation theorem). *Let $f : U \to \mathbb{R}$ be a twice continuously differentiable function over an open set $U \subseteq \mathbb{R}^n$, and let $\mathbf{x} \in U, r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$ there exists $\xi \in [\mathbf{x}, \mathbf{y}]$ such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\xi)(\mathbf{y}-\mathbf{x}).$$

Theorem 1.25 (quadratic approximation theorem). *Let $f : U \to \mathbb{R}$ be a twice continuously differentiable function over an open set $U \subseteq \mathbb{R}^n$, and let $\mathbf{x} \in U, r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + o(\|\mathbf{y}-\mathbf{x}\|^2).$$

Exercises

1.1. Show that $\|\cdot\|_{1/2}$ is not a norm.

1.2. Prove that for any $\mathbf{x} \in \mathbb{R}^n$ one has

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

1.3. Show that for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$

$$\|\mathbf{x}-\mathbf{z}\| \le \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}-\mathbf{z}\|.$$

1.4. Prove the Cauchy–Schwarz inequality (Lemma 1.5). Show that equality holds if and only if the vectors $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

1.5. Suppose that $\mathbb{R}^m$ and $\mathbb{R}^n$ are equipped with norms $\|\cdot\|_b$ and $\|\cdot\|_a$, respectively. Show that the induced matrix norm $\|\cdot\|_{a,b}$ satisfies the triangle inequality. That is, for any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$ the inequality

$$\|\mathbf{A}+\mathbf{B}\|_{a,b} \le \|\mathbf{A}\|_{a,b} + \|\mathbf{B}\|_{a,b}$$

holds.

1.6. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Show that the norm function $f(\mathbf{x}) = \|\mathbf{x}\|$ is a continuous function over $\mathbb{R}^n$.

1.7. **(attainment of the maximum in the induced norm definition).** Suppose that $\mathbb{R}^m$ and $\mathbb{R}^n$ are equipped with norms $\|\cdot\|_b$ and $\|\cdot\|_a$, respectively, and let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Show that there exists $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{x}\|_a \le 1$ and $\|\mathbf{Ax}\|_b = \|\mathbf{A}\|_{a,b}$.

1.8. Suppose that $\mathbb{R}^m$ and $\mathbb{R}^n$ are equipped with norms $\|\cdot\|_b$ and $\|\cdot\|_a$, respectively. Show that the induced matrix norm $\|\cdot\|_{a,b}$ can be computed by the formula

$$\|\mathbf{A}\|_{a,b} = \max_{\mathbf{x}}\{\|\mathbf{Ax}\|_b : \|\mathbf{x}\|_a = 1\}.$$

1.9. Suppose that $\mathbb{R}^m$ and $\mathbb{R}^n$ are equipped with norms $\|\cdot\|_b$ and $\|\cdot\|_a$, respectively. Show that the induced matrix norm $\|\cdot\|_{a,b}$ can be computed by the formula

$$\|\mathbf{A}\|_{a,b} = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_b}{\|\mathbf{x}\|_a}.$$

1.10. Let $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times k}$ and assume that $\mathbb{R}^m, \mathbb{R}^n$, and $\mathbb{R}^k$ are equipped with the norms $\|\cdot\|_c, \|\cdot\|_b$, and $\|\cdot\|_a$, respectively. Prove that

$$\|\mathbf{AB}\|_{a,c} \le \|\mathbf{A}\|_{b,c}\|\mathbf{B}\|_{a,b}.$$

1.11. Prove the formula of the ∞-matrix norm given in Example 1.9.

1.12. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Prove that

(i) $\frac{1}{\sqrt{n}}\|\mathbf{A}\|_\infty \le \|\mathbf{A}\|_2 \le \sqrt{m}\|\mathbf{A}\|_\infty$,

(ii) $\frac{1}{\sqrt{m}}\|\mathbf{A}\|_1 \le \|\mathbf{A}\|_2 \le \sqrt{n}\|\mathbf{A}\|_1$.

1.13. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Show that

(i) $\|\mathbf{A}\| = \|\mathbf{A}^T\|$ (here $\|\cdot\|$ is the spectral norm),

(ii) $\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \lambda_i(\mathbf{A}^T\mathbf{A})$.

1.14. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Show that

$$\max_{\mathbf{x}}\{\mathbf{x}^T\mathbf{Ax} : \|\mathbf{x}\|^2 = 1\} = \lambda_{\max}(\mathbf{A}).$$

1.15. Prove that a set $U \subseteq \mathbb{R}^n$ is closed if and only if its complement U^c is open.

1.16. (i) Let $\{A_i\}_{i\in I}$ be a collection of open sets where I is a given index set. Show that $\bigcup_{i\in I} A_i$ is an open set. Show that if I is finite, then $\bigcap_{i\in I} A_i$ is open.

(ii) Let $\{A_i\}_{i\in I}$ be a collection of closed sets where I is a given index set. Show that $\bigcap_{i\in I} A_i$ is a closed set. Show that if I is finite, then $\bigcup_{i\in I} A_i$ is closed.

1.17. Give an example of open sets $A_i, i \in I$ for which $\bigcap_{i\in I} A_i$ is not open.

1.18. Let $A, B \subseteq \mathbb{R}^n$. Prove that $\text{cl}(A\cap B) \subseteq \text{cl}(A) \cap \text{cl}(B)$. Give an example in which the inclusion is proper.

1.19. Let $A, B \subseteq \mathbb{R}^n$. Prove that $\text{int}(A\cap B) = \text{int}(A) \cap \text{int}(B)$ and that $\text{int}(A) \cup \text{int}(B) \subseteq \text{int}(A\cup B)$. Show an example in which the latter inclusion is proper.